A SYNTHESIS OF RECENT ADVANCES IN THE METHOD OF LEAST SQUARES

E. J. KRAKIWSKY

May 1975

LECTURE NOTES
42

A SYNTHESIS OF RECENT ADVANCES IN THE METHOD OF LEAST SQUARES

Edward J. Krakiwsky©

Department of Geodesy and Geomatics Engineering
University of New Brunswick
P.O. Box 4400
Fredericton, N.B.
Canada
E3B 5A3

Groupe de Recherches de
Géodésie Spatiale
Toulouse, France*

May 1975

Reprinted August 1976 with Corrections
Latest Reprinting October 1994

*On sabbatical leave from the University of New Brunswick 1974/1975 as a National Research Council Exchange Scientist

PREFACE

In order to make our extensive series of lecture notes more readily available, we have scanned the old master copies and produced electronic versions in Portable Document Format. The quality of the images varies depending on the quality of the originals. The images have not been converted to searchable text.

TABLE OF CONTENTS

PREFACE

This synthesis has evolved out of lecture notes for a graduate course in the Department of Surveying Engineering, the University of New Brunswick, Fredericton, Canada. It is a culmination of work done in preparation for guest lectures on the subject given by the author during his sabbatical year at the following agencies and institutions: Centre National d'Etudes Spatiales (GRGS), Toulouse; Centre d'Etudes et de Recherches Geodynamiques et Astronomiques, Grasse; Technishe Universität München; Technishe Universität Hannover; Rijkscommissie voor Geodesie, Delf. The first draft of the synthesis was written in Toulouse and the final work was finished later in Fredericton.

Several individuals took special interest in this project and submitted invaluable comments on the draft; they are Prof. W. Baarda, Dr. G. Blaha, Mr. J. Kouba, Dr. C. Reigber, and Dr. S. Pavdé. The writing of this work was made easier through the useful discussions with Dr. P. Vaníček, Dr. K. Lambeck, M. M. Lefebvre, Dr. G. Balmino, Mr. D.B. Thomson, M. R. Blais, Mr. S. Grant, and Mr. C. Chamberlain.

Typing of the report was so ably done by Mme. G. Raphanel, Mrs. D. Smith, and Mrs. S. Biggar. Proof-reading assistance came from Mrs. I. Paim and Messrs. S. El-Hakim, A. Mutajwaa, S. John, J. Adams and M. Dyment.

Even though several individuals participated in various ways in this project, the author holds himself solely responsible for the correctness of the material presented herein.

E. J. Krakiwsky

List of Figures

List of Tables

1. INTRODUCTION

In 1795 Karl Friederich Gauss and Adrien Marie Legendre simultaneously, but independently, were the first to use the method of least squares [Gauss, 1963]. Over the years mathematicians and experimental scientists generalized and extended the original method, and made comparisons of these advances (see Table 1-1). The purpose of this work is to make a synthesis of these advances and studies.

The early advancements made to the original method itself* are those of Krüger [1905], Tobey [1930], and Tienstra [1956]. In his textbook, Tienstra explains that a least squares problem can be rigorously solved in "phases" (in parts) simply by treating the already estimated parameters and corrected observations of a previous phase as quasi-observations in the subsequent phase. Other authors have studied the Tienstra phase technique such as Lambeck [1962], Kouba [1970, 1972], Ying Chung-Chi [1970] and Nassar [1972].

A significant advance came from Kalman [1960]. He extended least squares by:

(i) adding a second model which allows the vector of parameters (state vector) to vary with time;

(ii) inserting a variance-covariance matrix on this secondary model;

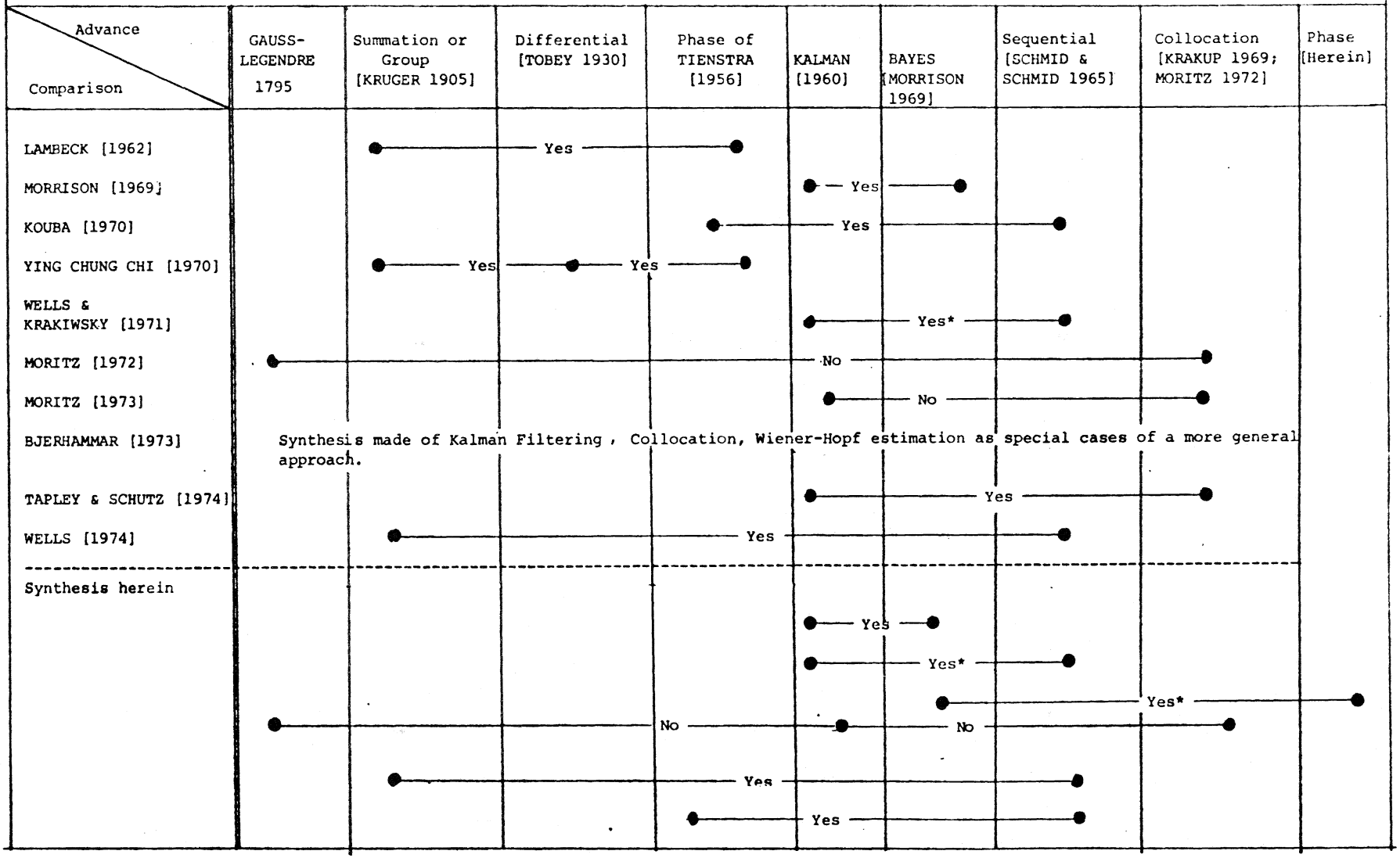
(iii) solving the problem in parts.

---

* This being quite different from the generalization made by Wiener [1949].

Table 1-1. ADVANCES IN LEAST SQUARES AND THEIR INTER-COMPARISONS.

("yes" indicates mathematical equivalence).



* Only true under the condition that the state vector does not vary with time.

It is because of this latter characteristic that the method has been called Kalman filtering. The following are some authors who have studied and employed this technique: Ott and Meder [1972]; Fail [1972]; Tapley [1972]; Tapley and Schutz [1974]. Bayes made a similar contribution and is called Bayes filtering in the literature, e.g. Morrison [1972].

