The Trouble with Constrained Adjustments

Charles R. Schwarz

ABSTRACT. *A constrained adjustment, in which a new survey is fit to existing control points, produces results that are at least as good as, and usually better than, the corresponding free adjustment. However, the proof of this property depends on the assumption that the uncertainty of the fixed control is much smaller than the uncertainties of the new survey. When this assumption is not fulfilled, the usual error-propagation equations must be extended to take into account the effects of the uncertainties of the fixed control points. The opposite conclusion then can be reached: It is possible for adjusted observations to have greater errors than the observed values, so the constrained-adjustment procedure can indeed degrade a perfectly good survey and produce results that are worse than the free adjustment.*

Introduction

Constrained adjustments are quite common in the processing of survey data. Every time we adjust a new survey into an existing coordinate system by using existing control points, we are performing a constrained adjustment.

The control network is intended to help surveyors place their surveys into some larger coordinate system, detect blunders in their observations, and control the build-up of the effect of observational errors on the adjusted coordinates. However, there are circumstances under which control networks become inadequate for their intended purpose. When this happens, surveyors may have difficulty fitting a new survey into the existing control network. Misclosures may be much larger than expected, and the difference between observed and adjusted values of observations may be much larger than can be explained by observational error.

Free and Constrained Adjustments

In the majority of least-squares adjustment problems, the unknown parameters are the coordinates of physical points. When coordinates are used, it is usually necessary to fix the coordinates of one or more points to define the coordinate system. The survey observations alone are not sufficient. Angle observations are completely independent of any coordinate system, and therefore cannot tell us anything about actual coordinates. Distance observations tell us only about the scale of a coordinate system, not its orientation or position.

In an adjustment one can fix a coordinate by including an appropriate equation that specifies the value to be assigned to the coordinate, such as $x_1 = 0$, $y_1 = 0$. Such equations have the same form as regular observation equations, but do not represent actual observations. They are sometimes called "direct observations of coordinates" and sometimes called "constraint equations."

Conventionally, we use the words "free adjustment" to describe an adjustment that uses just the number of constraint equations necessary to define the coordinate system, but no more. When more constraint equations are used, we say that we have a "constrained adjustment." The wording is perhaps a bit misleading, since a free adjustment indeed can include constraint equations (those necessary to define the coordinate system). Many authors prefer the phrase "minimal constraint adjustment" to denote a free adjustment; unfortunately, the use of this more descriptive phrase is not universal. When more than the minimum number of constraint equations are used, the resulting adjusted quantities are constrained not only to be in the proper coordinate system, but also to fit the additional constraints.

Consider the horizontal survey shown in Figure 1. Suppose that points I and J are pre-existing marks and we run a traverse between them, setting the new marks 1 and 2 in the process. We measure the distances I - 1, 1 - 2, and 2 - J, as well as the angles I - 1 - 2 and 1 - 2 - J. Thus we have five measurements with which to determine the four coordinates of the two new points—a redundancy of one.

There are at least two common ways of treating the coordinates of the old points. In a horizontal network that contains distance observations, we need three quantities to define the coordinate system—two to define the origin and one for the orientation. Thus we might perform a free adjustment by constraining both coordinates of point I and one of the two coordinates of point J. Alternatively, we might constrain the two coordinates of point I and the azimuth from I to J.

Free adjustments have the disturbing property that things move when they should stay fixed. In a free

Charles R. Schwarz is Chief of Systems Development for the National Geodetic Survey, 1315 East-West Highway, Silver Spring, MD 20910 and is editor of Surveying and Land Information Systems.

