GPS for Geodesy, p. 231-270, Springer-Verlag, Berlin (1998)

6. GPS DATA PROCESSING METHODOLOGY: FROM THEORY TO APPLICATIONS

Geoffrey Blewitt
Department of Geomatics, University of Newcastle, Newcastle upon Tyne, NE1 7RU, United Kingdom

6.1 INTRODUCTION

The idea behind this chapter is to use a few fundamental concepts to help develop a way of thinking about GPS data processing that is intuitive, yet has a firm theoretical foundation. Intuition is based on distilling an alarming array of information into a few core concepts that are basically simple. The fundamental concepts I have chosen to explore and develop here are generally based on equivalence principles and symmetry in problems. This involves looking at the same thing from different ways, or looking at apparently different things in the same way. Using symmetry and equivalence, we can often discover elegant explanations to problems.

The ultimate goal is that, the reader will be able to see answers to apparently complicated questions from first principles. An immediate goal, is to use this theoretical-intuitive approach as a vehicle to introduce a broad variety of algorithms and their application to high precision geodesy.

6.1.1 Background

It is useful to begin by placing this work into context briefly, by listing some of the common features of GPS data processing for high precision geodesy:

User Input Processing
- operating system interface
- interactive user control
- automation (batch control, defaults, contingency rules, etc.)
- help (user manual, on-line syntax help, on-line module guide, etc.)

Data Preprocessing
- GPS observation files, site database, Earth rotation data, satellite ephemerides, surface meteorological data, water vapour radiometer data
- formatting
- tools (satellite removal, data windowing, concatenation, etc.)
- editing (detecting and removing outliers and cycle slips)
- thinning (data decimation, data smoothing)
- data transformation (double differencing, ionosphere-free combination, etc.)
- ambiguity initialisation (and possible resolution)

Observation Models
- nominal values for model parameters
- (observed - computed) observations and partial derivatives
- orbit dynamics and satellite orientation
- Earth rotation and surface kinematics
- media propagation (troposphere, ionosphere)
- clocks
- relativistic corrections (clocks, spacetime curvature)
- antenna reference point and phase centre offset
- antenna kinematics
- phase modelling (phase centre variation, polarisation, cycle ambiguity)

Parameter Estimation
- parameter selection
- stochastic model and a priori constraints
- inversion (specific algorithms, filtering, blocking techniques, etc.)
- residual analysis (outliers, cycle slips) and re-processing
- sensitivity analysis (to unestimated parameters)

Solution Processing
- a priori constraints
- ambiguity resolution
- solution combination and kinematic modelling
- frame projection and transformation tools
- statistics (formal errors, repeatability, in various coordinate systems, etc.)

Output Processing
- archive solution files
- information for the user
- export formatting (RINEX, SINEX, IONEX, etc.)
- presentation formatting (e.g., graphics)

6.1.2 Scope and Content

This chapter introduces some theoretical ideas behind GPS data processing, leading to discussions on how this theory relates to applications. It is certainly not intended to review specific software, but rather to point to concepts underlying the software.

Obviously, it would be beyond the scope of this text to go into each of the above items in detail. Observation models have already been covered in depth in previous chapters. I have therefore chosen to focus on three topics that generally lie within the areas of data preprocessing, parameter estimation, and solution processing.

I'll start with a very practical equivalence, the *equivalence of pseudorange and carrier phase*, which can be used to develop data processing algorithms. Then I explain what I mean by the *equivalence of the stochastic and functional model*, and show how this leads to different (but equivalent) methods of estimating parameters.

Finally, I discuss *frame invariance and estimability* to (1) introduce geometry from a relativistic perspective, and (2) help the reader to distinguish between what can and what cannot be inferred from GPS data. In each case, I begin with a theoretical development of the concept, followed by a discussion, and then the ideas are used for a variety of applications.

6.2 EQUIVALENCE OF PSEUDORANGE AND CARRIER PHASE

To enhance intuition on the development of data processing algorithms, it can be useful to forget that carrier phase has anything to do with cycles, and instead think of it as a precise pseudorange with an additional bias. If we multiply the carrier phases from a RINEX file, which are in units of cycles, by their nominal wavelengths, the result is a set of data in distance units ($\Phi_1 \equiv \lambda_1 \varphi_1$ and $\Phi_2 \equiv \lambda_2 \varphi_2$). The advantage of expressing both pseudorange and carrier phase observables in the same units is that the symmetry in the observation equations is emphasised, thus assisting in our ability to visualise possible solutions to problems.

6.2.1 Theoretical Development

We start with an equation that will serve as a useful working model of the GPS observables, which can be manipulated to develop suitable data processing algorithms. In chapter 5, we see carrier phase developed in units of distance. Simplifying equations (5.23) and (5.32), the dual-frequency carrier phase and pseudorange data can be expressed in a concise and elegant form (where we purposely space the equations to emphasise the symmetry):

$$\begin{aligned} \Phi_1 &= \rho & - I + \lambda_1 N_1 + \delta m_1 \\ \Phi_2 &= \rho - (f_1/f_2)^2 I + \lambda_2 N_2 + \delta m_2 \\ P_1 &= \rho & + I & + dm_1 \\ P_2 &= \rho + (f_1/f_2)^2 I & & + dm_2 \end{aligned} \qquad (6.1a)$$

The reader should beware that all of parameters in this equation are generally biased, so should not be interpreted literally except in a few special cases which will be discussed. The term ρ is the satellite-receiver range; but it is biased by clock errors, S/A, and tropospheric delay. It is often called the *non-dispersive delay* as it is identical for all four data types. The term I is the ionospheric group delay at the L1 frequency, which has the opposite sign as phase delay. It is a biased parameter, as the L1 and L2 signals are transmitted at slightly different times for different satellites. The terms N_1 and N_2 are the ambiguity parameters which, it should be remembered, are biased by initialisation constants, and are therefore generally not integers; however they can change by integer amounts due to cycle slips. We call $\lambda_1 N_1$ and $\lambda_2 N_2$ the *carrier phase biases* (which have distance units). Finally, the last column of parameters are multipath terms, where it has been assumed that most of the error is due to multipath rather than receiver noise.

There are a few terms missing from equation (6.1a) which will be referred to below in a discussion on systematic errors. These errors will negligibly affect most algorithms developed from this equation, however, any limitations should be kept in mind.

Equation (6.1a) can be conveniently arranged into matrix form. Since this is really the same equation but in a matrix form, we denote it as equation (6.1b):

$$\begin{bmatrix} \Phi_1 \\ \Phi_2 \\ P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 & 0 \\ 1 & -(f_1/f_2)^2 & 0 & 1 \\ 1 & +1 & 0 & 0 \\ 1 & +(f_1/f_2)^2 & 0 & 0 \end{bmatrix} \begin{bmatrix} \rho \\ I \\ \lambda_1 N_1 \\ \lambda_2 N_2 \end{bmatrix} + \begin{bmatrix} \delta m_1 \\ \delta m_2 \\ dm_1 \\ dm_2 \end{bmatrix} \tag{6.1b}$$

We note that the above equation has been arranged so that the coefficient matrix has no units. This proves to be convenient when analysing the derived covariance matrix. It is worth commenting that, when performing numerical calculations, the coefficient for the L2 ionospheric delay should always be computed exactly using $f_1/f_2 \equiv 154/120$.

6.2.2 Discussion

Interpreting the Terms. As will now be explained, not only can we apply equation (6.1) to raw, undifferenced observation data, but also to single and double difference data, and to observation residuals. Depending on the application, the terms have different interpretations. In some cases, a particular term might have very predictable behaviour; in others, it might be very unpredictable, and require stochastic estimation.

For example, in the case of the double difference observation equations, the ambiguity parameters N_1 and N_2 are not biased, but are truely integers. Moreover, the ionosphere parameter I is truely an unbiased (but differential) ionospheric parameter. For short enough distances and depending on various factors that affect the ionosphere, it might be adequate to ignore I when using double differences. Chapter 13 goes in this in more detail.

Equation (6.1) might also be interpreted as a residual equation, where a model for the observations have been subtracted from the left hand side. In this case, the parameter terms are to be interpreted as residual offsets to nominal values. For example, if the equation is applied to double difference residuals, and if the differential tropospheric delay can be adequately modelled, then the range term ρ can be interpreted as a double difference range residual due to errors in the nominal station coordinates.

All parameters generally vary from one epoch to another, often unpredictably. Whether using undifferenced, single differenced, or double differenced data, any cycle slip or loss of lock that occurs will induce a change in the value of the ambiguity parameters, by exactly an integer. For undifferenced data, the range term ρ is typically extremely unpredictable due to the combined effects of S/A and receiver clock variation. The ionospheric delay I can often be predicted several minutes ahead using polynomials, but it can also exhibit wild fluctuations. Double

differencing will lead to smooth, predictable behaviour of ρ (for static surveying).

It is typical for carrier phase multipath, δm_1 and δm_2, to be at the level of a few millimetres, sometimes deviating as high as a few cm, level; whereas the level of pseudorange multipath dm_1 and dm_2 is generally greater by two orders of magnitude (decimetres to metres). It is extremely difficult to produce a functional model for multipath from first principles, and it is more typical to model it either empirically (from its daily repeating signature), or stochastically (which in its simplest form amounts to adjusting the data weight from stations with high multipath).

Using Equation (6.1). Can we use equation (6.1) to form a least-squares solution for the unknown parameters? Even if we interpret the multipath terms as residuals to be minimised, we would have 4 parameters at each epoch, and only 4 observational data. We can therefore construct an exact solution for each epoch, if we ignore the multipath terms. (Once again, we caution that any numerical computations should use exact values for $f_1/f_2 \equiv 154/120$).

$$
\begin{aligned}
\begin{bmatrix} \rho \\ I \\ \lambda_1 N_1 \\ \lambda_2 N_2 \end{bmatrix} &= \begin{bmatrix} 1 & -1 & 1 & 0 \\ 1 & -(f_1/f_2)^2 & 0 & 1 \\ 1 & +1 & 0 & 0 \\ 1 & +(f_1/f_2)^2 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ P_1 \\ P_2 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & +f_1^2/(f_1^2-f_2^2) & -f_2^2/(f_1^2-f_2^2) \\ 0 & 0 & -f_2^2/(f_1^2-f_2^2) & +f_2^2/(f_1^2-f_2^2) \\ 1 & 0 & -(f_1^2+f_2^2)/(f_1^2-f_2^2) & +2f_2^2/(f_1^2-f_2^2) \\ 0 & 1 & -2f_1^2/(f_1^2-f_2^2) & +(f_1^2+f_2^2)/(f_1^2-f_2^2) \end{bmatrix} \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ P_1 \\ P_2 \end{bmatrix} \\
&\cong \begin{bmatrix} 0 & 0 & +2.546 & -1.546 \\ 0 & 0 & -1.546 & +1.546 \\ 1 & 0 & -4.091 & +3.091 \\ 0 & 1 & -5.091 & +4.091 \end{bmatrix} \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ P_1 \\ P_2 \end{bmatrix}
\end{aligned}
\qquad (6.2)
$$

Note that the carrier phase biases are constant until lock is lost on the satellite, or until a cycle slip occurs. We can therefore use these equations to construct algorithms that (1) resolve ambiguities, and (2) detect and solve for cycle slips. The second point to notice, is that between cycle slips, we know that the ambiguities are constant. If we are interested in only the variation in the other parameters (rather than the absolute values), then we are free to ignore any constant terms due to the ambiguity parameters.