Schmid and Schmid [1965] showed that the least squares problem could be so formulated such that all variates receive weights ranging from zero to infinity. They proposed the notion also, that a least squares problem could be rigorously solved "sequentially" (in parts). This is achieved by updating the original estimate by a corrective term, the latter of which is a function of a matrix inverse already computed in the course of obtaining the original estimate. The sequential procedure has been studied by such authors as: Uotila [1967]; Krakiwsky [1968]; and Kouba [1970].

The independent works of Krarup [1969; 1970] and Moritz [1972] lead to the modification of the original least squares. In this method called collocation, one not only solves for the parameters as before but also the signal - the second of two random components in the observables. The residuals (noise), the first random component in the observables, do not play a primary role in collocation as in the original least squares. Moritz [1973a] has incorporated the idea of computation in parts in collocation and calls this new advancement stepwise and sequential collocation. Several authors have studied and employed collocation, e.g. Schwarz [1973, 1974, 1975]; Moritz [1973]; Rapp [1974]; and Tscherning [1974, 1975].

Listed in Table 1-1 are some advances to the original least squares of Gauss-Legendre. Several authors have compared these advances. Lambeck [1962] has shown the equivalence between the group and Tienstra phase techniques. Morrison [1969] has shown that the Kalman and Bayes filters are mathematically equivalent, but are significantly different from the computational point of view. Kouba [1970] has shown the mathematical equivalence between the Tienstra phase and sequential techniques for the condition case (observations only) of geodetic adjustments. Ying Chung-Chi [1970] has shown that Tobey's "differential" and Krüger's "group" adjustments of geodesy are mathematically equivalent with the Tienstra phase technique; He works out numerical examples for the three methods. Wells and Krakiwsky [1971] show that if the concept of time variation in the parameters is deleted, the Kalman filter equations are completely identical to the sequential expressions. Moritz [1972] emphasizes fundamental differences between the original least squares and collocation, some of which have already been mentioned above. A comparison will be given later in this work. Bjerhammer [1973]* has made a complete generalization of least squares where one can find, as special cases, Kalman filtering, collocation and Wiener-Hopf estimation [Wiener 1949]. Tapley and Schutz [1974] make a comparison of the original least squares (including weighted parameters - a priori data) with collocation, via the Kalman filtering expressions. They conclude that the two corresponding sets of expressions (Kalman and collocation) for

---

* This work will not be examined in this synthesis as it is a generalization of the least squares method.

the solution vector are equivalent. In another comparison study, Wells [1974] shows the mathematical equivalence between the group method of Krüger (addition of normal equations) and the sequential approach.

Using the equivalences established immediately above, the reader may deduce equivalences. For example, using the works of Kouba [1970] and Ying Chung-Chi [1970] one can deduce that the group and sequential methods are mathematical equivalent. If one uses the works of Moritz [1972] and Tapley and Schutz [1974] one arrives at a direct contradiction for the comparison of the original least squares, Kalman filtering and collocation. This contradiction has become a favorite topic for debate by other scientists as well during meetings in recent years. It is because of this type of controversy that the author has been motivated to do this synthesis.

In the synthesis we derive all the methods using the same standard least squares methodology and use exclusively modern matrix and linear algebra. By means of this rather elementary, consistent, and logical approach, we strive to discover the similarities and differences between the various methods.

This synthesis takes the following form:

(1) the standard cases (parametric, condition, combined, weighted parameters) of the original least squares adjustment are reviewed (Section 2);

(2) the Kalman filtering equations are derived from original least squares principles using the standard methodology (Section 3);

(3) the Bayes filtering equations are shown to be equivalent to the Kalman filtering equations (Section 4);

(4) the sequential and phase expressions are deduced respectively from the Kalman and Bayes filtering expressions (Section 4);

(5) the Tienstra phase and addition of normal equation approaches are derived from basic least squares, and then shown to be mathematically equivalent to the sequential and phase approaches (Section 4);

(6) the collocation equations are derived using the standard combined case equations of adjustments (Section 5);

(7) the methods are analyzed from the fundamental and computational points of view (Section 6);

(8) finally, the findings of the synthesis are summarized (Section 7).

2. REVIEW OF THE STANDARD CASES OF LEAST SQUARES

In this section we review the methods parametric, condition, combined, and weighted parameters - the standard cases of least squares adjustment. These cases have been studied by numerous authors, e.g. Schmid and Schmid [1965]; Uotila [1967]; Wolf [1968]; Kouba [1970]; Wells and Krakiwsky [1971]; Vaníček and Wells [1972]. Derived below is a general scheme from which all the cases are deduced.

The review of this material will serve to acquaint the reader with the notation, terminology, and methodology which is essential for the derivation of the Kalman filtering equations (Section 3), the collocation equations (Section 5), and the analysis of the advances made to the least squares method (Section 6).

2.1 The Least Squares Problem

Least squares estimation is the standard method to obtain a unique set of values for a set of unknown parameters ($\bar{X}$) from a redundant set of observables ($\bar{L}$) through a known mathematical model (F).

Before we treat the general situation, let us describe the least squares problem for the linear-explicit case, that is

$$\bar{L} = F(\bar{X}) \quad , \tag{2-1}$$

$$\underset{nxl}{\bar{L}} = \underset{nxu}{A} \; \underset{uxl}{\bar{X}} \quad , \tag{2-2}$$

where n observables are related to u unknown parameters through a design matrix A. We know the observables have some unknown correction (residuals). Denoting these by V, and the observed value of the observables by L, equation 2-2 becomes

$$\underset{nxl}{L} + \underset{nxl}{V} = \underset{nxu}{A} \; \underset{uxl}{\bar{X}} \quad . \tag{2-3}$$

The least squares estimate for $\bar{X}$ is obtained under the condition that the quadratic form

$$\hat{V}^T P \hat{V} = \text{minimum}, \tag{2-4}$$

where the weight matrix

$$P = \sigma_o^2 \, \Sigma_L^{-1} \tag{2-5}$$

is related to the a priori variance factor (σ_o^2) and variance-covariance matrix of the observations. It is through the "help" of this condition that two equations, in addition to equation 2-3, are obtained, and thereby yielding least squares estimates for $\bar{X}$ and V - denoted by $\hat{\bar{X}}$ and $\hat{V}$, respectively.

We now consider the general situation where our mathematical model of r equations is implicit and non-linear, that is

$$F(\bar{X}, \bar{L}) = 0 \quad , \tag{2-6}$$

that accuracy estimates exist for the observations (Σ_L), and the parameters are treated as quasi-observables (Σ_x). The corresponding weight matrices are

$$P = \sigma_o^2 \Sigma_L^{-1} , \tag{2-7}$$

and

$$P_x = \sigma_o^2 \Sigma_x^{-1} . \tag{2-8}$$

The least squares estimates $\hat{X}$ and $\hat{V}$ for this general situation are obtained under the condition that

$$(\hat{V}^T P \hat{V} + \hat{X}^T P_x \hat{X}) = \text{minimum}, \tag{2-9}$$

where $\hat{X}$ are corrections to the parameters as explained immediatel below.