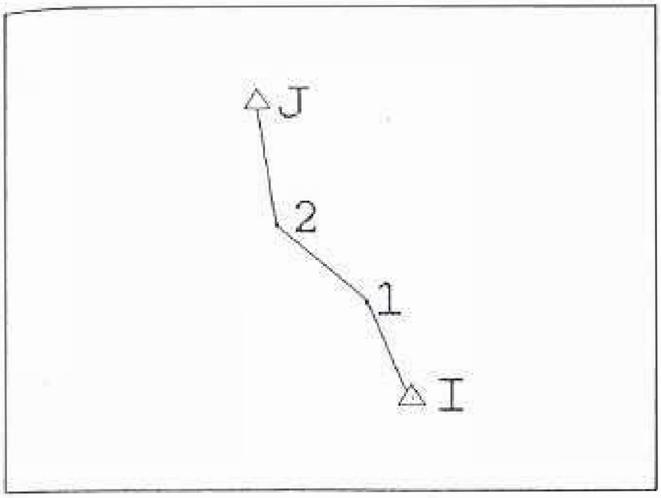

Figure 1. *Sample traverse.*

adjustment of the example network, point J is still free to move in one direction. This is not good, since the coordinates of point J have already been determined and published. It might be preferable to make sure that the existing control stays fixed by constraining both coordinates of both point I and point J in a constrained adjustment.

Why a Constrained Adjustment is Good

Our intent is that the coordinates of the old points I and J serve to "control" the new survey. These old coordinates actually accomplish this in three different ways. First, they serve to define the origin and orientation of the new survey so that the coordinates of the new points 1 and 2 are in the same coordinate system as the old points. Second, they provide a means of detecting blunders in the new survey. Third, the constrained adjustment dampens the build-up of the effect of accidental error.

The argument about constraining the effect of accidental observational errors goes like this: The coordinates of the existing points are assumed to be "correct." If the free adjustment has a misclosure at point J, it must be because of errors in the new survey. If the misclosure is large, we should look for a blunder in the observations. If it is within the tolerance allowed for this type of survey, we distribute the misclosure. The resulting adjusted observations are more accurate than the observed values, and the adjusted coordinates from the constrained adjustment are more accurate than those from the free adjustment.

We can show this mathematically. The constraint equations that are used to fix the coordinates of the control points can be treated as regular observations whose associated variance is zero. Thus we have nine observations altogether—five from the new survey and four "observations" of the coordinates of the two old points. We also have eight unknown parameters altogether—two coordinates for each of the four points. Let the total set of observation equations be written in standard notation as

$$AX = L + V \tag{1}$$

where A is the design matrix (partial derivatives of the observations with respect to the parameters), X is the vector of unknown parameters (or corrections to approximate values of parameters), L contains the observed values (observed minus computed terms), and V is the vector of residuals.

We partition these nine observation equations into three groups. Let

$A_1X = L_1 + V_1$ be the five observation equations arising from the new survey.

$A_2X = L_2 + V_2$ be the three observations of old coordinates (or functions of old coordinates) that are used in the free adjustment to define the coordinate system. Clearly these equations do not involve the coordinates of the new points 1 and 2, so A_2 will have zeroes in the columns corresponding to those coordinates in X.

$A_3X = L_3 + V_3$ be the remaining observation of an old coordinate (or function of an old coordinate).

Let the covariance matrices associated with these three sets of observations be denoted Σ_1, Σ_2, and Σ_3, respectively. Since the coordinates of the old control points are to be fixed, we will use $\Sigma_2 = 0$ and $\Sigma_3 = 0$. However, it will not hurt to carry these quantities symbolically.

If we perform an adjustment with only the first two sets of observations, we obtain the free-adjustment estimate X^- of X, with covariance matrix Σ^-. If we then sequentially add the third set, we obtain the updated (constrained) estimate

$$X^+ = X^- + \Sigma^- A_3^T \left(\Sigma_3 + A_3 \Sigma^- A_3^T\right)^{-1} (L_3 - A_3 X^-) \tag{2}$$

The covariance matrix of the updated estimate is

$$\Sigma^+ = \Sigma^- - \Sigma^- A_3^T (\Sigma_3 + A_3 \Sigma^- A_3^T)^{-1} A_3 \Sigma^- \tag{3}$$