We can rearrange equation (6.1) to reflect this idea, by attaching the ambiguity parameters to the carrier phase observations. Of course, we might not know these parameters perfectly, but that will have no effect on the estimated *variation* in the other parameters. Furthermore, we can explicitly introduce the pseudorange multipath terms into the parameter vector:

$$\begin{bmatrix} \tilde{\Phi}_1 \\ \tilde{\Phi}_2 \\ P_1 \\ P_2 \end{bmatrix} \equiv \begin{bmatrix} \Phi_1 - \lambda_1 N_1 \\ \Phi_2 - \lambda_2 N_2 \\ P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & -(f_1/f_2)^2 & 0 & 0 \\ 1 & +1 & 1 & 0 \\ 1 & +(f_1/f_2)^2 & 0 & 1 \end{bmatrix} \begin{bmatrix} \rho \\ I \\ dm_1 \\ dm_2 \end{bmatrix} + \begin{bmatrix} \delta m_1 \\ \delta m_2 \\ e_1 \\ e_2 \end{bmatrix} \tag{6.3}$$

where we have explicitly included pseudorange measurement noises e_1 and e_2. As we did for equation (6.1), equation (6.3) can be inverted:

$$\begin{bmatrix} \rho \\ I \\ dm_1 \\ dm_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & -(f_1/f_2)^2 & 0 & 0 \\ 1 & +1 & 1 & 0 \\ 1 & +(f_1/f_2)^2 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\Phi}_1 \\ \tilde{\Phi}_2 \\ P_1 \\ P_2 \end{bmatrix}$$

$$= \begin{bmatrix} +f_1^2/(f_1^2 - f_2^2) & -f_2^2/(f_1^2 - f_2^2) & 0 & 0 \\ +f_2^2/(f_1^2 - f_2^2) & -f_2^2/(f_1^2 - f_2^2) & 0 & 0 \\ -(f_1^2 + f_2^2)/(f_1^2 - f_2^2) & +2f_2^2/(f_1^2 - f_2^2) & 1 & 0 \\ -2f_1^2/(f_1^2 - f_2^2) & +(f_1^2 + f_2^2)/(f_1^2 - f_2^2) & 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{\Phi}_1 \\ \tilde{\Phi}_2 \\ P_1 \\ P_2 \end{bmatrix} \tag{6.4}$$

$$\cong \begin{bmatrix} +2.546 & -1.546 & 0 & 0 \\ +1.546 & -1.546 & 0 & 0 \\ -4.091 & +3.091 & 1 & 0 \\ -5.091 & +4.091 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{\Phi}_1 \\ \tilde{\Phi}_2 \\ P_1 \\ P_2 \end{bmatrix}$$

Notice the striking similarity in equations (6.2) and (6.4), and reversal of roles between carrier phase and pseudorange. One can see the familiar *ionosphere-free linear combination* of data as solutions for the range term; whereas in equation (6.2) it applies to pseudorange, in equation (6.4) it applies to carrier phase. Similarly, the ionospheric delay term is equal the familiar *ionospheric* or *geometry-free linear combination* of pseudorange in equation (6.2), and of carrier phase in equation (6.4).

The coefficients for the ambiguity and multipath estimates are symmetrical between equations (6.2) and (6.4). We can interpret this as follows. In equation (6.2), the pseudorange is being effectively used as a model for the carrier phase due in order to infer the carrier phase bias parameters. On the other hand, in equation (6.4) the carrier phase is effectively being used to model time variations in the pseudorange in order to infer pseudorange multipath variations. The symmetry of the coefficients in the two equations is therefore not surprising given this explanation.

Statistical Errors. Since the level of errors are strikingly higher for pseudorange as compared with carrier phase, we should look at the propagation of errors into the

above parameters. The method used here is similar to the familiar computation of *dilution of precision* for point positioning. The covariance matrix for the parameter estimates given by equation (6.2) can be computed by the usual procedure as follows:

$$\mathbf{C} = \left(\mathbf{A}^{\mathbf{T}} \mathbf{C}_{\text{data}}^{-1} \mathbf{A} \right)^{-1} \tag{6.5}$$

where $\mathbf{A}$ is the coefficient matrix in equations (6.1) or (6.3), and $\mathbf{C}_{\text{data}} = \mathbf{W}^{-1}$ is the data covariance matrix. If we assume that the data covariance is diagonal, that there is no difference between the level of errors on L1 and L2, and that the variance for carrier phase is negligible compared to the pseudorange then we write the data covariance:

$$\mathbf{C}_{\text{data}} = \lim_{\varepsilon \to 0} \begin{pmatrix} \varepsilon\sigma^2 & 0 & 0 & 0 \\ 0 & \varepsilon\sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix} \tag{6.6}$$

Recall that in a real situation, a typical value might be $\varepsilon \approx 10^{-4}$, which justifies our simplification by taking the limit $\varepsilon \to 0$. Applying equations (6.5) and (6.6) to equation (6.2), and substituting values for the frequencies, we find the parameter covariance:

$$\mathbf{C}_{(6.2)} = \sigma^2 \begin{pmatrix} 8.870 & -6.324 & -15.194 & -19.286 \\ -6.324 & 4.779 & 11.103 & 14.194 \\ -15.194 & 11.103 & 26.297 & 33.480 \\ -19.286 & 14.194 & 33.480 & 42.663 \end{pmatrix} \tag{6.7}$$

The formal standard deviations for the parameters are the square root of the diagonal elements:

Parameter	Standard Deviation
ρ	2.978σ
I	2.186σ
N_1	$5.128\sigma/\lambda_1$
N_2	$6.532\sigma/\lambda_2$

Table 6.1: Formal errors of parameter derived at a single epoch using dual frequency code and carrier phase data

The formal error for the ionosphere-free range term is approximately 3 times the level of the measurement errors. This result also applies to the estimates of range variation in equation (6.4) which uses the ionosphere-free carrier phase data. It illustrates the problem for short baselines, where there is a trade-off between raising the effective measurement error, versus reducing systematic error from the ionosphere (see chapter 13). The formal error for the ionospheric delay (at the L1 frequency) is

approximately 2 times the level of the measurement errors, which shows that the L1 and L2 signals are sufficiently well separated in frequency to resolve ionospheric delay. The large scaling factors of 5.128 to 6.532 for the carrier phase ambiguities shows that pseudorange multipath must be adequately controlled if there is any hope to resolve ambiguities (or detect cycle slips) using pseudorange data. For example, if we aim for an N_1 standard deviation of 0.25 cycles, then the pseudorange precision must be approximately 5 times smaller than this, which is less than 1 cm!

Systematic Errors. At this point, it is worth recalling that we have not used any functional model for the range term or the ionospheric term, other than that they satisfy the following assumptions:

- The range term (which includes range, tropospheric delay, and clock offsets) are identical for all observables.
- Ionospheric delay varies as the inverse square of the frequency, with the phase delay having the same magnitude but opposite sign to the group delay

Equations (6.2) and (6.4) tells us that we can form an estimators for the carrier phase ambiguities and pseudorange multipath variation, even in the extreme situation when we have no functional model for range, tropospheric delay, clocks, and ionospheric delay (other than the above simple assumptions). For example, no assumptions have been made concerning motion of the GPS antenna, and we can therefore derive algorithms to fix cycle slips and resolve carrier phase ambiguities that are suitable for kinematic applications. Similarly, pseudorange multipath can be assessed as an antenna is moved through the environment.

In the next section, we derive algorithms that can be considered application independent. This is only strictly true as far as the above assumptions are valid. For completeness, we list here reasons why the above assumptions might not be valid:

- The carrier signal is circularly polarised, and hence the model should really include the *phase wind up effect* caused by relative rotation between the GPS satellite's antenna and the receiver's antenna [*Wu et al.*, 1993]. This is particularly important for moving antennas with data editing algorithms operating on undifferenced data. It has also happened in the past that one of the GPS satellites began to spin due to some malfunction, thus causing a dramatic phase wind up effect in the carrier phase, which was of course not observed in the pseudorange. One way around such problems is to use the widelane phase combination, which is rotationally invariant $\varphi_W \equiv (\varphi_1 - \varphi_2) = (\Phi_1/\lambda_1 - \Phi_2/\lambda_2)$. The phase wind up effect can be almost eliminated by double differencing, or by using an iterative procedure to account for antenna orientation (which can often be modelled adequately using a preliminary solution for the direction of motion). An interesting twist on this is to try to use the observed phase wind up to help with models of vehicle motion. For this purpose, the single differenced ionospheric phase could be used between a nearby reference antenna, and an antenna on a moving vehicle. Over short distances, the ionospheric delay would almost cancel, leaving a clear signature due to antenna rotation.
- The model should really include antenna phase centre offsets and antenna phase centre variation. We are free to define any such errors under the umbrella term *multipath*, but it is advisable to correct the data for such effects. Double differencing over short baselines almost eliminates phase centre effects, provided

the same antenna types and method of mounting are used.

- There is generally an slight difference in the time of transmission for the L1 and L2 signals, which is different for each satellite (typically a few metres). Moreover, the effective time of reception might be slightly different in receivers which make no attempt to self-calibrate this interchannel delay. Certainly, for precise geodetic applications, such a bias is irrelevant, as it would either cancel in double differencing, or be harmlessly absorbed as a clock parameter; however, for ionospheric applications, these biases must be modelled.
- There might be a slight variable bias in the receiver between the different observable types due to any internal oscillators and electronics specific to the L1 and L2 frequencies. Hence the assumption that the range term is the same for all observables becomes invalid. Receivers are not supposed to do this, but hardware problems have been known to cause this effect, which is often temperature dependent. The effect is mitigated by double differencing, but can be problematic for undifferenced processing, for data processing algorithms, and for ionospheric estimation software.
- There might be slight variable biases due to systematic error in the receiver, such as tracking loop errors that are correlated with the Doppler shift. For well designed, geodetic-class receivers, these biases should be down at the millimetre level.

6.2.3 Applications

Multipath Observables. This is the simplest and most obvious application from the previous discussion. Equation (6.4) shows how pseudorange multipath can be estimated epoch by epoch. It relies on the assumption that the carrier phase biases are constant. If not, then the data should first be edited to correct for any cycle slips. It should also be remembered that such multipath estimates are biased; therefore, only multipath variation and not the absolute multipath can be inferred by this method. This method is particularly useful for assessing the quality of the environment at a GPS station. This might be used for site selection, for example. Another application is to look at the multipath statistics. These could then be used to compute pseudorange data weights in least squares estimation, or for other algorithms that use the pseudorange.

Data Editing. Data editing includes the process of outlier detection, cycle slip detection, and cycle slip correction. Equations (6.1-6.4) point to a possible method for data editing, as it shows that the parameters are correlated, and therefore perhaps at each epoch, the set of four observations can be assessed for self-consistency. But outlier detection requires data redundancy, which we do not have for individual epochs.

However, we can monitor the solution for the parameters, equation (6.2), and ask whether they are behaving as expected. This line of thought leads naturally to a sophisticated approach involving a Kalman filter to predict the solution at the next epoch, and then compare this prediction with the new data. If the data rate is sufficiently high that prediction becomes meaningful, then this approach might be useful.

However, experience by the author and those developing the GIPSY software at the

Jet Propulsion Laboratory showed this approach to be problematic, at least for undifferenced data. The presence of selective availability, the possible occurrence of high variability in ionospheric total electron content, and the poor frequency stability of receiver's internal oscillators limit the usefulness of Kalman filtering for typical geodetic data taken at a rate of 1 per 30 seconds. Even going to higher rates does not significantly improve the situation if the receiver clock is unpredictable. Moreover, results were difficult to reproduce if the analyst were allowed to tune the filter.