We chose to work with linear sets of equations thus we approximate our mathematical model (equation 2-6) by a linear Taylor series as follows:

$$F(\bar{X}, \bar{L}) = F(X, L) + \left.\frac{\partial F}{\partial \bar{X}}\right|_{X,L} (\bar{X}-X) + \left.\frac{\partial F}{\partial \bar{L}}\right|_{X,L} (\bar{L}-L) \tag{2-10}$$

$$= F(X, L) + \left.\frac{\partial F}{\partial \bar{X}}\right|_{X,L} \hat{X} + \left.\frac{\partial F}{\partial \bar{L}}\right|_{X,L} \hat{V} = 0 . \tag{2-10a}$$

The misclosure vector

$$\underset{rxl}{W} = F(X, L) \tag{2-11}$$

is the mathematical model evaluated with the quasi-observed values of the parameters (X) and the observed values of the observables (L). When F is evaluated with some approximate values of the parameters (X^o), we denote the misclosure vector as

$$W^o = F(X^o, L) . \tag{2-11a}$$

The first design matrix is

$$\underset{rxu}{A} = \left.\frac{\partial F}{\partial \bar{X}}\right|_{X,\ L} \tag{2-12}$$

and the second design matrix is

$$\underset{rxn}{B} = \left.\frac{\partial F}{\partial \bar{L}}\right|_{X,\ L} . \tag{2-13}$$

The final value (adjusted) of the vector of parameters is

$$\hat{\bar{X}} = X + \hat{X} \ . \tag{2-14}$$

The final value (adjusted) of the vector of observables is

$$\hat{\bar{L}} = L + \hat{V} \ . \tag{2-15}$$

The linearized mathematical model in symbolic form is

$$\underset{rxu\ uxl}{A\ \hat{X}} + \underset{rxn\ nxl}{B\ \hat{V}} + W = 0 \ , \tag{2-16}$$

where $\hat{X}$ and $\hat{V}$ are least squares estimates making $\hat{\bar{X}}$ and $\hat{\bar{L}}$ also least squares estimates.

2.2 The Least Squares Normal Equations

The least squares normal equations relating the unknown quantities $\hat{X}$ and $\hat{V}$ to the known quantities A, B, W, P and P_X is obtained from the variation function

$$\phi = \hat{V}^T P \hat{V} + \hat{X}^T P_X \hat{X} + 2\hat{K}^T (A\hat{X} + B\hat{V} + W) \ , \tag{2-17}$$

where the newly introduced unknown quantity $\hat{K}$ is the vector of r Lagrange correlates. To find the minimum of the two quadratic forms subject to the constraint function (linearized math model) is known as the extremal problem with constraints. The Lagrange method is the standard method of solving this problem.

First the derivatives of the variation function with respect to the variates $\hat{V}$ and $\hat{X}$ are taken and set equal to zero to determine the extremum, minimum in this case, namely

$$\frac{1}{2} \frac{\partial \phi}{\partial \hat{V}} = \hat{V}^T P + \hat{K}^T B = 0 \ , \tag{2-18}$$

$$\frac{1}{2}\frac{\partial \phi}{\partial \hat{X}} = \hat{X}^T P_X + \hat{K}^T A = 0 \ . \qquad (2\text{-}19)$$

The transpose of the above two equations and the linearized mathematical model constitute the three equations of the least squares normal equation system:

$$\begin{matrix} P & \hat{V} & + B^T & \hat{K} & = 0 \ , \\ nxn & nxl & nxr & rxl & \end{matrix}$$

$$\begin{matrix} P_X & \hat{X} & + A^T & \hat{K} & = 0 \ , \\ uxu & uxl & uxr & rxl & \end{matrix} \qquad (2\text{-}20)$$

$$\begin{matrix} A & \hat{X} & + B & \hat{V} & + W & = 0 \ . \\ rxu & uxl & rxn & nxl & rxl & \end{matrix}$$

The most expanded form of the least squares normal equation system in block matrix form is

$$\begin{bmatrix} P & B^T & 0 \\ B & 0 & A \\ 0 & A^T & P_X \end{bmatrix} \begin{bmatrix} \hat{V} \\ \hat{K} \\ \hat{X} \end{bmatrix} + \begin{bmatrix} 0 \\ W \\ 0 \end{bmatrix} = 0 \ , \qquad (2\text{-}21)$$

with a coefficient matrix of dimensions n+r+u. A solution for the vector comprising $\hat{V}$, $\hat{K}$, and $\hat{X}$ is possible by directly inverting the coefficient matrix. This is not efficient, thus a normal equation system is derived where the inversions are smaller.

We use a special elimination technique [e.g. Thompson 1969; Wells and Krakiwsky 1971]. Given a matrix equation system

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] \left[\begin{array}{c} X \\ \hline Y \end{array}\right] + \left[\begin{array}{c} U \\ \hline V \end{array}\right] = 0, \qquad (2\text{-}22)$$

where A, B, C and D constitute the known coefficient matrix, X and Y the unknown vector, and U and V the known vector, X is eliminated by forming a modified coefficient matrix and known vector as follows:

$$[D - CA^{-1} B]Y + [V - CA^{-1} U] = 0 \qquad (2\text{-}23)$$

(A must be non-singular). The proof is simple and left to the reader as an exercise.

We return to the problem at hand and first eliminate $\hat{V}$ from equation 2-21, where D becomes the lower two by two hyper-matrix of equation 2-21.

$$\left\{\begin{bmatrix} 0 & A \\ A^T & P_x \end{bmatrix} - \begin{bmatrix} B \\ 0 \end{bmatrix} P^{-1} \begin{bmatrix} B^T & 0 \end{bmatrix}\right\} \begin{bmatrix} \hat{K} \\ \hat{X} \end{bmatrix} + \left\{\begin{bmatrix} W \\ 0 \end{bmatrix} - \begin{bmatrix} B \\ 0 \end{bmatrix} P^{-1} \begin{bmatrix} 0 \end{bmatrix}\right\} = 0, \qquad (2\text{-}24)$$

$$\begin{bmatrix} -BP^{-1}B^T & A \\ A^T & P_x \end{bmatrix} \begin{bmatrix} \hat{K} \\ \hat{X} \end{bmatrix} + \begin{bmatrix} W \\ 0 \end{bmatrix} = 0 \ . \qquad (2\text{-}25)$$

Then eliminate $\hat{K}$ using the same technique:

$$[P_x - A^T (-BP^{-1}B^T)^{-1} A]\hat{X} + 0 - A^T (-BP^{-1}B^T)^{-1} W = 0, \qquad (2\text{-}26)$$

$$[A^T (BP^{-1} B^T)^{-1} A + P_x] \hat{X} + A^T (BP^{-1} B^T)^{-1} W = 0 \ , \qquad (2\text{-}27)$$

and

$$\boxed{\hat{X} = -[A^T (BP^{-1}B^T)^{-1} A + P_x]^{-1} A^T (BP^{-1}B^T)^{-1} W \ .} \qquad (2\text{-}28)$$

The solution for $\hat{K}$ is made using the first expression from equation system 2-25, namely

$$-(BP^{-1}B^T)\hat{K} + A\hat{X} + W = 0 \ ,$$

$$\boxed{\hat{K} = (BP^{-1}B^T)^{-1}(A\hat{X} + W) \ .} \qquad (2\text{-}29)$$

The solution for $\hat{V}$ is made using the first expression from equation system 2-21, namely,

$$P\hat{V} + B^T\hat{K} = 0 \ ,$$

$$\boxed{\hat{V} = -P^{-1}B^T\hat{K}.} \qquad (2\text{-}30)$$

Equations 2-28, 2-29 and 2-30 represent the alternative solution to the least squares normal equation 2-21.

2.3 Derivation of Variance-Covariance Matrices

In this section we derive the variance-covariance matrices for the residual vector $\hat{V}$, the parameter vector $\hat{X}$, the final value of the parameter vector $\hat{\bar{X}}$, and the final value of the observable vector $\hat{\bar{L}}$.

We make extensive use of the covariance law which states that, given a functional relationship

$$Y = F(Z) \qquad (2\text{-}31)$$

between two random vectors Y and Z along with the variance-covariance matrix of Z (Σ_Z), the variance-covariance matrix of Y is given by

$$\boxed{\Sigma_Y = \left(\frac{\partial F}{\partial Z}\right) \Sigma_Z \left(\frac{\partial F}{\partial Z}\right)^T \ .} \qquad (2\text{-}32)$$

Dividing both sides of the above equation by the a priori variance factor (σ_o^2) we have the covariance law in terms of weight coefficients:

$$\boxed{Q_Y = \left(\frac{\partial F}{\partial Z}\right) Q_Z \left(\frac{\partial F}{\partial Z}\right)^T \ .} \qquad (2\text{-}33)$$

Note the following relationship between the variance-covariance (Σ), weight (P) and weight coefficient (Q) matrices:

$$P = \sigma_o^2 \Sigma^{-1} = Q^{-1}$$

or

$$P^{-1} = \frac{1}{\sigma_o^2} \Sigma = Q \ .$$

Since we have formulated the least squares problem in terms of weights (P) we will be consistent and use the covariance law in terms of weight coefficients (inverse of weights) instead of variances-covariances.