This is a well-known equation. With a change of notation, it is equation (4.118) in Leick (1990) or equation (12.5a) in Mikhail (1976). The second term on the right is a positive semidefinite matrix (whether or not $\Sigma_3 = 0$). Positive semidefinite matrices are analogous to numbers that are greater than or equal to zero. Since Σ^+ is equal to Σ^- minus a positive semidefinite matrix, we say that $\Sigma^+ \leq \Sigma^-$. This means that the variance of any scaler function of X^+ is less than or equal to the variance of the same function evaluated at X^-. Intuitively, it means that by adding new information (the third set of equations) to an old set, we cannot make things worse, and generally make things better.

In principle, it is possible to make a new observation that gives no new information about the parameters. For instance, we could make an additional observation of a parameter that is already fixed, such as one of the coordinates of point I in the example. This is why the second term on the right of equation (3) can be zero. In practice, this almost never happens. In practice, almost all new observations (including redundant constraints) help. Sometimes they help only a little, but more often they make the results much better.

Why a Constrained Adjustment May Not Be So Good

The previous section seems to prove that the constrained adjustment is at least as good as, and may be much better than, the free adjustment. Furthermore, the constrained adjustment uses all the information available to us, which is intuitively preferable to a procedure that ignores some data. Why, then, do we hear surveyors complain that they have to "distort" or "degrade" highly accurate GPS surveys to fit the existing NAD 83 control?

The answer is that the error-propagation equations given above, and indeed all the error-propagation equations usually associated with least-squares adjustments, depend on the assumption that the adjustment was performed with a weight matrix that is inversely proportional to the covariance matrix of the observations (i.e., $W = \sigma_0^2\Sigma^{-1}$). This assumption does not hold when we fix the control points, since we then carry out the adjustment as if the variances of the coordinates of these points were all zero, while we know that these points are not known perfectly.

Least-squares estimates are often said to be optimal estimates or, equivalently, minimum variance linear unbiased estimates. This means that the least-squares algorithm can be derived from the principle that the covariance matrix of the estimated parameters must be smallest among all possible linear unbiased estimates that satisfy the observation equations. The principle of minimum variance really goes to the heart of the matter—it says that we should pick the estimate that is the most accurate. For this reason, many analysts find the principle of minimum variance to be more satisfying than the principle that simply says to minimize the sum of squares of the residuals. However, when the least-squares equations are derived from the principle of minimum variance, we must explicitly use a weight matrix that is inversely proportional to the covariance matrix of the observations (Appendix C).

This means that least-squares adjustments using a weight matrix that is not inversely proportional to the covariance matrix of the observations do not have the minimum variance property. Since they are not optimal, we can say that they are suboptimal. In spite of being less than optimal, such adjustments are done all the time. In fact, every constrained adjustment in which the control points are held fixed is suboptimal.

Effect of Uncertainties of the Fixed Control

The familiar equation

$$\Sigma_{XX} = \sigma_0^2 N^{-1} = \sigma_0^2 (A^T W A)^{-1} \qquad (4)$$

which says that the covariance matrix of the parameters is proportional to the inverse of the normal equations, does not apply without modification to constrained adjustments. The modified equation is

$$\Sigma_{XX} = \sigma_0^2 (A^T W A)^{-1} + (A^T W A)^{-1} A^T W B \Sigma_{CC} B^T W A (A^T W A)^{-1} \qquad (5)$$

where B contains the partial derivatives of the five new observations with respect to the four coordinates of the two control points I and J, and Σ_{CC} is the correct 4x4 covariance matrix of the coordinates of the control points. Since this equation is not well known, a derivation is given in Appendix B.