As a result, a simpler, fully automatic algorithm was developed known as TurboEdit [*Blewitt*, 1990], which uses the positive aspects of filtering (i.e., noise reduction through averaging, and using prediction as a means of testing new data). The new algorithm attempted to minimise sensitivity to unusual, but acceptable circumstances, by automatically adapting its procedures to the detected level of noise in the data. The specific TurboEdit algorithm will not be described in detail here, but rather we shall next focus on some principles upon which data editing algorithms can be founded.

Firstly, we shall look at the use of pseudorange to help detect and correct for cycle slips. (In this context, by cycle slip we mean a discontinuity in the integer ambiguity parameter, which can be caused by the receiver incorrectly counting the cycles, or if the receiver loses lock on the signal). From Table 1, we see that a pseudorange error of 1 cm would result in an instantaneous estimate for the carrier phase biases at the level of 5 to 6 cm, which corresponds to approximately one quarter of a wavelength. Therefore, it would seem that pseudorange multipath would have to be controlled at this level if we were to simply use these estimates to infer whether the integer ambiguity had changed by one cycle. This seems like a dire situation.

This situation can be improved, if we realise three useful observations of experience: (1) Most often, cycle slips are actually caused by loss of lock, in which case the slip is much greater than one cycle; therefore, detection is simpler than correction. (2) Unless we have conditions so adverse that any result would be questionable, most often we have many epochs of data until we reach a cycle slip; therefore, we can average the results from these epochs to estimate the current value of the carrier phase bias. (3) If we have an initial algorithm that flags epochs where it suspects a cycle slip may have occurred, we can seperately average the carrier phase bias solutions either side of the suspected cycle slip, and test the hypothesis that the carrier phase bias has changed by at least an integer number of wavelengths.

Following this logic, we can estimate how many epochs of data are required either side of the cycle slip so that the pseudorange errors can be averaged down sufficiently for us to test the hypothesis that a cycle slip has ocurred. Using the results in Table 1, and assuming the errors average down as the square root of the number of epochs, we can, for example, write the error in the estimated cycle slip on L1 as:

$$\sigma_{\text{slip}}(n) = \sqrt{2}\,\frac{5.128\sigma}{\sqrt{n}} \tag{6.8}$$

where n is the number of epochs either side of the hypothesised cycle slip. For example, if we insist that this computed error be less than one quarter of a cycle, i.e., approximately 5 cm, and if we assume that the pseudorange error σ is approximately 50 cm, then we see that the number of epochs must be greater than approximately 5000. This is clearly an unrealistic approach as it stands.

We can use the above approach, if instead we use the well known widelane carrier phase ambiguity. From equation (6.7) and using the law of propagation of errors, we can compute the formal variance in the widelane ambiguity, $N_{\mathrm{W}} \equiv N_1 - N_2$

$$\begin{aligned}
\mathbf{C}_{\mathrm{W}} &= \sigma^2 \begin{pmatrix} 1/\lambda_1 & -1/\lambda_2 \end{pmatrix} \begin{pmatrix} 26.297 & 33.480 \\ 33.480 & 42.663 \end{pmatrix} \begin{pmatrix} 1/\lambda_1 \\ -1/\lambda_2 \end{pmatrix} \\
&= \sigma^2 \left(26.297/\lambda_1{}^2 - 2 \times 33.480/\lambda_1\lambda_2 + 42.663/\lambda_2{}^2 \right) \\
&= \left(\sigma/\lambda_1\right)^2 \left(26.297 - 66.960 \times (120/154) + 42.663 \times (120/154)^2 \right) \\
&= \left(0.15720\sigma/\lambda_1 \right)^2
\end{aligned} \tag{6.9}$$

This derivation uses the exact relation $(\lambda_1/\lambda_2) = (f_2/f_1) = (120/154)$. (As a rule for numerical stability, it is always wise to substitute explicitly for the L1 carrier wavelength only at the very last step).

Remarkably, the standard error in the widelane wavelength does not reach 1 cycle until the pseudorange errors approach $\sigma = \lambda_1/0.15720 = 6.3613\lambda_1 \approx 120\mathrm{cm}$. We can therefore use such widelane estimates on an epoch by epoch basis as an algorithm to flag possible cycle slips. The hypothesis can then be tested by averaging down the pseudorange noise either side of the proposed slip, as discussed previously.

Widelaning data editing methods are generally very successful for modern geodetic receivers, which have well behaved pseudorange. However, they do not distinguish as to whether the slip occurred on L1, L2, or both.

This problem can be resolved by looking at either the biased range parameter ρ or biased ionospheric parameter I in equation (6.4). Note that the carrier phase ambiguity parameters appear to the right side of this equation. Were there to be a cycle slip, it would manifest itself as a discontinuity in both parameters I and ρ. Using this method requires that either one of these parameters be predictable, to effectively bridge the time period during which the receiver lost lock on the signal. For double differenced data, this should be rather straightforward for both parameters, particularly for short baselines. For undifferenced data, parameter ρ tends to be too unpredictable due to S/A and receiver clock variation; however, I is usually sufficiently well behaved that time periods of up to several minutes can be bridged. A low order polynomial predictor could be used for this purpose.

Data editing algorithms can be designed to be adaptive to the changing quality of the data and the predictability of the parameters. The level of pseudorange noise can be easily monitored, as discussed, using equation (6.4) to estimate the multipath terms (taking care to correct for cycle slips detected so far).

The predictability of the parameters can be tested by applying the prediction algorithm backwards in time to *previous* data which are known to be clean or corrected for cycle slips. For example, if we have a loss of lock and subsequent data outage of 5 minutes, we might want to test a simple algorithm which predicts the I parameter using a second order polynomial on 15 minutes of data prior to the loss of lock. The test could be conducted by extrapolating the polynomial backwards, and comparing it with existing data.

Data Thinning. For static GPS where data are collected over a period of several hours, a carrier phase data rate of 1 epoch every 5 minutes should be more than sufficient to achieve high precision results. In fact, using higher rate data is unlikely to improve the result significantly. The reason for this is that if we continue to increase the data rate, we may well be able to reduce the contribution of measurement error to errors in the parameter estimates; however, we will do little to reduce the effect of systematic error, for example, low frequency components of multipath.

Therefore, if we are presented with a file with carrier phase data every 30 seconds, a simple and effective way to speed up the processing is to decimate the data, only accepting one point every 5 minutes.

For pseudorange data, however, a higher data rate often leads to improved results, presumably because measurement error and high frequency components of multipath continue to be significant error sources. A better approach than decimation would be to interpolate the high rate pseudorange data to every 5 minute data epoch, because the interpolation process would help average down the high frequency noise. For subsequent least squares analysis to be valid, the interpolator should strictly only independent 5 minute segments of high rate data, so that no artificial correlations are introduced (which could, for example, confound other algorithms in your software).

A convenient method of interpolation is to use the carrier phase as a model for the pseudorange. The multipath expressions in Equation (6.4) provides us the solution to this problem. For example, we can rearrange (6.4) to express a model of the pseudorange data in terms of the carrier phase data and the pseudorange multipath:

$$\begin{aligned} P_1 &= 4.091\tilde{\Phi}_1 - 3.091\tilde{\Phi}_2 + dm_1 \\ &= 4.091\Phi_1 - 3.091\Phi_2 + dm_1 + \mathrm{B} \end{aligned} \tag{6.10}$$

The carrier phase here is effectively being used to mimic the time variation in the pseudoranges, correctly accounting for variation in range and ionospheric delay. The constant B is due to the (unknown) carrier phase biases. We can procede to express an estimator for P_1 as the expected value:

$$\begin{aligned} \hat{P}_1 &\equiv E\left(P_1\right) \\ &= E\left(4.091\Phi_1 - 3.091\Phi_2 + dm_1 + B\right) \\ &= 4.091\Phi_1 - 3.091\Phi_2 + E\left(B\right) \end{aligned} \tag{6.11}$$

where, for the carrier phase data, the expected values are simply the actual data recorded at the desired 5-minute epoch, and we have assumed that the expected value for multipath is zero. If we wish our resulting estimate $\hat{P}_1$ to be truely independent from one 5 minute epoch to the next, then $E(B)$ can only be based on data found in a 5 minute window surrounding this epoch, giving us the following expression:

$$\hat{P}_1 = 4.091\Phi_1 - 3.091\Phi_2 + \left\langle P_1 - 4.091\Phi_1 + 3.091\Phi_2 \right\rangle \tag{6.12}$$

(where the angled brackets denote the time average operator). The result is a smoothed estimate for the pseudoranges. Remember, only the smoothed pseudorange

that falls on the 5 minute epoch should to be saved.

Parameter Estimation. It is common in geodetic software to first use the pseudorange to produce a receiver clock solution, and then use the double differenced carrier phase to produce a precise geodetic solution. The reason we need to know the receiver clock time precisely is to determine the time the data were collected, and hence fix the geometry of the model at that time. Once this has been done, the clock parameters are then effectively removed from the problem by double differencing. The disadvantage to this scheme, is that we might be interested in producing a high precision clock solution.

One way of approaching this is to estimate clock parameters explicitly along with the geodetic parameters, using undifferenced carrier phase data. The problem with this is that there is an extremely high correlation with the (undifferenced) carrier phase bias parameters.

An alternative, elegant one-step procedure is to process undifferenced carrier phase and pseudorange data simultaneously, effectively using the pseudorange to break the correlation. For example, suppose we process dual frequency ionosphere-free carrier phase and pseudorange together. The models for both types of observables are identicle, apart from the carrier phase bias (which can, in any case, assume a nominal value of zero), and apart from the lack of a phase wind-up effect in the pseudorange data. Similarly, the parameters we estimate would be identical; that is, no extra parameters are required. Apart from expanding the possible applications (where clocks are involved), this method provides an extra degree of reliability, especially for kinematic applications, where the pseudorange effectively provides a consistency check on the carrier phase solution.

One very interesting new application of this idea is called *precise point positioning*. Developed by researchers at JPL [*Zumberge et al.*, 1996], this technique is identical to conventional pseudorange point positioning, except that (1) both pseudorange and carrier phase data are processed simultaneously, and (2) precise satellite ephemerides are used. Precise point positioning allows a single receiver to be positioned with 1 cm accuracy in the global frame (ITRF). We return to this exciting new tool later in this chapter.

Ambiguity Resolution. The application of equation (6.2) to ambiguity resolution is basically very similar to the application to data editing and the correction of cycle slips. It must be remembered, however, that only the double differenced carrier phase biases are an integer number of wavelengths. Therefore, equation (6.2) should be interpreted as for double differenced data and parameters. Alternatively, undifferenced parameters can be estimated, and subsequently the estimates can be double differenced.

As discussed for cycle slip correction, the pseudorange multipath too large for reliable ambiguity resolution using equation (6.2) directly. On the other hand, the widelane carrier phase ambiguity $N_{\mathrm{W}} \equiv N_1 - N_2$ can be fixed very reliably using pseudoranges, even of mediocre quality. The advantage to this method is that it is independent of baseline length.