Variance-Covariance Matrix for $\hat{\bar{X}}$

We follow Krakiwsky [1968], Kouba [1970] and Wells and Krakiwsky [1971] in deriving the variance-covariance matrix for the final value of the parameters $\hat{\bar{X}}$.

According to equations 2-14 and 2-28

$$\hat{\bar{X}} = X + \hat{X} \ , \tag{2-34}$$

$$\hat{\bar{X}} = X - [A^T M^{-1} A + P_X]^{-1} A^T M^{-1} W, \tag{2-35}$$

where we have let

$$M = BP^{-1} B^T \ . \tag{2-36}$$

Recalling equation 2-11,

$$W = F(X, L), \tag{2-37}$$

we see that $\hat{\bar{X}}$ is a function of two independent random variables - the a priori estimate of the parameters (quasi-observable) (X), and the observed value of the observables (L).

Applying the covariance law to equation 2-34 yields

(assuming henceforth that X and L are statistically independent)

$$Q_{\hat{x}} = \left(\frac{\partial \hat{\bar{X}}}{\partial X}\right) Q_x \left(\frac{\partial \hat{\bar{X}}}{\partial X}\right)^T + \left(\frac{\partial \hat{\bar{X}}}{\partial L}\right) Q_L \left(\frac{\partial \hat{\bar{X}}}{\partial L}\right)^T , \tag{2-38}$$

$$Q_{\hat{x}} = \left(\frac{\partial \hat{\bar{X}}}{\partial X}\right) P_x^{-1} \left(\frac{\partial \hat{\bar{X}}}{\partial X}\right)^T + \left(\frac{\partial \hat{\bar{X}}}{\partial L}\right) P^{-1} \left(\frac{\partial \hat{\bar{X}}}{\partial L}\right)^T . \tag{2-39}$$

$$\frac{\partial \hat{\bar{X}}}{\partial X} = I - [A^T M^{-1} A + P_x]^{-1} A^T M^{-1} A \tag{2-40}$$

since from equation 2-12

$$\frac{\partial W}{\partial X} = \frac{\partial F(X, L)}{\partial X} = A . \tag{2-41}$$

$$\frac{\partial \hat{\bar{X}}}{\partial L} = -[A^T M^{-1} A + P_x]^{-1} A^T M^{-1} B \tag{2-42}$$

since from equation 2-13

$$\frac{\partial W}{\partial L} = \frac{\partial F(X, L)}{\partial L} = B . \tag{2-43}$$

Before proceeding further with the derivation, it will prove useful to derive the weight coefficient matrix for $\hat{X}$ (equation 2-28), and to do this, we first need the weight coefficient matrix of W.

Applying the covariance law to equation 2-37 yields

$$Q_W = \left(\frac{\partial F}{\partial X}\right) Q_x \left(\frac{\partial F}{\partial X}\right)^T + \left(\frac{\partial F}{\partial L}\right) Q_L \left(\frac{\partial F}{\partial L}\right)^T , \tag{2-44}$$

$$\boxed{Q_W = A P_x^{-1} A^T + B P^{-1} B^T .} \tag{2-45}$$

From equation 2-35

$$\hat{X} = - [A^T M^{-1} A + P_x]^{-1} A^T M^{-1} W . \tag{2-46}$$

Applying the covariance law to the above equation and taking into account equation 2-45

$$Q_{\hat{X}} = \left(\frac{\partial \hat{X}}{\partial W}\right) Q_W \left(\frac{\partial \hat{X}}{\partial W}\right)^T \tag{2-47}$$

$$- \{-[A^T M^{-1} A + P_X]^{-1} A^T M^{-1}\}\{A P_X^{-1} A^T + B P^{-1} B^T\}\{-[A^T M^{-1} A + P_X]^{-1} A^T M^{-1}\}^T \tag{2-48}$$

$$Q_{\hat{X}} = [A^T M^{-1} A + P_X]^{-1} A^T M^{-1} (A P_X^{-1} A^T + M) M^{-1} A [A^T M^{-1} A + P_X]^{-1} \tag{2-49}$$

and letting

$$N = A^T M^{-1} A , \tag{2-50}$$

$$Q_{\hat{X}} = [N + P_X]^{-1} A^T M^{-1} (A P_X^{-1} A^T + M) M^{-1} A [N + P_X]^{-1} \tag{2-51}$$

and expanding terms

$$\boxed{Q_{\hat{X}} = [N+P_X]^{-1} N P_X^{-1} N [N+P_X]^{-1} + [N+P_X]^{-1} N [N+P_X]^{-1}.} \tag{2-52}$$

We now turn to the task of determining $Q_{\hat{\tilde{X}}}$. Substituting equations 2-40 and 2-42 into 2-39 and using 2-50, we find

$$\begin{aligned} Q_{\hat{\tilde{X}}} &= \{I - [N+P_X]^{-1} N\} P_X^{-1} \{I - [N+P_X]^{-1} N\}^T \\ &\quad + \{-[N+P_X]^{-1} A^T M^{-1} B\} P^{-1} \{-[N+P_X]^{-1} A^T M^{-1} B\}^T \\ &= \{I - [N+P_X]^{-1} N\} P_X^{-1} \{I - N [N+P_X]^{-1}\} \\ &\quad + [N+P_X]^{-1} A^T M^{-1} B\, P^{-1} B^T M^{-1} A [N+P_X]^{-1} \qquad (2\text{-}53) \\ &= P_X^{-1} - P_X^{-1} N [N+P_X]^{-1} - [N+P_X]^{-1} N P_X^{-1} \\ &\quad + [N+P_X]^{-1} N P_X^{-1} N [N+P_X]^{-1} + [N+P_X]^{-1} N [N+P_X]^{-1}. \qquad (2\text{-}54) \end{aligned}$$

Noting that the last two terms of the above equation are identical to $Q_{\hat{X}}$ (equation 2-52), we can write

$$\boxed{Q_{\hat{\tilde{X}}} = P_X^{-1} + Q_{\hat{X}} - P_X^{-1} N [N+P_X]^{-1} - [N+P_X]^{-1} N P_X^{-1} .} \tag{2-55}$$

The above expression can be shown to be equivalent to the inverse of the coefficient matrix of the normal equation system 2-28

for $\hat{X}$, that is

$$Q_{\hat{x}} = [A^T(BP^{-1}B^T)^{-1} A + P_x]^{-1} , \tag{2-56}$$

$$= [A^T M^{-1} A + P_x]^{-1} , \tag{2-57}$$

$$\boxed{Q_{\hat{x}} = [N + P_x]^{-1} .} \tag{2-58}$$

To prove this we begin by factoring-out the term

$$[N + P_x]^{-1}$$

from equation 2-54, then multiply terms, and finally cancel like terms. The details are the following:

$$Q_{\hat{x}} = [N+P_x]^{-1}\{[N+P_x]P_x^{-1} - [N+P_x]P_x^{-1} N[N+P_x]^{-1} - NP_x^{-1}$$

$$+ NP_x^{-1}N [N+P_x]^{-1} + N[N+P_x]^{-1}\} ,$$

$$= [N+P_x]^{-1} \{NP_x^{-1}+I-NP_x^{-1}N[N+P_x]^{-1} - N[N+P_x]^{-1}$$

$$- NP_x^{-1} + NP_x^{-1}N [N+P_x]^{-1} + N[N+P_x]^{-1}\},$$

$$= [N+P_x]^{-1}I = [N+P_x]^{-1} .$$

Finally, we write the variance-covariance matrix as

$$\boxed{\Sigma_{\hat{x}} = \sigma_o^2 Q_{\hat{x}} ,} \tag{2-59}$$

and the estimated variance-covariance matrix as

$$\boxed{\hat{\Sigma}_{\hat{x}} = \hat{\sigma}_o^2 Q_{\hat{x}} ,} \tag{2-60}$$

where the estimated variance factor

$$\boxed{\hat{\sigma}_o^2 = \frac{\hat{v}^T P \hat{v} + \hat{X}^T P_x \hat{X}}{\nu} .} \tag{2-61}$$

The degrees of freedom*

$$\boxed{\nu = r - u + u_x ,} \tag{2-62}$$

* equation 2-62 is only an approximation, see Bossler [1972] for a complete and rigorous treatment.

where r is the number of equations in F, u the number of parameters to be estimated, and u_x the number of parameters weighted. The proof of equation 2-61 is beyond the scope of this work, see for example Hamilton [1964], Wells and Krakiwsky [1971].