Equation (5) says that the 4x4 covariance matrix of the coordinates of the two new points is the sum of two terms. The first term gives the contribution of the variance of the five new observations, and might be called the internal error; the second gives the contribution of the real uncertainty of the fixed control, and might be called the external error. Thus we might write

$$\Sigma_{XX} = \Sigma_{int} + \Sigma_{ext} \qquad (6)$$

Equation (5) provides a mathematical explanation of how control networks become inadequate. The classical concept, of course, is that the control network is supposed to be much more accurate than the new densification survey. Mathematically, this means that Σ_{CC} should be so small (in comparison with Σ) that the second term in equation (5) is much smaller than the first term. As long as this is so, equation (4) can be used as a reasonable approximation of equation (5).

This is indeed how classical control networks are developed. We expect a rough correlation between purpose and accuracy: Primary networks should be surveyed to first-order accuracy; secondary networks to second-order, etc. As long as this rough correlation holds, we can use equation (4) instead of (5).

The concept falls apart if the accuracy of the new survey approaches or exceeds that of the existing control points. For instance, if we try to fit a second-order traverse between two third-order points, the result is not what is expected of second-order work. The uncertainty of the new points must be computed by equation (5), not equation (4). Unfortunately, this is almost never done in practice, with the result that we often do not know how to describe the accuracy of such points.

We also can look at what happens to the adjusted observations when the existing control points are held fixed. As shown in Appendix B, the covariance matrix of the adjusted observations also consists of two terms. For example,

$$\Sigma_{L_aL_a} = \sigma_0^2 A(A^TWA)^{-1}A^T + [A(A^TWA)^{-1}A^TW - I] \times B\Sigma_{CC}B^T[A(A^TWA)^{-1}A^TW - I]^T \quad (7)$$

If the second term in this equation vanishes, then we are left with the conventional expression

$$\Sigma_{L_aL_a} = \sigma_0^2 A(A^TWA)^{-1}A^T \quad (8)$$

In this case, the difference between the covariance matrix of the actual observations and that of the adjusted observations is

$$\Sigma - \Sigma_{L_aL_a} = \Sigma - \sigma_0^2 A(A^TWA)^{-1}A^T = [I - A(A^TWA)^{-1}A^TW] \times \Sigma[I - A(A^TWA)^{-1}A^TW]^T \quad (9)$$

This is a positive semidefinite matrix. Thus we can write

$$\Sigma_{L_aL_a} \leq \Sigma \quad (10)$$

which says that the variance of an adjusted observation is always at least as small as the variance of the actual observation (i.e., the adjusted observations are better).

If the second term in equation (7) does not vanish, equation (10) does not necessarily hold. In fact, it is quite possible that the variances of the adjusted observations could be larger than the variances of the corresponding actual observations. In other words, if we fix the control points, we might cause the adjusted values of the observations to be worse than the actual observed values.

The same arguments apply when we try to fit GPS vectors accurate to 1:1,000,000 into the existing NAD 83 network, accurate to about 1:300,000. We can indeed adjust these vectors while holding the existing control fixed, but the covariance matrix of the new points must then be computed by equation (5), not equation (4). The covariance matrix of the adjusted observations must be computed by equation (7), and equation (10) may not hold.

Effects on Free Adjustments

Equation (5) also holds for a free adjustment. We might perform a free adjustment by fixing only those coordinates necessary to define the coordinate system. Following the normal least-squares algorithm, we would compute the covariance matrix in equation (4). However, this only gives us the uncertainty in the adjusted coordinates that is due to the uncertainties of the new observations. It tells us how well the coordinates of the new points are known relative to the fixed control, but not how well they are known relative to the datum as a whole. The second term in equation (5) accounts for the contribution of the uncertainty of the fixed control.

A free adjustment can be shown to have the property that the columns of matrix B are linear combinations of the columns of matrix A, say $B = AH$ for some matrix H. Then

$$(A^TWA)^{-1}A^TWB = (A^TWA)^{-1}A^TWAH = H$$

and equation (5) becomes

$$\Sigma_{XX} = \sigma_0^2(A^TWA)^{-1} + H\Sigma_{CC}H^T \quad (11)$$

Even more interesting, we then have

$$[A(A^TWA)^{-1}A^TW - I]B = [A(A^TWA)^{-1}A^TWAH - AH] = 0$$

so that the second term in equation (7) vanishes. This means that equation (10) holds for all free adjustments, irrespective of how the coordinate system is defined and of the uncertainty of the fixed control. The coordinates obtained in a free adjustment may be affected by the errors in the fixed control, but the adjusted observations are not. This is the sense in which these adjustments are "free."