As with correcting cycle slips, we need to address how we resolve the ambiguities for N_1 and N_2 separately. Assuming we know the widelane, the problem reduces to

finding the correct value for N_1. Once again, one possible answer lies in the solutions for the the (double differenced) ionospheric term *I*. Using equation (6.4), and assuming we have a very good model for *I*, we can find the best fitting values of the ambiguities for $\tilde{\Phi}_1 \equiv \Phi_1 - \lambda_1 N_1$ and $\tilde{\Phi}_2 \equiv \Phi_2 - \lambda_2 N_2$, subject to the constraint $N_1 - N_2 = \hat{N}_W$, where $\hat{N}_W$ is the widelane ambiguity, previously resolved using equation (6.2)

$$\begin{aligned} I &= 1.546\tilde{\Phi}_1 - 1.546\tilde{\Phi}_2 \\ I/1.546 &= \left(\Phi_1 - \lambda_1 N_1\right) - \left(\Phi_2 - \lambda_2 N_2\right) \\ &= \Phi_1 - \Phi_2 - \lambda_1 N_1 + \lambda_2 N_2 \\ &= \Phi_1 - \Phi_2 - \lambda_2 \hat{N}_w + \left(\lambda_2 - \lambda_1\right) N_1 \end{aligned} \tag{6.13}$$

This situation is relatively easy over baselines of a few km, where it can be assumed that, to a good approximation, *I* = 0. However, the coefficient $\left(\lambda_2 - \lambda_1\right) \approx 5.4\text{cm}$ is very small, so we can easily run into problems over 5 km during daylight hours, and over 30 km at night. However, it is an almost an instantaneous technique, and was used successfully for rapid static surveying of post-seismic motion following the Loma Prieta earthquake of 1989 [*Blewitt et al.*, 1990].

Over longer baselines, *Melbourne* [1985] suggested an approach that uses the ionosphere-free phase combination of equation (6.4) and a good model for the range term. Later experience has shown that the range model must be based on a preliminary bias-free solution (since our a priori knowledge of the troposphere is generally inadequate). From equation (6.4), we can find the value of N_1 that best fits the range model, subject to the usual widelane constraint:

$$\begin{aligned} \rho &= 2.546\tilde{\Phi}_1 - 1.546\tilde{\Phi}_2 \\ &= 2.546\left(\Phi_1 - \lambda_1 N_1\right) - 1.546\left(\Phi_2 - \lambda_2 N_2\right) \\ &= 2.546\Phi_1 - 1.546\Phi_2 - 2.546\lambda_1 N_1 + 1.546\lambda_2 N_2 \\ &= 2.546\Phi_1 - 1.546\Phi_2 - 1.546\lambda_2 \hat{N}_w + \left(1.546\lambda_2 - 2.546\lambda_1\right) N_1 \end{aligned} \tag{6.14}$$

The coefficient $\left(1.546\lambda_2 - 2.546\lambda_1\right) \approx 10.7\text{cm}$ shows that this method will work provided we can control our estimated (double differenced) range errors to within a few centimetres. Using precise orbit determination and stochastic tropospheric estimation, this method has proved successful over thousands of km [*Blewitt*, 1989], and even over global scales [*Blewitt and Lichten*, 1992].

6.3 EQUIVALENCE OF STOCHASTIC AND FUNCTIONAL MODELS

We are familiar with standard least squares theory, where the observations have both a functional model, which tells us how to compute the observation, and a stochastic model, which tells us the expected statistics of the errors. If we decide to

augment the functional model with extra parameters, an equivalent result can be obtained if instead we modify the stochastic model. As we shall see, this equivalence introduces great flexibility into estimation algorithms, with a wide variety of geodetic applications..

6.3.1 Theoretical Development

Terminology. Consider the linearised observation equations:

$$z = Ax + v \tag{6.15}$$

where z is the column vector of observed minus computed observations, A is the design matrix, x is the column vector of corrections to functional model parameters, and v is a column vector of errors. Let us assume the stochastic model

$$\begin{aligned} E(v) &= 0 \\ E(vv^T) &= C \equiv W^{-1} \end{aligned} \tag{6.16}$$

Assuming a well conditioned problem, the *best linear unbiased estimator* of x is:

$$\hat{x} = \left(A^T W A\right)^{-1} A^T W z \tag{6.17}$$

which has the following statistical properties:

$$\begin{aligned} E(\hat{x}) &= E(x) = x \\ E(\hat{x}\hat{x}^T) &= \left(A^T W A\right)^{-1} \equiv C_{\hat{x}} \end{aligned} \tag{6.18}$$

If we use a Bayesian approach to estimation, we may make the a priori assumption $E(x) = 0$ where we implicitly introduce pseudo-data $x = 0$ with an a priori covariance C_0. In this case, the estimator becomes:

$$\hat{x} = \left(A^T W A + C_0^{-1}\right)^{-1} A^T W z \tag{6.17b}$$

We see that (6.17b) approaches (6.17) in the limit $C_0 \to \infty$, hence we can consider (6.17) the special case of (6.17b) where we have no a priori information.

Augmented Functional Model. Suppose we aim to improve our solution by estimating corrections to an extra set of functional model parameters y. We therefore consider the augmented observation equations:

$$z = Ax + By + v \tag{6.19}$$

We can write this in terms of partitioned matrices:

$$z = \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + v \tag{6.20}$$

We can therefore see by analogy with (6.17) that the solution for the augmented set of parameters will be

$$\begin{aligned} \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} &= \left(\begin{pmatrix} A^T \\ B^T \end{pmatrix} W \begin{pmatrix} A & B \end{pmatrix} \right)^{-1} \begin{pmatrix} A^T \\ B^T \end{pmatrix} W z \\ &= \begin{pmatrix} A^T W A & A^T W B \\ B^T W A & B^T W B \end{pmatrix}^{-1} \begin{pmatrix} A^T W z \\ B^T W z \end{pmatrix} \end{aligned} \tag{6.21}$$

We now use the following lemma on matrix inversion for symmetric matrices, which can easily be verified:

$$\begin{pmatrix} \Lambda_1 & \Lambda_{12} \\ \Lambda_{21} & \Lambda_2 \end{pmatrix}^{-1} = \begin{pmatrix} \left(\Lambda_1 - \Lambda_{12} \Lambda_2^{-1} \Lambda_{21} \right)^{-1} & \left(\Lambda_{12} \Lambda_2^{-1} \Lambda_{21} - \Lambda_1 \right)^{-1} \Lambda_{12} \Lambda_2^{-1} \\ \left(\Lambda_{21} \Lambda_1^{-1} \Lambda_{12} - \Lambda_2 \right)^{-1} \Lambda_{21} \Lambda_1^{-1} & \left(\Lambda_2 - \Lambda_{21} \Lambda_1^{-1} \Lambda_{12} \right)^{-1} \end{pmatrix} \tag{6.22}$$

Applying this lemma, we can derive the following elegant result for the estimates of *x*, the parameters of interest (defining the projection operator *P*):

$$\hat{x} = \left(A^T W P A \right)^{-1} A^T W P z \quad \text{where} \quad P \equiv I - B \left(B^T W B \right)^{-1} B^T W \tag{6.23}$$

That is, we have derived a method of estimating parameters *x*, without having to go to the trouble of estimating *y*. The result (6.23) is remarkable, in that it has exactly the same form as equation (6.17), where we substitute the original weight matrix for *the reduced weight matrix*:

$$\begin{aligned} W' &\equiv WP \\ &= W - WB \left(B^T W B \right)^{-1} B^T W \end{aligned} \tag{6.24}$$

If we are in fact interested in obtaining estimates for *y* at each batch, we can backsubstitute $\hat{x}$ into (6.21) (for each batch) to obtain:

$$\hat{y} = \left(B^T W B \right)^{-1} B^T W \left(z - A\hat{x} \right) \tag{6.25}$$

Augmented Stochastic Model. We need to find a stochastic model that gives rise to the reduced weight matrix (6.24). A stochastic model is correctly stated in terms of the expectation values (6.16), but unfortunately, the reduced weight matrix is singular (because *P* is an idempotent matrix: *PP*=*P*). However, an interesting interpretation arises if we derive the stochastic model from first principles. If we treat the augmented part of the model as a source of noise (called *process noise*) rather than as

part of the functional model, we can write the *augmented stochastic model* as follows:

$$\begin{aligned} C' &= E\left(v'v'^T\right) \\ &= E\left((By+v)(By+v)^T\right) \\ &= E\left(vv^T\right) + BE\left(yy^T\right)B^T \\ &= C + BC_yB^T \end{aligned} \tag{6.26}$$

Where C_y is, by definition, an *a priori covariance matrix* for the parameter, *y*. Note that we can choose C_y to be arbitrarily large, if we wish the data to completely influence the result. If we now invert this expression, it ought to correspond to the reduced weight matrix of (6.24). But first, we need to know another very useful matrix inversion lemma (worth remembering!):

$$\left(\Lambda_1 \pm \Lambda_{12}\Lambda_2^{-1}\Lambda_{21}\right)^{-1} = \Lambda_1^{-1} \mp \Lambda_1^{-1}\Lambda_{12}\left(\Lambda_2 \pm \Lambda_{21}\Lambda_1^{-1}\Lambda_{12}\right)^{-1}\Lambda_{21}\Lambda_1^{-1} \tag{6.27}$$

where $\Lambda_{12} \equiv \Lambda_{21}^T$. Applying this lemma to equation (6.26), we find:

$$\begin{aligned} W' &= \left(C + BC_yB^T\right)^{-1} \\ &= C^{-1} - C^{-1}B\left(BC^{-1}B^T + C_y^{-1}\right)^{-1}B^TC^{-1} \\ &= W - WB\left(BWB^T + C_y^{-1}\right)^{-1}B^TW \end{aligned} \tag{6.28}$$

Comparing this expression with (6.24), we find the only difference is the presence of the a priori covariance matrix for parameters *y*. The functional and stochastic approach are equivalent in the limit that the a priori stochastic parameter covariance is made sufficiently large. (See the discussion following (6.17b)). For the two models to be equivalent, the augmented stochastic model should only account for correlations introduced by data's functional dependence on the process noise (as defined by the matrix *B*), with no a priori information on the actual variance of the process noise.

Stochastic Estimation. In the context of least squares analysis, a parameter in general is defined in terms of its linear relationship to the observable (i.e., through the design matrix). A *stochastic parameter* has the same property, but is allowed to vary in a way that can be specified statistically. In computational terms, a *constant parameter* is estimated as a constant over the entire data span, whereas a stochastic parameter is estimated as a constant over a specified batch interval, and is allowed to vary from one interval to the next. For example, a special case of this is where the stochastic parameter is allowed to change at every data epoch.

The least squares estimator includes specific a priori information on the parameter, in terms of (1) how its value propagates in time from one batch to the next, and (2) how its variance propagates to provide an a priori constraint on the next batch's estimate. Here, we introduce two of the most important models used in stochastic

estimation for precise GPS geodesy:

(1) The simplest is the *white noise* model, which can be specified by:

$$\begin{aligned} E(y_i) &= 0 \\ E(y_i y_j) &= \sigma^2 \delta_{ij} \end{aligned} \tag{6.29}$$

(2) The random walk model can be specified by

$$\begin{aligned} E(y_i - y_j) &= 0 \\ E\left((y_i - y_j)^2\right) &= \vartheta(t_i - t_j) \end{aligned} \tag{6.30}$$

In the absence of data, the white noise parameters become zero with a constant assumed variance, whereas the random walk parameters retain the last estimated value, with a variance that increases linearly in time. The white noise model is useful where we wish to impose no preconceived ideas as to how a parameter might vary, other than (perhaps) its expected average value. As (6.30) does not require us to specify $E(y_i)$, the random walk model is particularly useful for cases where we do expect small variations in time, but we might have little idea on what to expect for the overall bias of the solution.

The white noise and random walk models are actually special cases of the first order Gauss-Markov model of process noise [*Bierman*, 1977], however, this general model is rarely used in GPS geodesy.