Variance-Covariance Matrix for $\hat{\underline{L}}$

We begin from the definition of the final (adjusted) observables (equation 2-15),

$$\hat{\underline{L}} = \underline{L} + \hat{\underline{V}} \ . \tag{2-63}$$

Using equations 2-29 and 2-30

$$\hat{\underline{L}} = \underline{L} - P^{-1}B^T\hat{K} \tag{2-64}$$

$$= \underline{L} - P^{-1}B^TM^{-1}(A\hat{X} + W) \tag{2-65}$$

$$= \underline{L}-P^{-1}B^TM^{-1}A\hat{X}-P^{-1}B^TM^{-1}W \ , \tag{2-66}$$

and after using equation 2-46

$$\hat{\underline{L}} = \underline{L}+P^{-1}B^TM^{-1}A\ [A^TM^{-1}A + P_x]^{-1}A^TM^{-1}W - P^{-1}B^TM^{-1}\ W \ . \tag{2-67}$$

Applying the covariance law to the above equation yields

$$Q_{\hat{\underline{L}}} = (\frac{\partial\hat{\underline{L}}}{\partial\underline{L}})Q_L\ (\frac{\partial\hat{\underline{L}}}{\partial\underline{L}})^T + (\frac{\partial\hat{\underline{L}}}{\partial X})\ Q_x(\frac{\partial\hat{\underline{L}}}{\partial X})^T \ , \tag{2-68}$$

where

$$\frac{\partial\hat{\underline{L}}}{\partial\underline{L}} = I+P^{-1}B^TM^{-1}A\ [A^TM^{-1}A+P_x]^{-1}\ A^TM^{-1}B-P^{-1}B^TM^{-1}\ B \ , \tag{2-69}$$

$$\frac{\partial\hat{\underline{L}}}{\partial X} = P^{-1}B^TM^{-1}A\ [A^TM^{-1}A+P_x]^{-1}\ A^TM^{-1}A - P^{-1}B^TM^{-1}\ A \ . \tag{2-70}$$

Substituting equations 2-69 and 2-70 in 2-68, noting equation 2-52, and collecting terms we get

$$
\begin{aligned}
Q_{\hat{L}} = {} & P^{-1} + P^{-1}B^T M^{-1} A\, Q_{\hat{x}}\, A^T M^{-1} BP^{-1} \\
& + P^{-1}B^T M^{-1}\, A\, P_x^{-1}\, A^T M^{-1}\, BP^{-1} \\
& - P^{-1}\, B^T M^{-1}\, A(A^T M^{-1} A + P_x)^{-1}\, A^T M^{-1} AP_x^{-1}\, A^T M^{-1} BP^{-1} \\
& - P^{-1} B^T M^{-1} AP_x^{-1}\, A^T M^{-1}\, A(A^T M^{-1} A + P_x)^{-1}\, A^T M^{-1} BP^{-1} \\
& - P^{-1} B^T M^{-1} BP^{-1} \quad . \qquad (2\text{-}71)
\end{aligned}
$$

Using equation 2-55, the above equation can be written in terms of $Q_{\hat{x}}$, namely

$$
\boxed{Q_{\hat{L}} = P^{-1} + P^{-1}B^T M^{-1}\, A\, Q_{\hat{x}}\, A^T M^{-1} BP^{-1} - P^{-1}B^T M^{-1} BP^{-1} \quad .} \qquad (2\text{-}72)
$$

Weight and Coefficient Matrix for $\hat{V}$

Expressing $\hat{V}$ as a function of W (analogous to equation 2-35), and applying the covariance law, we get after considering equation 2-58 that

$$
\boxed{
\begin{aligned}
Q_{\hat{V}} &= P^{-1} B^T M^{-1} BP^{-1} - P^{-1} B^T M^{-1}\, A\, Q_{\hat{x}}\, A^T M^{-1} BP^{-1} \quad , \\
&= Q_L - Q_{\hat{L}} \quad .
\end{aligned}
} \qquad (2\text{-}73)
$$

The above equation corresponds to common sense, namely the variances of the observables after the adjustment are smaller than the variances of the observables (observations) before the adjustment since

$$Q_{\hat{L}} = Q_L - Q_{\hat{V}} \quad .$$

See equations 2-121 and 2-122 as an example.

2.4 Generation of the Standard Cases

Combined Case with Weighted Parameters (A, B, P, $P_x \neq 0$)

The general case of a non-linear implicit model with weighted parameters is known as the combined case with weighted parameters. It has a solution given by the following equations (2-28, 2-14, 2-29, 2-30, 2-45, 2-52, 2-55, 2-15, 2-61, 2-62, 2-60, 2-72, 2-73):

$$\hat{X} = -[A^T (BP^{-1}B^T)^{-1} A + P_x]^{-1} A^T (BP^{-1}B^T)^{-1} W$$

$$= - [N + P_x]^{-1} A^T M^{-1} W \qquad (2\text{-}74)$$

$$= - [N + P_x]^{-1} U , \qquad (2\text{-}75)$$

$$\hat{\bar{X}} = X + \hat{X} , \qquad (2\text{-}76)$$

$$\hat{K} = (BP^{-1} B^T)^{-1} (A\hat{X} + W) \qquad (2\text{-}77)$$

$$= M^{-1} (A\hat{X} + W) , \qquad (2\text{-}78)$$

$$\hat{V} = - P^{-1} B^T \hat{K} , \qquad (2\text{-}79)$$

$$\hat{\bar{L}} = L + \hat{V} , \qquad (2\text{-}80)$$

$$Q_{\hat{x}} = [A^T(BP^{-1}B^T)^{-1}A+P_x]^{-1}A^T(BP^{-1}B^T)^{-1}AP_x^{-1}A^T(BP^{-1}B^T)^{-1}A$$
$$[A^T(BP^{-1}B^T)^{-1}A+P_x]^{-1}$$
$$+ [A^T(BP^{-1}B^T)^{-1} A+P_x]^{-1} N[A^T(BP^{-1}B^T)^{-1}A + P_x]^{-1} \qquad (2\text{-}81)$$

$$= [N+P_x]^{-1} NP_x^{-1}N[N+P_x]^{-1} + [N+P_x]^{-1} N[N+P_x]^{-1} , \qquad (2\text{-}82)$$

$$Q_{\hat{\underline{x}}} = P_x^{-1} + Q_{\hat{x}} - P_x^{-1}A^T(BP^{-1}B^T)^{-1}A[A^T(BP^{-1}B^T)^{-1}A + P_x]^{-1}$$

$$- [A^T(BP^{-1}B^T)^{-1}A + P_x]^{-1} A^T(BP^{-1}B^T)^{-1} A P_x^{-1} \quad (2\text{-}83)$$

$$= P_x^{-1} + Q_{\hat{x}} - P_x^{-1} N[N + P_x]^{-1} - [N + P_x]^{-1}NP_x^{-1} \quad (2\text{-}84)$$