Practical Implications

Many surveyors have an intuitive grasp of these mathematical results. They say that the constrained adjustment "distorts" their observations. This does not mean that the observed values are actually changed; it means that the adjusted values of the observations are more uncertain, and could, therefore, have greater errors than the observed values. They rebel against this possibility; no one wants his or her work to be "degraded" by putting it through a process that can produce worse results than one started with.

Thus many surveyors processing GPS vectors are rejecting constrained adjustments in favor of free adjustments, for which equation (10) holds. Others are required by contract to fit their GPS surveys into the existing control network, but are uncomfortable with this requirement to do so.

The problems described here mathematically are indeed the trouble with constrained adjustments, and the trouble with the entire concept of a hierarchy of control networks in which the more accurate networks control the lower-order surveys. From time to time, new technology comes along that allows new surveys to be performed with higher accuracy than the existing control network. When this happens, the extended error-propagation equations developed in this article must be used, with the unhappy result that equation (10) may not hold.

This situation has arisen twice in this century. In the 1960s, the introduction of electronic distance measurement equipment allowed new surveys to be performed with greater accuracy than the existing NAD 27. This eventually led to the creation of NAD 83. Now the same situation is occurring again. GPS surveys can be performed with greater accuracy than NAD 83. It is likely that this situation sooner or later will lead to the computation of a new continental datum.

ACKNOWLEDGMENT

The author thanks William E. Strange, chief geodesist of the National Geodetic Survey, who arranged for two anonymous reviews of this article.

REFERENCES

Leick, Alfred. 1990. *GPS Satellite Surveying*. New York: John Wiley and Sons.

Mikhail, Edward M. 1976. *Observations and Least Squares*. New York: IEP—A Dun-Donnelley Publisher.

Appendix A: Linear Error Propagation

Let X be a vector of random variables and let $Y = f(X)$ be a vector of functions of X. Assume that the covariance matrix Σ_{XX} is known. Then the covariance matrix of Y is

$$\Sigma_{YY} = \left(\frac{\partial Y}{\partial X}\right) \Sigma_{XX} \left(\frac{\partial Y}{\partial X}\right)^T \qquad (12)$$

With a change of notation, this is equation (4.34) of Leick (1990) and equation (4.40) of Mikhail (1976).

Appendix B: Effect of Unestimated Parameters

In the example traverse shown in Figure 1, we have four points and eight coordinates altogether. Let us partition these into two sets. Let X_N be the four coordinates of the two new points 1 and 2, and let X_C be the four coordinates of the two existing control points I and J.

Mathematical Development

The five observations in the traverse shown in Figure 1 involve all eight unknowns. This set of five observation equations can be written

$$AX_N + BX_C = L + V \qquad (13)$$

where the covariance matrix associated with these five observations is Σ.

We also wish to add four constraint equations for the coordinates of the existing control points. We write

$$X_C = L_C + V_C \qquad (14)$$

where the covariance matrix associated with these four constraint equations is Σ_{CC}.

The total set of all nine equations is now

$$\begin{pmatrix} A & B \\ 0 & I \end{pmatrix} \begin{pmatrix} X_N \\ X_C \end{pmatrix} = \begin{pmatrix} L \\ L_C \end{pmatrix} + \begin{pmatrix} V \\ V_C \end{pmatrix} \qquad (15)$$

with covariance matrix

$$\begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma_{CC} \end{pmatrix} \qquad (16)$$

The most correct way to treat all these data is to perform the minimum-variance adjustment, which is an adjustment of the complete system (15) using a 9x9 weight matrix that is inversely proportional to (16). Of course, this is almost never done, since it might result in changes to the coordinates of the existing control points.