6.3.2 Discussion

Model Equivalence. The equivalence of (6.23) with (6.21), and (6.24) with (6.26) proves the correspondence between modifying the functional model and modifying the stochastic model. Instead of estimating extra parameters, we can instead choose to modify the stochastic model so as to produce the reduced weight matrix, or equivalently, an augmented covariance. Note that, as we would expect, the weight matrix is reduced in magnitude, which is why it is said that estimating extra parameters *weakens the data strength*. It follows that the corresponding covariance matrices for the data and for the estimated parameters will increase.

We can summarise the above theoretical development by the maxim:

(covariance augmentation) ≡ (weight matrix reduction) ≡ (parameter estimation)

That is, augmenting the stochastic model can be considered implicit estimation of additional parameters, with the advantage of that there is a saving in computation. The only disadvantage is that the full covariance matrix between all *x* and *y* parameters is not computed. Fortunately, there are many applications where the full covariance matrix is of little interest, particularly for problems that are naturally localised in space and time.

Stochastic Parameters. The above theory indicates possible ways to deal with

stochastic parameters, that are allowed to vary in time according to some stochastic model. Equations (6.23), (6.26) and (6.28) provides a simple mechanism for us to estimate a special class of stochastic parameters called *white noise parameters*, that are allowed to vary from one (specified) batch of data to the next, with no a priori correlation between batches. The a priori covariance matrix in (6.28) can be ignored if we wish, but it can be useful if we believe we know the parameter variance a priori (from some other source), and we do not wish to weaken the data strength unnecessarily.

For example, if we have some parameters which are stochastic in time, we could group the data into batches covering a set time interval, and apply equation (6.23) to estimate the x parameters at each batch interval. The final x parameter estimates could then be derived by accumulating the normal equations from every batch, and then inverting.

The formalism presented above also suggests a method for implementing random walk parameter estimation. Specifically, (6.28) allows for the introduction of an a priori covariance, which could come from the previous batch interval solution, augmented by the model (6.30). Several convenient formulisms have been developed for the step-by-step (batch sequential) approach to estimation, including algorithms such as the Kalman Filter. It is beyond the scope of this chapter to go into specific algorithms, but we shall describe filtering in general terms.

Filtering. In *filtering* algorithms, the a priori estimate for each batch is a function of the current running estimate mapped from the previous batch. The current estimate is then specified as a weighted linear combination of the a priori estimate, and the data from the current batch. The relative weights are determined by the *gain matrix*, which can also account for the a priori correlations between stochastic parameters in accordance with the user-specified stochastic model (6.29) or (6.30). The principles of separating stochastic from global parameters are the same as described earlier. The process of backsubstitution in this context is called *smoothing*, which is essentially achieved by running the filter backwards to allow earlier data to be influenced by the later data in a symmetric way.

Algorithm Equivalence. Whatever algorithm is used, we should always remember that it is the underlying stochastic and functional models that determine the solution. That is, it is possible to construct a conventional weighted least-squares estimator to produce the same answer as, say, a Kalman filter. The choice of algorithm is largely one of computational efficiency, numerical stability, and convenience in being able to control the stochastic model.

Although we have not shown it here, there is a similar equivalence between stochastic estimation and applying a transformation to remove so-called *nuisance parameters*. A simple example of this is the ionospheric linear combination of data, which removes ionospheric delay. This is equivalent to estimating ionospheric delay as *white noise* for each observation. Likewise, the double differencing transformation is equivalent to estimating white noise clock parameters (assuming all available data are effectively transformed). There are parameter estimation algorithms that make use of this kind of equivalence, for example, the use of the Householder transformations in the square root information filter (SRIF) to produce a set of statistically uncorrelated linear combinations of parameters as a function of linear combinations of

data. Hence, in the SRIF, there is no longer a distinction between stochastic and functional model, and algorithm development becomes extremely easy (for example, as used by *Blewitt* [1989] to facilitate *ambiguity bootstrapping*).

In summary, one can effectively implement the same functional and stochastic model in data estimation using the following methods:

(1) explicit estimation by augmenting the functional model;

(2) implicit estimation by augmenting the stochastic model;

(3) parameter elimination by transforming the data and stochastic model.

This presents a rich variety of possible techniques to deal with parameters, which partly explains the very different approaches that software packages might take. Specific algorithms, such as the square root information filter may effectively embody approaches at once, which illustrates the point that the algorithm itself is not fundamental, but rather the underlying functional and stochastic model.

6.3.3 Applications

Global and Arc Parameters. We sometimes call *x global parameters* and *y local parameters* (if they are localised in space, e.g., for a local network connected to a global network through a subset of stations) or *arc parameters* (if they are localised in time, e.g., coordinates for the Earth's pole estimated for each day). More generally *y* can be called *stochastic parameters*, since it allows us to estimate a parameter that varies (in some statistical way) in either space or time. As we have seen, we don't actually have to explicitly estimate *y*, if all we are interested in are the global parameters, *x*.

Earth Rotation Parameters. A typical daily solution for a global GPS network might contain coordinates for all the stations, plus parameters to model the orientation of the Earth's spin axis in the conventional terrestrial frame and its rate of rotation (for example, X and Y pole coordinates, and length of day). We can then combine several day's solutions for the station coordinates, in which case the station coordinates can be considered global parameters. It is also possible to estimate station velocity at this stage, to account for tectonic motion. Next, we can orient this station coordinate (and velocity) solution to a conventional frame, such as the ITRF (IERS Terrestrial Reference Frame). If we then wished to produce improved estimates for daily Earth rotation parameters in this frame, we could then apply equation (6.25) to compute the corrections:

$$\Delta\hat{y} = -\left(B^T W B\right)^{-1}\left(B^T W A\right)\Delta\hat{x} \tag{6.31}$$

This can easily be done if the coefficient matrix relating $\Delta\hat{y}$ to $\Delta\hat{x}$ is stored along with each daily solution. This is an example of smoothing, without having to resort to the full Kalman filter formulism. Effectively, the Earth rotation parameter have been estimated as white noise parameters. The length of day estimates can then be integrated to form an estimate of variation in the Earth's hour angle (UT1-UTC), which would effectively have been modelled as random walk (which can be defined

as integrated white noise).

Helmert Wolf Method. The spatial analogy to the above is sometimes called the *Helmert-Wolf Method* or *Helmert Blocking.*. The data are instead batched according to geographic location, where the stochastic *y* parameters are the station coordinates of a local network. The *x* parameters comprise station coordinates at the overlap (or *nodes*) between local networks. The *x* parameters are first estimated for each network according to (6.23); then these estimates are combined; finally the *y* parameters can obtained using (6.25). Helmert Blocking seen in this context is therefore simply a specific application of a more general concept.

Troposphere Estimation. The random walk model is commonly used for tropospheric zenith bias estimation, because (1) this closely relates to the expected physics of atmospheric turbulence [*Truehaft and Lanyi*, 1987], and (2) surface meteorological measurements don't provide sufficient information for us to constrain the overall expected bias.

Filtering algorithms can easily allow the tropospheric zenith bias to vary from one data epoch to the next. Traditional least-squares algorithms can also estimate the troposphere stochastically by either explicit augmentation of the set of parameters, or by using the reduced weight matrix (6.24). However, the traditional method is too cumbersome for dealing with a separate parameter at every data epoch, which is why it is common to estimate tropospheric biases which are constant over time periods of an hour or so. Results indicate that this works well for geodetic estimation, but of course, it might be unsatisfactory for tropospheric research.

Clock Estimation. The white noise model is commonly used for clock estimation when processing undifferenced data. This is partly because one does not have to worry about any type of glitch because the a prior correlation is assumed to be zero. As already mentioned, white noise clock estimation is an alternative to the double differencing approach. The advantage, of course, is that we obtain clock estimates, which leads us naturally to the application *precise point positioning*.

Precise Point Positioning. We can consider receiver coordinates as local parameters, connected to each other only through the global parameters (that affect all spatially separated observations), which include orbit, satellite clock, and Earth rotation parameters. The global network of permanent GPS stations is now reaching the point that the addition of an extra station would do very little to change the estimated orbit and satellite clock parameters. We can therefore take the global solution to be one using the current global network, and consider a user's receiver coordinates as the local parameters. Application of (6.25) to a single receivers carrier phase and pseudorange data using the global parameter solution for *x* would therefore give us a precise point position solution for *y*.

This can actually be simplified further. The term $\left(z - A\hat{x}\right)$ in (6.25) is simply the user's receivers data, minus a model computed using the global parameters (orbits, clocks, and Earth rotation parameters). Therefore, we can solve for *y* by only storing the global parameters and the single receiver's data. Further still, the orbit and Earth rotation parameters can be processed to produce a table of orbit positions in the Earth fixed frame.

Putting all of this together, we can therefore see that (1) producing a single receiver point position solution using a precise ephemerides in the Earth fixed frame is essentially equivalent to (2) processing the station's data as double differences together in a simultaneous solutions with the global network's data. The only difference is that the user's receiver cannot influence the orbit solution. This is astonishingly simple, and has revolutionised high precision geodetic research due to the very short time it takes to produce high precision results, which is typically a few minutes for a 24 hour data set [*Zumberge et al.,* 1996].

6.4 FRAME INVARIANCE AND ESTIMABILITY

Strange as it may seem, station coordinates are generally not estimable parameters. This statement may appear ludicrous, given that GPS is supposedly designed to allow us to position ourselves. But position relative to what? In the simple case of point positioning, we are positioning ourselves relative to the given positions of the GPS satellites, in the reference frame known as WGS-84. How are the orbits known? They are determined using the Control Segment's tracking stations at known coordinates. How are these tracking station coordinates known? By using GPS. And so the questions continue, in a circular fashion.

To view this problem clearly, we consider the general case of the one step procedure, estimating all the satellite orbits and all the station coordinates at the same time. In this section, we consider the nature of these coordinates, and consider exactly what is estimable when attempting to position a global network of GPS stations.

6.4.1 Theoretical Development

Insight from Physics. "A view advanced by Einstein, there is a widespread belief among modern physicists that the fundamental equations of physics should possess the same form in all coordinates systems contemplated within the physical context of a given theory" [*Butkov,* 1968]. Although coordinates are essential for the computation of observable models, our intuition is better served if we think of *geometrical objects*, which we can define as *frame invariant* objects. In general terms, such geometrical objects are called *tensors*. Tensors are classified according to their rank. Formally, a tensor of rank *r* is defined as an *invariant linear function of r directions* [*Butkov,* 1968]. The rank of a tensor (not to be confused with the rank of a matrix) tells you the number of indices required to specify the tensor's coordinates. Here, we stick to familiar objects; *vectors* which are tensors of rank 1 (which have a single direction in space), and *scalars* are tensors of rank zero (which have no directionality).

Equations can be explicitly expressed in terms of tensors without reference to coordinates. One must therefore be careful not to confuse the true vector, which is a geometrical object, with the column vector, which is a representation of the vector in a specific coordinate frame. For example, although the coordinates represented in a column vector change under a frame transformation, the true vector does not change.