$$= [A^T (BP^{-1}B^T)^{-1} A + P_x]^{-1} = [N + P_x]^{-1} , \quad (2\text{-}85)$$

$$Q_{\hat{\underline{L}}} = P^{-1} + P^{-1}B^TM^{-1} AQ_{\hat{\underline{x}}}A^TM^{-1}BP^{-1} - P^{-1}B^TM^{-1}BP^{-1}, \quad (2\text{-}86)$$

$$Q_{\hat{V}} = P^{-1}B^TM^{-1}BP^{-1} - P^{-1}B^TM^{-1}A\, Q_{\hat{\underline{x}}}\, A^T\, M^{-1}B\, P^{-1} , \quad (2\text{-}87)$$

$$Q_W = A\, P_x^{-1}\, A^T + BP^{-1}\, B^T , \quad (2\text{-}88)$$

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T\, P\, \hat{V} + \hat{X}^T P_x\, \hat{X}}{\nu} , \quad (2\text{-}89)$$

$$\nu = r - u + u_x , \quad (2\text{-}90)$$

$$\Sigma_{\hat{x}} = \sigma_o^2\, Q_{\hat{x}} , \quad (2\text{-}91)$$

$$\Sigma_{\hat{\underline{x}}} = \sigma_o^2\, Q_{\hat{\underline{x}}} , \quad (2\text{-}92)$$

$$\Sigma_{\hat{\underline{L}}} = \sigma_o^2\, Q_{\hat{\underline{L}}} , \quad (2\text{-}93)$$

$$\Sigma_{\hat{V}} = \sigma_o^2\, Q_{\hat{V}} , \quad (2\text{-}94)$$

$$\Sigma_W = \sigma_o^2\, Q_W . \quad (2\text{-}94a)$$

In the above expressions

$$M = BP^{-1}\, B^T , \quad (2\text{-}95)$$

$$N = A^T\, (BP^{-1}B^T)^{-1}\, A = A^TM^{-1}A , \quad (2\text{-}96)$$

$$U = A^T\, (BP^{-1}B^T)^{-1}\, W = A^TM^{-1}W . \quad (2\text{-}97)$$

An intriguing mathematical fact concerning the variance-covariance matrices for the parameters is that the inverse of the coefficient matrix of the normal equations is the weight coefficient matrix of the final (adjusted) parameters $\hat{\hat{X}}$ and not of the solution vector $\hat{X}$ [Kouba 1970].

If all the parameters are weighted, note that the number of degrees of freedom becomes equal to the number of equations. This is analogous to the condition case below in which all observables are weighted. This is not surprising, for in the present case all quantities are also weighted. Schmid and Schmid [1965] call this "generalized least squares".

One should also note that the variance-covariance matrix of W is defined with the a priori variance factor σ_o^2 which allows statistical testing before the adjustment takes place, if σ_o^2 is indeed known. In the case that σ_o^2 is not known, then an estimate may be obtained from the adjustment itself. Hamilton [1964] shows that in the latter case (σ_o^2 not known), the confidence region for the adjusted parameters ($\hat{\hat{X}}$) are given in terms of the Fischer distribution, while if σ_o^2 is known, the confidence region is described through the multivariate Chi-squared distribution.

Combined Case (A, B, P, P_x = 0)

The combined case is characterized by a non-linear implicit mathematical model with no weights on the parameters. We deduce the corresponding set of expressions from the general case by considering that if there are no weights then P_x is

equal to zero. This implies that X is a constant vector (now denoted by X^o), and its variance covariance matrix Σ_{x^o} does not exist. As a consequence, both P_x and P_x^{-1} (Q_x) are null matrices. Also note that the partial derivatives of $\hat{\bar{X}}$ with respect to X^o will also be a null matrix. Upon substitution of the three null matrices into equations 2-9 through 2-73, we get the desired results:

$$\hat{X} = - [A^T(BP^{-1}B^T)^{-1}A]^{-1} A^T(BP^{-1} B^T)^{-1} W^o \quad (2\text{-}98)^*$$

$$= - N^{-1} U \ , \quad (2\text{-}99)$$

$$\hat{\bar{X}} = X^o + \hat{X} \ , \quad (2\text{-}100)$$

$$\hat{K} = (BP^{-1}B^T)^{-1} (A\hat{X} + \overset{o}{W}) \quad (2\text{-}101)$$

$$= M^{-1} (A\hat{X} + \overset{o}{W}) \ , \quad (2\text{-}102)$$

$$\hat{V} = - P^{-1} B^T \hat{K} \ , \quad (2\text{-}103)$$

$$\hat{\bar{L}} = L + \hat{V} \ , \quad (2\text{-}104)$$

$$Q_{\hat{X}} = N^{-1} = [A^T (BP^{-1} B^T)^{-1} A]^{-1} = Q_{\hat{\bar{X}}} \ , \quad (2\text{-}105)$$

$$Q_{\hat{\bar{L}}} = P^{-1} + P^{-1}B^TM^{-1} A N^{-1}A^TM^{-1}BP^{-1} - P^{-1}B^TM^{-1}BP^{-1} \ , \quad (2\text{-}106)$$

$$Q_{\hat{V}} = P^{-1}B^TM^{-1}BP^{-1} - P^{-1}B^TM^{-1} A N^{-1}A^TM^{-1}BP^{-1} \ , \quad (2\text{-}107)$$

$$Q_{W^o} = BP^{-1} B^T \ , \quad (2\text{-}108)$$

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{\nu} \ , \quad (2\text{-}109)$$

$$\nu = r - u \ , \quad (2\text{-}110)$$

---

* $W^o = F\ (X^o, L)$

$$\Sigma_{\hat{x}} = \Sigma_{\hat{\bar{x}}} = \sigma_o^2 Q_x \ , \tag{2-111}$$

$$\Sigma_{\hat{\bar{L}}} = \sigma_o^2 Q_{\hat{\bar{L}}} \tag{2-112}$$

$$\Sigma_{\hat{V}} = \sigma_o^2 Q_{\hat{V}} \ , \tag{2-113}$$

$$\Sigma_{W^o} = \sigma_o^2 Q_{W^o} . \tag{2-114}$$

We witness that the weight coefficient matrix of the correction vector $\hat{X}$ and adjusted vector $\hat{\bar{X}}$ are identical and equal to the inverse of the coefficient matrix of the normal equations. The degrees of freedom is calculated as the difference between the number of equations and the number of unknown parameters.

Parametric Case (A, B = -I, P, P_x = 0)

The parametric case is characterized by a non-linear explicit model. This means that the observables can be explicitly expressed as some non-linear function of the parameters thus the reason for the second design matrix be to be equal to a minus identity matrix. Setting B equal to -I in the combined case with no weights on the parameters we get the following expressions:

$$\hat{X} = -[A^T P A]^{-1} A^T P W^o \ , \tag{2-115}$$

$$\hat{\bar{X}} = X^o + \hat{X} \ , \tag{2-116}$$

$$\hat{K} = P(A\hat{X} + W^o) \ , \tag{2-117}$$

$$\hat{V} = P^{-1} \hat{K} \ , \tag{2-118}$$

$$\hat{\bar{L}} = L + \hat{V} \ , \tag{2-119}$$

$$Q_{\hat{x}} = Q_{\hat{\underline{x}}} = [A^T P A]^{-1} , \tag{2-120}$$

$$Q_{\hat{\underline{L}}} = A\,[A^T P A]^{-1}\,A^T , \tag{2-121}$$

$$Q_{\hat{V}} = P^{-1} - A\,[A^T P A]^{-1}\,A^T , \tag{2-122}$$

$$Q_{W^o} = P^{-1} , \tag{2-123}$$

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{\nu} , \tag{2-124}$$

$$\nu = r - u = n - u , \tag{2-125}$$

$$\Sigma_{\hat{x}} = \Sigma_{\hat{\underline{x}}} = \sigma_o^2\, Q_{\hat{x}} , \tag{2-126}$$

$$\Sigma_{\hat{\underline{L}}} = \sigma_o^2\, Q_{\hat{\underline{L}}} , \tag{2-127}$$

$$\Sigma_{\hat{V}} = \sigma_o^2\, Q_{\hat{V}} , \tag{2-128}$$

$$\Sigma_{W^o} = \sigma_o^2\, Q_{W^o} . \tag{2-129}$$

Note that the number of equations equals the number of observations; this is not true for the combined case. We see that the weight coefficient matrix of the adjusted observables has the form of a propagation of errors (covariance law) from the adjusted parameters into these quantities.