To perform a constrained adjustment, we arbitrarily (i.e., without mathematical justification) set the residuals V_C in (14) to zero. The result $X_C = L_C$ is substituted into equation (13), which is rearranged to read

$$AX_N = L - BL_C + V \qquad (17)$$

This system of five observation equations in four unknowns is adjusted with a weight matrix W that is inversely proportional to Σ, yielding the estimate

$$\hat{X}_N = (A^TWA)^{-1}A^TW(L - BL_C) \qquad (18)$$

Since the coordinates of the existing control points X_C should have been carried as unknowns but were not, they are called "unestimated parameters." Even though these coordinates are not estimated in the constrained adjustment, we can still take account of their effect when we perform error propagation.

The estimate in (18) has two sources of error—the errors in the five traverse observations L and the errors in the coordinates of the existing control L_C. Since these two groups of quantities were determined by different people at different times, we can reasonably assume that they are independent. Thus the total set of independent variables is

$$\bar{L} = \begin{pmatrix} L \\ L_C \end{pmatrix}$$

and the covariance matrix of this vector is given by (16). The partial derivatives are

$$\frac{\partial \hat{X}_N}{\partial \mathbf{L}} = \begin{pmatrix} \frac{\partial \hat{X}_N}{\partial \mathbf{L}} & \frac{\partial \hat{X}_N}{\partial \mathbf{L}_C} \end{pmatrix}$$
$$= ((A^TWA)^{-1}A^TW \quad -(A^TWA)^{-1}A^TWB) \qquad (19)$$

Thus the covariance matrix Σ_{XX} of the estimate in (18) is

$$\Sigma_{XX} = ((A^TWA)^{-1}A^TW \; -(A^TWA)^{-1}A^TWB)$$
$$\times \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma_{CC} \end{pmatrix} \begin{pmatrix} WA(A^TWA)^{-1} \\ -B^TWA(A^TWA)^{-1} \end{pmatrix}$$
$$= (A^TWA)^{-1}A^TW\Sigma WA(A^TWA)^{-1}$$
$$+ (A^TWA)^{-1}A^TWB\Sigma_{CC}B^TWA(A^TWA)^{-1}$$
$$= \sigma_0^2(A^TWA)^{-1}$$
$$+ (A^TWA)^{-1}A^TWB\Sigma_{CC}B^TWA(A^TWA)^{-1} \qquad (20)$$

Similarly, the adjusted value of the five traverse observations is

$$\mathbf{L}^a = AX_N + B\mathbf{L}_C$$
$$= A(A^TWA)^{-1}A^TW\mathbf{L}$$
$$- [A(A^TWA)^{-1}A^TW - I]B\mathbf{L}_C \qquad (21)$$

and the covariance matrix of the adjusted observations is

$$\Sigma_{L^aL^a} = \sigma_0^2A(A^TWA)^{-1}A^T$$
$$+ [A(A^TWA)^{-1}A^TW - I]$$
$$\times B\Sigma_{CC}B^T[A(A^TWA)^{-1}A^TW - I]^T \qquad (22)$$

A Numerical Example

To keep the numerical example small, we reinterpret Figure 1 to be a drawing of a leveling network. Points G and J are now assumed to be benchmarks in the national vertical network. The object of the new survey is to determine the elevations of the new points 1 and 2. Observed elevation differences are accumulated, setup by setup, between the marked points, resulting in the following observations:

Obs.	Model	Value (m)	Distance (km)
l_1	$H_1 - H_G$	5.013	100
l_2	$H_2 - H_1$	−17.062	200
l_3	$H_J - H_2$	42.771	100

The published elevations of points G and J are H_G = 123.113 meter and H_J = 153.805 meter. From the adjustment of the national network, we have

$$\sigma_G^2 = 0.010 \text{ m}^2$$
$$\sigma_J^2 = 0.010$$
$$\sigma_{GJ} = 0.0075$$

or, in matrix form,

$$\Sigma_{CC} = \begin{pmatrix} 0.010 & 0.0075 \\ 0.0075 & 0.010 \end{pmatrix}$$

The leveling is done to specifications that result in an uncertainty of elevation difference of $0.004\sqrt{K}$ meters, where K is the length of the line in kilometers.