Vectors and Transformations. A vector can be defined as an *invariant linear*

function of direction [*Butkov,* 1968]. We should really think of the vector as an axiomatic geometrical object, which represents something physical, and is therefore unaffected by frame transformations. We can write the vector **x** in frame F as [*Mathews and Walker,* 1964]:

$$\mathbf{x} = \sum_{\mathbf{i}} x_{\mathbf{i}} \mathbf{e}_{\mathbf{i}} \tag{6.32}$$

in terms of coordinates x_i and base vectors $\mathbf{e}_i$ (vectors which define the direction of the coordinate axes). The same vector can be written in frame F' as:

$$\mathbf{x} = \sum_{\mathbf{i}} x'_{\mathbf{i}} \mathbf{e}'_{\mathbf{i}} \tag{6.33}$$

We can, for analytical convenience, write this equivalence in matrix form, where we define the row matrix **e** and column matrix *x* as follows:

$$\begin{aligned} \mathbf{x} &= \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \equiv \mathbf{e}x \\ &= \begin{pmatrix} \mathbf{e}'_1 & \mathbf{e}'_2 & \mathbf{e}'_3 \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} \equiv \mathbf{e}'x' \end{aligned} \tag{6.34}$$

Notice that both coordinates and the base vectors must change such that the vector itself remains unchanged. This axiomatic invariance of a vector requires that the transformation for the base vectors is accompanied by a related (but generally different) transformation of the coordinates. We can start by following the convenient matrix form of (6.34) to define each base vector of the new frame $\mathbf{e}'_i$ as a vector in the old frame with coordinates γ_{ji}. Coordinates γ_{ji} are elements of the transformation matrix Γ:

$$\begin{aligned} \mathbf{e}'_i &= \sum_j \mathbf{e}_j \gamma_{ji} \\ \mathbf{e}' &= \mathbf{e}\Gamma \end{aligned} \tag{6.35}$$

Using the equivalence relation (6.34), we find the corresponding transformation for coordinates:

$$
\begin{aligned}
\mathbf{x} &= \mathbf{e}x \\
&= \mathbf{e}\left(\Gamma\Gamma^{-1}\right)x \\
&= \left(\mathbf{e}\Gamma\right)\left(\Gamma^{-1}x\right) \\
&= \mathbf{e}'x' \\
&\therefore\ x' = \Gamma^{-1}x
\end{aligned}
\tag{6.36}
$$

Objects such as the coordinates are said to transform *contragradiently* to the base vectors. Objects which transform in the same way are said to transform *cogradiently*.

Scalar Functions and Transformations. The frame transformation, represented by Γ, is called a *vector function*, since it transforms vectors into vectors. In contrast, geodetic measurements can be generally called *scalar functions* of the vectors. The dot product between vectors is an example of a scalar function. Simply take a look at typical functional models used in geodesy, and you will find objects such as the dot product between vectors. We therefore need to look at the theory of scalar functions of vectors and how they transform.

A *linear scalar function* of a vector can be defined in terms of its effect on the basis vectors. For example, in 3-dimensional space, we can define the scalar function as a list of 3 numbers known as the *components* of the scalar function:

$$
\begin{aligned}
\phi\left(\mathbf{e}_1\right) &= \alpha_1 \\
\phi\left(\mathbf{e}_2\right) &= \alpha_2 \\
\phi\left(\mathbf{e}_3\right) &= \alpha_3
\end{aligned}
\tag{6.37}
$$

When the scalar function is applied to a general vector **x**, the result can be written

$$
\begin{aligned}
\phi(\mathbf{x}) &= \phi\left(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3\right) \\
&= \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 \\
&= \begin{pmatrix}\alpha_1 & \alpha_2 & \alpha_3\end{pmatrix}\begin{pmatrix}x_1 \\ x_2 \\ x_3\end{pmatrix} \\
&\equiv \alpha x
\end{aligned}
\tag{6.38}
$$

The result must be independent of reference frame, because the geometrical vector **x** is frame invariant. Therefore we can derive the law of transformation for the scalar components α:

$$\begin{aligned}\phi(\mathbf{x}) &= \alpha x \\ &= \alpha\left(\Gamma\Gamma^{-1}\right)x \\ &= (\alpha\Gamma)\left(\Gamma^{-1}x\right) \\ &= \alpha' x' \\ &\therefore \alpha' = \alpha\Gamma\end{aligned} \tag{6.39}$$

This proves that the scalar components transform cogradiently with the base vectors, and contragradiently with the coordinates.

It would appear that the scalar functions have very similar properties to vectors, but with slightly different rules about how to transform their components. The scalar function is said to form a *dual space*, with the same dimensionality as the original vectors.

Supposing we have a vector **y** in our geodetic system, we can define a special the scalar function that always forms the dot product with **y**. The result can be expressed in matrix form:

$$\begin{aligned}\phi_{\mathbf{y}}(\mathbf{x}) &= \mathbf{y}.\mathbf{x} \\ &= (\mathbf{e}y)^T.(\mathbf{e}x) \\ &= y^T\left(\mathbf{e}^T.\mathbf{e}\right)x \\ &= y^T g x\end{aligned} \tag{6.40}$$

where g is the matrix representation of the *metric tensor*, which can be thought of as describing the unit of length for possible directions in space (here represented in 3 dimensions) [*Misner, Thorne, and Wheeler*, 1973]:

$$g \equiv \begin{pmatrix} \mathbf{e}_1.\mathbf{e}_1 & \mathbf{e}_1.\mathbf{e}_2 & \mathbf{e}_1.\mathbf{e}_3 \\ \mathbf{e}_2.\mathbf{e}_1 & \mathbf{e}_2.\mathbf{e}_2 & \mathbf{e}_2.\mathbf{e}_3 \\ \mathbf{e}_3.\mathbf{e}_1 & \mathbf{e}_3.\mathbf{e}_2 & \mathbf{e}_3.\mathbf{e}_3 \end{pmatrix} \tag{6.41}$$

Comparing (6.38) and (6.40), we see that the components of the dot product scalar function are given in matrix form by

$$\alpha_y = y^T g \tag{6.42}$$

Proper length. One can therefore easily construct such a scalar function for every vector, simply using the vector's coordinates, and the metric properties of the space.

$$\begin{aligned}l_{\mathbf{x}} &= \left|\phi_{\mathbf{x}}(\mathbf{x})\right|^{\frac{1}{2}} \\ &= \left(x^T g x\right)^{\frac{1}{2}}\end{aligned} \tag{6.43}$$

It is easy to prove using all the above definitions, that the length of a vector, defined by (6.43) is completely frame invariant, no matter what kind of transformation is performed. For example, if the frame were scaled up so that a different *unit of length* were being used, the metric tensor would be scaled down to compensate.

In the language of relativity, such a length defined using a 4-dimensional spacetime metric, is called a *proper length*. Proper length which is said to be a *scalar invariant* (i.e., a tensor of rank 0). The geometry expressed by (6.43) is known as a *Riemann geometry*. In a *Riemannian space* (e.g., the surface of a sphere), length is calculated along geodesics, which are in turn defined by the metric tensor. It reduces to Euclidean geometry in the special case that the metric is the identity matrix, in which case we have cartesian coordinates.

In physics, the metric tensor (6.41) is a property of spacetime, to be inferred by experiment. According to special relativity, a natural consequence of the universality of the speed of light is that spacetime according to an *inertial observer* has the metric

$$g_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -c^2 \end{pmatrix} \tag{6.44}$$

It might seem odd that a dot product $\mathbf{e}_t.\mathbf{e}_t$ has a negative value, but if we accept that any reasonable definition of "length" must be frame invariant, that's what experiment tells us! [*Schutz*, 1990]. The proper length between two points of relative coordinates in the rest frame (Δx, Δy, Δz, Δt) is therefore defined as:

$$l_0 \equiv \left(\Delta x^2 + \Delta y^2 + \Delta z^2 - c^2\Delta t^2\right)^{\frac{1}{2}} \tag{6.45}$$

which reduces to our normal concept of *spatial length* between two points at the same time coordinate (Pythagoras Theorem):

$$s_0 \equiv l_0(\Delta t = 0) = \left(\Delta x^2 + \Delta y^2 + \Delta z^2\right)^{\frac{1}{2}} \tag{6.46}$$

Proper length is known experimentally to be frame invariant, as is evidenced by the independence of the speed of light on the motion of the source; so in two different frames moving at constant velocity with respect to each other, we can write:

$$l' \equiv \left(\Delta x'^2 + \Delta y'^2 + \Delta z'^2 - c^2\Delta t'^2\right)^{\frac{1}{2}} = l_0 \equiv \left(\Delta x^2 + \Delta y^2 + \Delta z^2 - c^2\Delta t^2\right)^{\frac{1}{2}} \tag{6.47}$$

But, in general, our normal concept of the *spatial length* would be different!

$$s' \equiv \left(\Delta x'^2 + \Delta y'^2 + \Delta z'^2\right)^{\frac{1}{2}} \neq s_0 \equiv \left(\Delta x^2 + \Delta y^2 + \Delta z^2\right)^{\frac{1}{2}} \tag{6.48}$$

In general relativity, spatial length is affected not only by relative motion, but also

by the gravitational potential. The geometrical error in assuming a 3-dimensional Euclidean space amounts to about 2 cm in the distance between satellites and receivers. The relative geometrical error amounts to 1 part per billion, which is not insignificant for today's high precision capabilities. For convenience, three dimenstional Euclidean space underlies GPS spatial models, with relativity applied to geometrical delay as corrections to Pythagoras Theorem (and with relativistic corrections applied to compute the coordinate time of signal transmission).

Scalar Function Equivalence in Euclidean Space. We therefore procede assuming 3-dimensional Euclidean geometry is adequate, assuming relativistic corrections are applied. In this case, the metric is represented by a 3×3 identity matrix. By inspection of (6.41), we see that the basis vectors would be orthonormal, hence defining a Euclidean space where points are represented by cartesian coordinates.

In Euclidean space, the components of the scalar function are simply the cartesian components of the vector. For each and every cartesian vector, there is a corresponding scalar function with identical components as the vector. One can think of this scalar function as the operator that projects any vector onto itself. In a sense, the vector has been redefined in terms of its own functionality. The set of scalar functions defining the dot product operator with respect to each vector in the geodetic system completely describes the geometry as far as scalar observations are concerned. We can therefore conclude that the dot product operator has an equal footing in representing geometry as the vectors themselves. (Length *is* geometry! [*MTW*, 1973]).

Measurements and Geometry. From a modern physical point of view, measurement models should be independent of the selected coordinate frame. Therefore, the measurement model must be a function of objects which are frame invariant. In geodesy, we are seeking to estimate spatial vectors in a Euclidean space. This seems at first to be problematic, since measurements are scalars, not spatial vectors. This situation can be theoretically resolved, once we realise that vectors have a one to one correspondence with scalar functions defined as the dot products between the vectors. These dot products define the length of a vector, and the angles between them. Fundamentally, the above theoretical development proves that the dot products contain all the geometrical information than did the original vectors. A pertinent example, is the differential delay of a quasar signal between two VLBI telescopes. It represents a projection of the baseline vector connecting the two telescopes in the quasar direction. Many observations would reveal the vector as it projects onto different directions.

We must therefore be careful when it comes to estimating coordinates. From what has been said, only the dot products between vectors are estimable, not the coordinates themselves. Since we have freedom to select different frames it is clear that the set of coordinates in a given frame must be redundant, which is to say, there are more coordinates than is necessary to define the vectors, and therefore scalar functions and modelled observations.

Assuming the above Euclidean model, we can show explicitly the source of this redundancy. Following again the VLBI analogy, consider a baseline vector **x**, and a quasar direction **y**, which is represented in two different solutions, one in frame F, the other in frame F'. Projection is formally represented by the dot product between the

two vectors. The equivalence relation (6.34) tells us that the dot product between any two vectors must give the same answer no matter which frame is used. Since we assume we can construct orthonormal base vectors (of unit length, and at right angles to each other, as in a Cartesian frame), we can write the dot product for both frames as:

$$s \equiv \mathbf{x}.\,\mathbf{y} = x^T y = x'^T y' \tag{6.49}$$

What types of coordinate transformation are allowed that satisfy (6.41)? It can easily be shown that the transformation matrix must be orthogonal; that is its transpose equals its inverse. In matrix notation, let us consider equation (6.49), where we apply a transformation R to go from frame F coordinates to frame F' coordinates:

$$\begin{aligned} x^T y &= x'^T y' \\ &= (Rx)^T (Ry) \\ &= x^T \left(R^T R\right) y \\ &\therefore R^T = R^{-1} \end{aligned} \tag{6.50}$$

Such transformations are called *rotations*, and (6.50) shows the property of rotation matrices. We therefore deduce that global rotations have no effect on dot products computed according to (6.49) (which assumed Euclidean frames).