Condition Case (A = 0, B, P, P_x = 0)

The condition case is characterized by a non-linear model consisting of only observables, thus the first design matrix A in the combined case with no weights on the parameters vanishes, yielding:

$$\hat{K} = (B P^{-1} B^T)^{-1} W \tag{2-130}$$

$$= M^{-1} W , \tag{2-131}$$

$$\hat{V} = - P^{-1} B^T \hat{K} , \tag{2-132}$$

$$\hat{\underline{L}} = L + \hat{V} , \tag{2-133}$$

$$Q_{\hat{\underline{L}}} = P^{-1} - P^{-1} B^T (BP^{-1}B^T)^{-1} BP^{-1} , \tag{2-134}$$

$$Q_{\hat{\underline{V}}} = P^{-1} B^T M^{-1} BP^{-1} , \tag{2-135}$$

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{\nu} , \tag{2-136}$$

$$\nu = r , \tag{2-137}$$

$$\Sigma_{\hat{\underline{L}}} = \sigma_o^2 Q_{\hat{\underline{L}}} , \tag{2-138}$$

$$\Sigma_{\hat{V}} = \sigma_o^2 Q_{\hat{V}} . \tag{2-139}$$

The above case can be used to solve the combined case with weighted parameters as follows: given

$$B^* V^* + W = 0, P^* \tag{2-140}$$

where

$$B^* = [A \mid B] , \tag{2-141}$$

$$V^* = \begin{bmatrix} \hat{X} \\ \hat{V} \end{bmatrix} \tag{2-142}$$

$$P^* = \begin{bmatrix} P_x & 0 \\ 0 & P \end{bmatrix} , \tag{2-143}$$

then substitute these definitions of B^* and P^* into the equations for the condition case immediately above. Note that the design matrix A and weight matrix P_x both pertain to the weighted parameters.

This terminates the review of the standard cases of least squares adjustments. We draw heavily on these equations in the subsequent sections.

3. DERIVATION OF THE KALMAN FILTER EQUATIONS

One of the highlights of this synthesis is the derivation of the Kalman filter equations using the same conventional methodology employed in the derivation of the standard cases of least squares adjustment in the previous section. As a result of using the same methodology, we have ready at our finger tips a clear and relatively elementary manner for comparing the Kalman filter equations with the standard cases, and with other advances made in the least squares method. Our derivation below follows closely that given in [Grant and Krakiwsky 1974]. The reader can find alternative derivations in [Morrison 1969; Moritz 1973a], while the original derivation is given in [Kalman 1960].

3.1 Definition of the Problem

The standard least squares problem of the previous section differs from the present one in two ways. Firstly, we consider the vector of parameters (state vector)* as varying with time. Consequently we must group the observations according to the same time epoch because of this variation. Below we introduce the Kalman

---

* Kalman filter terminology.

Two satellite passes

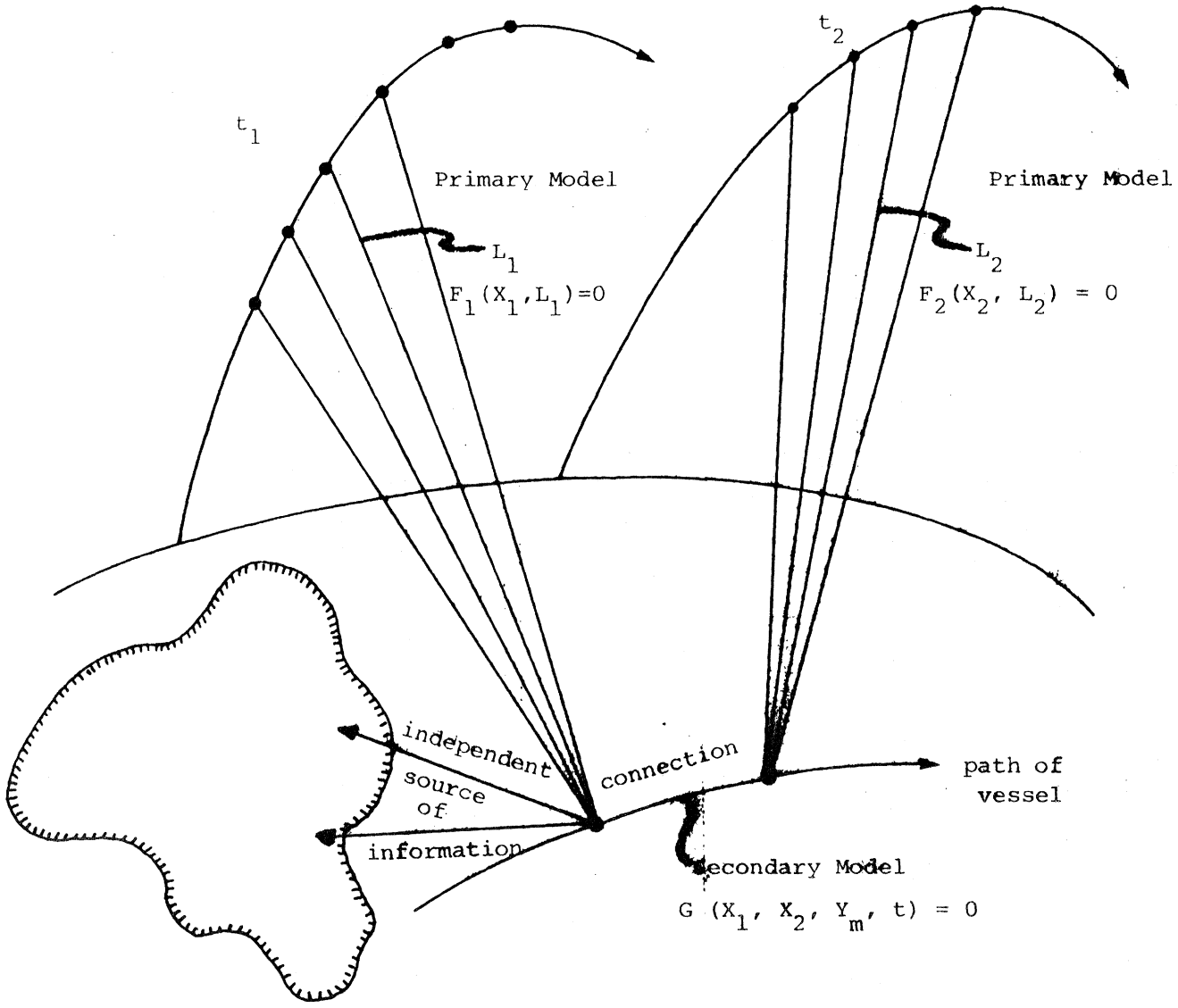


Figure 3-1. Four Sources of Information in Kalman Filtering

- Land determination of X_1 at t_1;
- Satellite determination of X_1 at t_1;
- Navigation determination of X_2 relative to X_1;
- Satellite determination of X_2 at t_2.

filtering problem via an obvious example in which the coordinates (state vector) of a moving point vary with time. There are other problems in which it is not necessary to have"physical movements" to induce a variation in the state vector, that is, simply the passage of time is sufficient to change the state vector.