Of course, in practice we are not usually given formal standard errors of the elevations of points in the national network. It would be even more unusual (almost unheard of) were we actually to be given a formal covariance between two elevations. Nevertheless, such numbers do exist in principle, and the numbers given here are reasonable estimates of what might be obtained in a real network. Note that the elevation errors at points G and J have a significant positive correlation (0.75). This says that points close together share some of the same error sources.

We select a value of the reference variance of σ_0^2 = 0.0016 and compute the weights as

Obs.	Model	Value (m)	Distance (km)	σ^2	w
l_1	$H_1 - H_G$	5.013	100	0.0016	1
l_2	$H_2 - H_1$	−17.062	200	0.0032	½
l_3	$H_J - H_2$	42.771	100	0.0016	1

The observation equations are then

$$\begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \begin{pmatrix} 5.013 \\ -17.062 \\ 42.771 \end{pmatrix} - \begin{pmatrix} -1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 123.113 \\ 153.805 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

This is in the form of equation (17), so that we immediately identify

$$A = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} \qquad \mathbf{L} = \begin{pmatrix} 5.013 \\ -17.062 \\ 42.771 \end{pmatrix}$$

$$B = \begin{pmatrix} -1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \qquad \mathbf{L}_C = \begin{pmatrix} 123.113 \\ 153.805 \end{pmatrix}$$

The weight matrix is

$$W = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \tfrac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

We compute

$$(A^TWA)^{-1} = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$$

and, by equation (18),

$$\hat{X}_N = \begin{pmatrix} H_A \\ H_B \end{pmatrix} = \begin{pmatrix} 128.1185 \\ 111.0415 \end{pmatrix}$$

The true covariance matrix is computed by equation

(20). We get

$$A^T W B = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

and

$$\Sigma_{xx} = 0.0016 \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix} + \begin{pmatrix} 0.0090625 & 0.0084375 \\ 0.0084375 & 0.0090625 \end{pmatrix}$$

$$= \begin{pmatrix} 0.0012 & 0.0004 \\ 0.0004 & 0.0012 \end{pmatrix} + \begin{pmatrix} 0.0090625 & 0.0084375 \\ 0.0084375 & 0.0090625 \end{pmatrix}$$

$$= \begin{pmatrix} 0.0102625 & 0.0088375 \\ 0.0088375 & 0.0102625 \end{pmatrix}$$

As expected, the uncertainty of the fixed control points dominates this expression. The uncertainties of the elevations of the new points are much larger than would have been expected from the accuracy with which the new survey was performed. The elevations of two new points are also highly correlated, since they share the uncertainties of the control points.

The covariance matrix of the adjusted observations can be found by evaluating equation (22). This yields

$$\Sigma_{\hat{L}\hat{L}} = \frac{0.0016}{4} \begin{pmatrix} 3 & -2 & -1 \\ -2 & 4 & -2 \\ -1 & -2 & 3 \end{pmatrix} + \frac{0.010}{16} \begin{pmatrix} 2 & 4 & 2 \\ 4 & 8 & 4 \\ 2 & 4 & 2 \end{pmatrix}$$

$$= \begin{pmatrix} -0.0012 & -0.0008 & -0.0004 \\ -0.0008 & 0.0016 & -0.0008 \\ -0.0004 & -0.0008 & 0.0012 \end{pmatrix}$$

$$+ \begin{pmatrix} 0.00125 & 0.0025 & 0.00125 \\ 0.0025 & 0.0050 & 0.0025 \\ 0.00125 & 0.0025 & 0.00125 \end{pmatrix}$$

$$= \begin{pmatrix} 0.00245 & 0.0017 & 0.00085 \\ 0.0017 & 0.0066 & 0.0017 \\ 0.00085 & 0.0017 & 0.00245 \end{pmatrix}$$