The analogy in special relativity is the *Lorentz transformation*, which can be considered as a rotation in 4-dimensional spacetime (3 rotations + 3 velocity boosts). Relativistic transformations preserve proper length, but can change spatial length. The only change in scale which is physically acceptable is that due to the relativistic choice of reference frame, which depends on relative speed (special relativity) and the gravitational potential (general relativity). For example, VLBI solutions computed in the barycentric frame (origin at the centre of mass of the solar system) produce baselines with a different scale to SLR solutions computed in the geocentric frame.

6.4.2 Discussion

Space Geodetic Consequences. The corollary of equation (6.50) and the correspondence between vectors and scalar functions, is that space geodetic data cannot provide any information on global rotations of the *entire* system. Since this arbitrariness has 3 degrees of freedom (a rotation), it results in a 3-rank deficient problem. This justifies our original statement that coordinates are not estimable.

In the case of VLBI, the entire system includes quasars, so the rotation of the Earth is still accessible. However, in GPS, the system has satellites that can be only be considered approximately fixed (through dynamical laws which have systematic errors), and therefore we can conclude the data only weakly constrains long period components of Earth rotation. As for relative rotation of the Earth's plates, we can conclude that, on purely geometrical grounds, there is no advantage of VLBI over GPS (this is not to exclude other classes of arguments to favour one or the other).

Geometrical Paradigm. The paradigm for this section is geometrical. In the spirit of Einstein, although conventions and coordinates provide a convenient representation of reality for computational purposes, our intuition is often better served by a geometrical model that is independent of these conventions. The relationship between the geometrical paradigm and the conventional model is discussed below, where we refer the reader to Chapter 1 for a more complete description of the conventional terms.

Consider a network of GPS stations, tracking all the GPS satellites. Using the GPS data, we can estimate the *geometrical* figure defined by the stations and the satellite orbits. That is, GPS provides information on internal geometry, including the distances between stations, and the angles between baselines, and how these parameters vary in time. The geometrical figure defined by the stations is sometimes called *the polyhedron*, particularly in IGS jargon. This is to remind us that, fundamentally, the data can tell us precisely the shape of the figure described by the network of points. For permanent tracking networks, the data can also tell us how the polyhedron's internal geometry changes over time. The elegant aspect of this geometrical picture, is that it more closely relates to quantities that can actually be measured in principle, such as the time it takes for light to travel from one station to another. This is in contrast to coordinates which are frame dependent.

Since GPS orbits can be well modelled over an arc length of a day (2 complete orbits), we have access to an instantaneous inertial frame, which by definition, is the frame in which Newton's laws appear to be obeyed. In historical terminology, GPS data together with dynamical orbit models give us access to an *inertial frame of date*. A frame determined this way, cannot rotate significantly relative to inertial space, otherwise the orbits would not appear to obey Newton's laws.

The system can therefore determine the direction of the instantaneous spin axis of the Earth with respect to the polyhedron. Although the spin axis is not tangible like stations and satellites, it is an example of an estimable vector. For example, GPS can tell us unambiguously the angles between any baseline and the instantaneous spin axis (called the Celestial Ephemeris Pole, CEP). We can therefore determine a station's latitude relative to the CEP without any problem. However, the direction of the CEP as viewed by the polyhedron does wander from one day to the next, a phenomenon known as *polar motion*. It would therefore be impractical to define a station's latitude this way, so instead, a conventional reference pole direction is defined (called the conventional terrestrial pole, or CTP).

The problem is, the choice of CTP is arbitrary, and fundamentally has nothing to do with GPS data. Therefore, conventional station latitudes (relative to the CTP) strictly cannot be estimated, but only true latitudes can (relative to the CEP). This state of affairs is not hopeless; for example, the CTP can be defined by constraining at 2 station latitudes. If we allow for time evolution of the polyhedron (which we must), then we must also specify the time evolution of the CTP with respect to the polyhedron, which again goes beyond GPS, and into the domain of conventions.

GPS is also sensitive to the rate of rotation of the Earth about the CEP. Again, this is because the satellites are forced to obey Newton's laws in our model. Since the spin rate can be estimated, our model can map the time series of station positions in the instantaneous inertial frame back to an arbitrary reference time. We can therefore determine the relative longitude between stations, as angles subtended around the CEP. However, just as for latitudes, the longitudes determined this way would

wander from one day to the next due to polar motion (an effect that is maximum near the poles, and negligible at the equator). Longitudes are therefore also dependent on the choice of CTP. Moreover, only relative longitude can be inferred, since GPS data has no way to tell us exactly the location of the Prime Meridian. Once again, we would have to go beyond GPS, and arbitrarily fix some station's longitude to a conventional value (preferably, near the equator), thus effectively defining the Prime Meridian.

We note in passing that the CEP also varies in inertial space (by nutation and precession). We only need to model this variation over the period for which we are modelling the satellite dynamics, which is typically over a day or so. GPS is therefore insensitive to nutation and precession errors longer than this period, because, in effect, we are defining a brand new inertial frame of date every time we reset the orbit model. The reason for having relatively short orbit arcs (as compared to SLR) is not because of fears about nutation, but rather because of inadequacies in the orbit model. But an orbit arc of a day is sufficient for the purpose of precisely determining the polyhedron, which implicitly requires a sufficiently precise determination of polar motion and Earth spin rate. (The Earth's spin rate is often parameterised as the excess length of day, or variation in UT1-UTC, that is the variation in the Earth's hour angle of rotation relative to atomic time).

Finally, there is another geometrical object to which GPS is sensitive, and that is the location of the Earth's centre of mass within the geometrical figure of the polyhedron. In Keplerian terms, the Earth centre of mass is at the focus for each and every GPS elliptical orbit. Of course, Kepler's laws are only approximate. More precisely, the Earth's centre of mass is the dynamical origin of the force models used to compute the GPS satellite orbits.

If we arbitrarily displaced the polyhedron relative to this origin, we would find the satellite orbits appearing to violate Newton's laws. We therefore can say that GPS can locate *the geocentre*, which is to say that it can determine a displacement of the centre of figure with respect to the centre of mass [*Vigue et al.,* 1992]. Effectively, GPS therefore allows us to estimate geocentric station height, which is the radial distance from the Earth centre of mass. However, it should be kept in mind, that the geocentre estimate is very sensitive to the accuracy of orbit force models, and is not determined as precisely as the geometry of the figure. In fact, vary rarely is true geocentric height variation shown from GPS analyses, but rather height relative to the average figure, which is an order of magnitude more precise, with the (apparent) geocentre variation often displayed separately as a global parameter.

6.4.3 Applications

Free Network Solutions. If we estimated all station coordinates and satellite positions, the 3-rank deficiency in the problem would imply that a solution could not be obtained. However, suppose we apply very loose a priori constraints to the station coordinates. The above theory predicts that our resulting coordinates would still be ill-defined, however the geometry of the figure would be estimable [*Heflin et al.,* 1992]. That is, if we were to compute the dot product between any two vectors in the system, we would find it to be well constrained by the data. Such a solution has been called a *free network solution*, a *fiducial-free solution*, or a *loose solution.*

We discuss below several applications of free network solutions, which for example can be used directly to estimate geophysical parameters of interest, since geophysical parameters depend on scalar functions of the vectors, not the coordinates. For some applications, though, a frame definition may be necessary.

Frame Definition. Although *conventional reference systems* may have ideal notions of the basis vectors and origin, *conventional terrestrial frames* today are *defined* through the coordinates of a set of points co-located with space geodetic instruments. These points serve to define implicitly the directions of the coordinate axes (i.e., the basis vectors). Such frames contain an extremely redundant number of points, and so might also serve as a source of a priori geometrical information.

One could choose to fix all these points, which might be advantageous for a weak data set, where a priori information may improve the parameters of interest. If the data set is strong, finite errors in the frame's geometry will conflict with the data, producing systematic errors. To avoid this problem, we should only impose *minimal constraints*. The large redundancy allows one to define the frame statistically, so that errors on the redundant set of definitions for the X,Y, and Z axis directions are averaged down. This can be achieved using either 3 equations of constraint for orientation, or by using the fiducial free, or free network approach, where all station coordinates are estimated, with the final indeterminate solution being rotated to agree on average with the frame.

Quality Assessment. Internal quality assessment involves looking at the residuals to the observations, after performing the least-squares solution, and assessing the significance of deviations. Residuals are estimable even in the absence of frame definition, and so it is recommended to assess internal quality of the data using free network solutions, prior to applying any constraints, otherwise it would be impossible to distinguish data errors from systematic errors arising from the constraints.

External quality assessment involves comparing geodetic solutions. How can we tell if the geometry of the solutions (i.e., the vectors) are the same, if the solution is only represented by coordinates?

If we do happen to know how the base vectors are related between the two systems, then we can simply transform one set of coordinates and do a direct comparison. This is rarely the case, unless the base vectors are implicitly defined through constrained coordinates in the system.

Alternatively, we can use the fact that, fundamentally, a vector reveals itself through its scalar functions, and check all of the dot products in the system. Obvious candidates for this include baseline length, and angles between baselines.

Thirdly, one can solve for a rotation transformation between the two frames, apply the transformation, and compare all the vector coordinates, which it must be stressed, are to be constructed between *physical* points to which the observations are sensitive.

This last point requires clarification. The position coordinates of a point, for example, do not constitute the coordinates of a physical vector, unless the origin has some physical significance in the model. For VLBI it does not, for SLR it does, and for GPS, there is a degree of sensitivity which depends on global coverage, and other issues. We are allowed to use the Earth centre of mass as a physical point for satellite systems, so "position vector" does become physically meaningful in the special case that the origin is chosen to be at the Earth centre of mass. So, strictly, the theory does

not permit a direct comparison of VLBI station coordinates with, say, GPS; however, it does permit a comparison of the vectors. However, if one were to insist on comparing single station positions, one could remove an estimated translational bias between the frames, but the resulting station coordinates would logically then be some linear combination of all estimated station coordinates, making interpretation potentially difficult.

Finally, as already discussed, a change in scale is not considered an acceptable transformation between frames assuming Euclidean space. Apart from relativistic considerations, scaling between solutions must be considered as systematic error rather than a valid frame transformation.

Coordinate Precision. We should now be able to see that coordinates could not be estimated unless we have observational access to any physical objects that might have been used to define the unit vectors. (For example, a physical inscription marking the Prime Meridian). Coordinate precision therefore not only reflects the precision to which we have determined the true geometry of the figure, but also the precision to which we have attached ourselves to a particular frame. Coordinate precision (e.g., as formally given by a covariance matrix computation) can therefore be a very misleading measure of the geometrical precision.

Geophysical Interpretation. Our ability to attach ourselves to a particular frame has absolutely no consequence to the fundamental physics to be investigated (say, of the Earth, or of satellite orbit dynamics). However particular frames may be easier to express the dynamic models. For example, the inertial frame is better for describing the satellite equations of motion. A terrestrial (co-rotating) frame is easier for describing motion of the Earth's crust. Nevertheless, the estimable quantities will be frame independent.