The uncertainty of the fixed control points, responsible for the second term, also dominates this expression. Furthermore, remembering that the covariance matrix of the observed quantities is

$$\Sigma = \begin{pmatrix} 0.0016 & 0 & 0 \\ 0 & 0.0032 & 0 \\ 0 & 0 & 0.0016 \end{pmatrix}$$

we see that the second term causes the covariance matrix of the adjusted observations to be larger than the covariance matrix of the actual observations.

Appendix C: Minimum Variance Adjustment (Gauss-Markov Theorem)

Consider the linear model

$$AX = L + V \tag{1}$$

in which the observations are unbiased and have covariance matrix Σ. We look for an estimate $\hat{X}$ of X that is

1. Best (in the sense of minimum variance), so that $\Sigma_{xx} = E[(\hat{x}-x)(\hat{x}-x)^T]$ is a minimum
2. Linear in the observations L, so that $\hat{x} = BL$ for some matrix B
3. Unbiased, so that $E[\hat{x}] = x$

We must define what we mean by minimizing a covariance matrix. Since there is no strict ordering of matrices, we must minimize some scaler measure of the matrix. A common choice is to minimize the trace $Tr\Sigma_{xx}$.

Since the observations are unbiased, $E[V] = 0$ and $E[L] = AX$. Then

$$E[\hat{\chi}] = E[BL] = BE[L] = BA\chi$$

and by the unbiased property, we must have $BAx = x$. Since this must hold irrespective of the value of X, we must have

$$BA - I = 0 \tag{23}$$

If there are u unknown parameters X, (23) represents u^2 separate equations. Let Λ be a matrix of u^2 Lagrange multipliers. Then

$$Tr[(BA - I)\Lambda]$$

represents the sum of all u^2 equations in (23), each multiplied by a Lagrange multiplier.

Furthermore, since $X = E[\hat{X}] = E[BL] = BE[L]$, we have

$$\hat{X} - X = BL - BE[L] = B(L - E[L])$$

so that

$$\Sigma_{xx} = E[(\hat{X} - X)(\hat{X} - X)^T]$$
$$= BE[(L - E[L])(L - E[L])^T]B^T = B\Sigma B^T \tag{24}$$

Now the problem is to minimize the augmented cost function

$$\Phi = Tr(B\Sigma B^T) + 2Tr[(BA - I)\Lambda] \tag{25}$$

This is done by differentiating (25) with respect to B and Λ, and setting each set of partial derivatives to zero. We get

$$\frac{\partial \Phi}{\partial B} = 0 \Rightarrow 2\left(\sum B^T + A\Lambda\right) = 0 \tag{26}$$

and

$$\frac{\partial \Phi}{\partial \Lambda} = 0 \Rightarrow BA - I = 0 \tag{27}$$

From (26) we obtain

$$B = -\Lambda^T A^T \Sigma^{-1}$$

and using (27)

$$BA = -\Lambda^T A^T \Sigma^{-1} A = I$$

so

$$\Lambda^T = -\left(A^T \Sigma^{-1} A\right)^{-1}$$

and

$$B = \left(A^T \Sigma^{-1} A\right)^{-1} A^T \Sigma^{-1}$$

Thus

$$\hat{X} = (A^T \Sigma^{-1} A)^{-1} A^T \Sigma^{-1} L \qquad (28)$$

is the best linear unbiased estimator. As a final modification, we can write $\Sigma^{-1} = (1/\sigma_0^2)\,W$ in (28). The two appearances of σ_0^2 cancel each other, yielding the familiar form

$$\hat{X} = (A^T W A)^{-1} A^T W L \qquad (29)$$