One pertinent example here is the relative Euler pole of rotation between two plates, with the estimable quantities being, for example, the relative velocity along the direction of a baseline crossing a plate boundary. Another example is crustal deformation due to strain accumulation. Here, the invariant geometrical quantity is the symmetic strain tensor, with the invariant estimable quantities being scalar functions of the strain tensor. However, space geodesy cannot unambiguously state, for example, the velocity coordinates of a point, since that requires arbitrarily defined axes.

When comparing geophysically interesting parameters, one must take care to ensure frame invariance, or at least, approximate frame invariance. For example, comparing station velocity components between solutions, or Euler poles of individual plates will generally show discrepancies that relate to frame definition.

Ambiguity Resolution. This section could more generally refer to all inherently scalar parameters, such as tropospheric or clock parameters. Like these parameters, the carrier phase ambiguities are manifestly frame independent quantities. As a consequence, no frame constraints are necessary at all to estimate ambiguity parameters. In fact, there are good reasons for not including frame constraints. Frame constraints, if not minimal, can distort solutions due to systematic error in the a priori geometry. This can be very undesirable where the a priori information is suspect.

As a test of this concept, *Blewitt and Lichten* [1992] solved for ambiguities on a

global scale using a network solution free of frame constraints, and found they could resolve ambiguities over even the longest baselines (up to 12,000 km).

Covariance Projection. One might wish to compare coordinates after applying a rotation between solutions. Or perhaps one wishes to assess the geometrical strength of the free network solution. In both cases, it is useful to consider the coordinate error as having a component due to internal geometry, and external frame definition. A free network solution is ill-defined externally, but well defined internally. How can we compute a covariance matrix that represents the internal errors?

We can apply a *projection operator*, which is defined as the estimator for coordinate residuals following a least squares solution to rotation. Consider the linearised observation equation which rotates the coordinates into another frame, accounting for possible measurement error:

$$x = Rx' + v \tag{6.51}$$

This can be rearranged so that the 3 unknown angles contained in R are put into a column vector θ, and defining a rectangular matrix A as a linear function of θ such that:

$$A\theta \equiv Rx' \tag{6.52}$$

Therefore, after substitution in to (6.23), we find the least squares estimator for the errors:

$$\begin{aligned}\hat{v} &= x - A\hat{\theta} \\ &= x - A\left(A^T W A\right)^{-1} A^T W x \\ &= \left(I - A\left(A^T W A\right)^{-1} A^T W\right)x\end{aligned} \tag{6.53}$$

The covariance matrix for the estimated errors is therefore:

$$\begin{aligned}C_{\hat{v}} &= \left(I - A\left(A^T W A\right)^{-1} A^T W\right) C_x \left(I - A\left(A^T W A\right)^{-1} A^T W\right)^T \\ &= C_x - A\left(A^T W A\right)^{-1} A^T \\ &= C_x - A C_{\hat{\theta}} A^T\end{aligned} \tag{6.54}$$

This is called *projecting* the covariance matrix onto the space of errors [*Blewitt et al.*, 1992]. Since these errors are scalar quantities (independent of frame), they represent the geometrical errors of the system. Therefore, the projected covariance matrix is a formal computation of the precision to which the geometry has been estimated, without us having to define a frame.

Note from (6.54) that we can write the original covariance matrix for coordinates as:

$$\begin{aligned} C_x &= C_{\hat{v}} + A\left(A^T W A\right)^{-1} A^T \\ &= C_{\hat{v}} + A C_{\hat{\theta}} A^T \\ &= C_{\text{internal}} + C_{\text{external}} \end{aligned} \tag{6.55}$$

This shows explicitly that the coordinate covariance can be decomposed into a covariance due to internal geometrical errors, and an external term which depends on the level of frame attachment.

Loosening Transformation. In the case of *loose solutions*, in effect the orientation parameters have loose a priori constraints. If this constraint can be represented by the (large) a priori covariance matrix E (*external*), equation (6.54) would more correctly read (see 6.17b):

$$\begin{aligned} C_x &= C_{\hat{v}} + A\left(A^T W A + E^{-1}\right)^{-1} A^T \\ &\approx C_{\hat{v}} + A E A^T \end{aligned} \tag{6.56}$$

where we use the fact that the data themselves provide no information on global orientation, hence the components of $A^T W A = C_{\hat{\theta}}^{-1}$ can be considered negligibly small.

We call (6.56) a *loosening transformation*, or a *covariance augmentation*. The resulting covariance is often called a *loosened covariance*, or *loosened solution* (even though we have not changed the estimates themselves). It can be applied, for example, to network solutions that have a well defined constraint in orientation, for applications where we wish to effectively remove the frame definition. Once augmented in this way, the coordinate covariance can then be projected onto another frame, applying the projection operator.

Equation (6.55) should look familiar. We have actually seen it before in equation (6.26), in the context of augmenting the stochastic model as an alternative to estimating extra parameters. Effectively, this is telling us that a combination of loosened solutions is equivalent to estimating and removing a relative rotation between constrained networks and combining them. It also tells us that it is unnecessary to estimate and remove relative rotations between loose solutions prior to combination.

This has very practical applications when combining network solutions from various analysis groups, who might apply different minimal coordinate constraints. Upon receiving a coordinate solution with full covariance matrix, one can procede to loosen the covariance matrix prior to combination with other solutions. Therefore, one does not have to estimate and apply transformation parameters every time the coordinate solution is processed. Moreover, covariance loosening has the elegant aspect in that the fundamental rank-3 deficiency is represented in an obvious way to the user, as the diagonal elements of the covariance matrix will be large, with the off-diagonal elements containing the geometrical information to which the data are truely sensitive.

As an example, the IGS Densification Program (IDP) currently uses the above concept of combining loose covariance matrices from a variety of analysis centres.

Algorithm development for manipulating such solutions becomes very straightforward, when one does not have to worry about solutions being contrained to different frames. The IDP also, once again, illustrates the concept of global and local parameters, with each regional network being connected to the global network using 3 common *anchor stations*. Using 3 anchor stations allows for the implicit estimation of relative rotation when combining regional and global network solutions [*Blewitt et al*, 1993 and 1995].

6.5 SUMMARY AND CONCLUSIONS

We are now in a position to summarise some of the most important conclusions in terms of a few maxims, which are purposely expressed in an informal way to appeal to an intuitive mode of thinking.

6.5.1 Equivalence of Pseudorange and Carrier Phase

Models for the pseudorange can be constructed using carrier phase, and visa versa
This allows us to develop algorithms that use both data types to:

- Estimate pseudorange multipath
- Edit carrier phase and pseudorange data for outliers
- Edit carrier phase data for cycle slips
- Smooth the pseudorange using the carrier phase
- Process undifferenced data without a preliminary point position solution
- Resolve carrier phase ambiguities in a model independent way

6.5.2 Equivalence of the Stochastic and Functional Models

(covariance augmentation) ≡ (weight reduction) ≡ (estimation)=(data combination)
This allows us to develop algorithms to:

- separately estimate global and local parameters
- partition problems in time
- partition problems in space
- remove nuisance parameters
- implicitly estimate parameters
- estimate stochastic parameters
- estimate precise point positions using single receivers

6.5.3 Frame Invariance and Estimability

(invariant geometry)=(tensors, vectors, scalars) ≡ (scalar functions)=(observations)
This allows us to understand:

- a geometrical paradigm for space geodesy
- which parameters are estimable
- problems with coordinate estimability

- frame definition
- the importance and utility of free network solutions
- internal and external components of coordinate error
- covariance projection as a means to quantify geometrical error
- loosening transformation to remove rotational information
- network combination analysis

6.5.4 Concluding Remark

What I hope to have achieved in this chapter is (1) specifically, to impart an intuitive understanding of certain aspects of GPS data processing and estimation, and (2) more generally, to have inspired a change in the way we might think about GPS data processing problems in general, by looking for patterns, symmetries, and equivalences, and exploiting these so that answers to questions become more obvious.

Acknowledgement

I would like to thank the late Professor Richard P. Feynman as a teacher and researcher during my years at the California Institute of Technology, for his inspiration.

References

Bierman, G., *Factorization methods for discrete sequential estimation*, Academic Press, New York, NY (1977).

Blewitt, G., Carrier phase ambiguity resolution for the Global Positioning System applied to geodetic baselines up to 2000 km, *Journ. Geophys. Res., Vol. 94*, No. B8, pp. 10187-10283 (1989)

Blewitt, G., An automatic editing algorithm for GPS data, *Geophys. Res. Lett., Vol. 17*, No. 3, pp. 199-202 (1990).

Blewitt, G., A new tool for dense monitoring of crustal strain: GPS rapid static surveying, *Eos. Trans. Am. Geophys. U., Vol. 71*, No. 17, p. 483, (1990).

Blewitt, G., and S.M. Lichten, Carrier phase ambiguity resolution up to 12000 km: Results from the GIG'91 experiment, in *Proc. of the Sixth Int. Symp. on Satellite Positioning*, Columbus, Ohio State University, USA (1992)

Blewitt, G., M.B. Heflin, F.H. Webb, U.J. Lindqwister, and R. P. Malla, Global coordinates with centimeter accuracy in the International Terrestrial Reference Frame using the Global Positioning System, *Geophy. Res. Lett., 19*, pp. 853-856 (1992).

Blewitt, G., Y. Bock, and G. Gendt, Regional clusters and distributed processing, in *Proc. of the IGS Analysis Center Workshop*, Ed. by J. Kouba, pp. 62-91, Ottawa, Canada (1993).

Blewitt, G., Y. Bock, and J. Kouba, "Constructing the IGS polyhedron by distributed processing," in Proc. of the IGS Workshop, ed. by J. Zumberge, IGS Central Bureau, Pasadena, Calif., USA, p. 21-36 (1995)

Butkov, E., *Mathematical physics*, Addison-Wesley, Amsterdam (1968)

Heflin, M.B., W.I. Bertiger, G. Blewitt, A.P. Freedman, K.J. Hurst, S.M. Lichten, U.J. Lindqwister, Y. Vigue, F.H. Webb, T.P. Yunck, and J.F. Zumberge, Global geodesy using

GPS without fiducial sites, *Geophysical Research Letters, Vol. 19*, pp. 131-134 (1992).

Mathews, J., and R.L. Walker, *Mathematical methods of physics*, Benjamin Cummings, Amsterdam (1964)

Melbourne, W.G., The case for ranging in GPS based geodetic systems, in Proc. of 1st Int. Symp. on Precise Positioning with the Global Positioning System, U.S. Dept. of Commerce, Rockville, MD (1985).

Misner, C. W., K.S. Thorne, and J.A. Wheeler, *Gravitation*, Freeman, San Francisco (1973)

Wu, J.T., S.C. Wu, G.A. Hajj, W.I. Bertiger, and S.M. Lichten, Effects of antenna orientation on GPS carrier phase, *Manuscripta Geodaetica, 18*, pp. 91-93 (1993).

Schutz, B.F., A first course in general relativity, Cambridge Univ. Press, Cambridge (1990)

Truehaft, R.N. and G.E. Lanyi, The effects of the dynamic wet troposphere on radio interferometric measurements, *Radio Science, 22*, pp. 251-265 (1987).

Vigue, Y., S.M. Lichten, G. Blewitt, M.B. Heflin, and R.P. Malla, Precise determination of the Earth's center of mass using measurements from the Global Positioning System," *Geophys. Res. Lett., 19*, pp. 1487-1490 (1992)

Zumberge, J.F., M.B. Heflin, D.C. Jefferson, M.M. Watkins, and F.H. Webb, Precise point positioning for the efficient and robust analysis of GPS data from large networks, *Journ. Geophys. Res., 102*, No. B3, pp. 5005-5018 (1997).