Imagine a moving vessel seeking to accurately position itself (Figure 3-1). One can determine its coordinates at t_1 from measurements made to land (X_1 and Σ_{X_1}). Still at time t_1, observations L_1 can be made to a passing satellite and the determination of X_1 made from the model

$$F_1 (X_1, L_1) = 0 . \qquad (3\text{-}1)$$

Navigation devices on board recording velocity can serve to determine the position X_2 at t_2 relative to X_1 from a function

$$G (X_1, X_2, t) = 0 . \qquad (3\text{-}2)$$

Still at time t_2 observations L_2 can be made to another passing satellite and the determination made through the model

$$F_2 (X_2, L_2) = 0 . \qquad (3\text{-}3)$$

There is one characteristic in common to all these four determinations; they are independent from one another. The problem of rigorously combining all four sources of information in one mathematical system is in essence what the Kalman filter does. Variance and covariances are rigorously propagated from one stage to the next, and at any given time, the solution (state vector) contains all the information up-to and including the present. You can even "predict" the value of state vector since you have information on the changing of the parameters with time.

The Kalman filtering problem is then to solve the system of equations

$$\begin{array}{lll} r_1 & F_1(\bar{X}_1, \bar{L}_1) = 0, & t_1 \\ r_2 & F_2(\bar{X}_2, \bar{L}_2) = 0, & t_2 \\ u & G(\bar{X}_1, \bar{X}_2, \bar{Y}_m, t) = 0, & \end{array} \qquad (3\text{-}4)$$

for $\bar{X}_1$ and $\bar{X}_2$. The overhead bar denotes final (adjusted) value as in Section 2. We make strict distinction between F_1 and F_2, the primary mathematical models giving the functional relationship between the parameters and observables, and the secondary mathematical model G giving the functional relationship among the parameters themselves. The former model is usually referred to as the observation equations in adjustments, and the latter model is known in Kalman filtering as the dynamical model. Also note the separation of the two types of models according to time epochs t_1 and t_2. In essence, F_1 and F_2 are linked via G. The model errors $\bar{Y}_m$ will be explained later.

We will assume the primary models to be of the non-linear implicit variety. Choosing to work with linear systems of equations we approximate F_1 and F_2 by a linear Taylor series as follows:

$$F_1(\bar{X}_1, \bar{L}_1) = F_1(X_1, L_1) + \left.\frac{\partial F_1}{\partial \bar{X}_1}\right|_{X_1, L_1} (\bar{X}_1 - X_1) + \left.\frac{\partial F_1}{\partial \bar{L}_1}\right|_{X_1, L_1} (\bar{L}_1 - L_1) = 0, \qquad (3\text{-}5)$$

$$= \underset{r_1 x 1}{W_1} + \underset{r_1 x u}{A_1} \underset{u x 1}{\hat{X}_1} + \underset{r_1 x n_1}{B_1} \underset{n_1 x 1}{\hat{V}_1} = 0, \qquad (3\text{-}6)$$

and

$$F_2(\bar{X}_2,\bar{L}_2) = F_2(X_2,L_2) + \left.\frac{\partial F_2}{\partial \bar{X}_2}\right|_{X_2,L_2} (\bar{X}_2-X_2) + \left.\frac{\partial F_2}{\partial \bar{L}_2}\right|_{X_2,L_2} (\bar{L}_2-L_2) = 0 \qquad (3\text{-}7)$$

$$= \underset{r_2x1}{W_2} + \underset{r_2xu}{A_2}\ \underset{uxl}{\hat{X}_2} + \underset{r_2xn_2}{B_2}\ \underset{n_2xl}{\hat{V}_2} = 0 \ , \qquad (3\text{-}8)$$

The observational vectors L_1 and L_2 (that is between two time epochs) are assumed to be statistically independent. Their respective weight matrices are related to the a priori variance factor and variance-covariance matrices by the following:

$$P_1 = \sigma_o^2 \Sigma_{L_1}^{-1} \ , \qquad (3\text{-}9)$$

$$P_2 = \sigma_o^2 \Sigma_{L_2}^{-1} \ . \qquad (3\text{-}10)$$

The observational residuals (corrections) are assumed to be random with zero mean; $\hat{V}_1$ and $\hat{V}_2$ denote the least squares estimates for these quantities

The vectors $\hat{X}_1$ and $\hat{X}_2$ are respectively the least squares estimates for the differences $\bar{X}_1-X_1$ and $\bar{X}_2-X_2$, thus allowing us to write the following expressions for the least squares estimates of the complete quantities:

$$\hat{\bar{X}}_1 = X_1 + \hat{X}_1 \ ,$$

$$\hat{\bar{X}}_2 = X_2 + \hat{X}_2 \ .$$

It should be noted that the entire problem is based on one value of the vector of parameters - that being X_1^o or X_1. We will see from equation 3-40 that X_2^o or X_2 is related to X_1^o or X_1. Further, the vector of parameters are stipulated in Kalman filtering to have the same dimensions* from one time epoch (t_1) to another epoch (t_2).

---

* See Section 4.2 for other possibilities regarding the dimensions of the vectors of parameters.

An independent a priori estimate of $\bar{X}_1$ is denoted by X_1 with weight matrix

$$P_x = \sigma_o^2 \Sigma_{x_1}^{-1} . \qquad (3\text{-}10a)$$

In this case X_1 is said to be a quasi-observable. The "observational" corrections to X_1 are assumed to be random with zero mean. X_1 denotes the least squares estimate for these corrections.

We now turn to the secondary model, and begin our derivation by assuming G to be non-linear, explicit in $\bar{X}_2$ and the model corrections $\bar{Y}_m$, and implicit in $\bar{X}_1$. Thus we write

$$G(\bar{X}_1, \bar{X}_2, \bar{Y}_m, t) = \bar{X}_2 - F_3(\bar{X}_1, t) - \bar{Y}_m = 0 . \qquad (3\text{-}11)$$

Linearizing as before we get

$$G(\bar{X}_1, \bar{X}_2, \bar{Y}_m, t) = \bar{X}_2 + (\bar{X}_2 - \bar{X}_2)$$
$$= [F_3(X_1, t) + \frac{\partial F_3}{\partial \bar{X}_1}\Big|_{X_1} (\bar{X}_1 - X_1)] - Y_m - (\bar{Y}_m - Y_m) = 0.$$

We know that

$$X_2 = F_3 (X_1, t) \qquad (3\text{-}12)$$

since this is how the parameters at time t_2 are usually calculated. Also

$$Y_m = 0$$

since the expected value of the model corrections are assumed to be random variables with zero mean. After making the equivalences

$$\frac{\partial F_3}{\partial \bar{X}_1}\Big|_{X_1} = \Phi ,$$

which is termed the **transition matrix**, and

$$\hat{\bar{X}}_1 - X_1 = \hat{X}_1 ,$$
$$\hat{\bar{X}}_2 - X_2 = \hat{X}_2 ,$$
$$\hat{\bar{Y}}_m - Y_m = \hat{Y}_m ,$$

which are the least squares estimates for the differences, the secondary model becomes approximated by

$$G(\bar{X}_1, \bar{X}_2, \bar{Y}_m, t) = - \underset{uxu}{\Phi} \underset{uxl}{\hat{X}_1} + \underset{uxl}{\hat{X}_2} - \underset{uxl}{\hat{Y}_m} = 0 \cdot \qquad (3\text{-}13)$$

The a priori weight matrix associated with the model errors is

$$P_m = \sigma_o^2 \Sigma_m^{-1} \quad . \qquad (3\text{-}13a)$$

These model errors are assumed to be statistically independent from one application of the secondary model for a given time interval to the consecutive application for another time interval.

Thus we have seen how we pass from a model in terms of the full value of the parameters (equation 3-11) to one in terms of the corrections (equation 3-13). This was done to be consistent with the primary models (equations 3-6 and 3-8).

3.2 Derivation of the System of Equations

The variation function from which we get the least squares normal equations and the Kalman filter equations is

$$\begin{aligned} \phi = & \hat{V}_1^T P_1 \hat{V}_1 + \hat{V}_2^T P_2 \hat{V}_2 + \hat{Y}_m^T P_m \hat{Y}_m + \hat{X}_1^T P_x \hat{X}_1 \\ & + 2\hat{K}_1^T (A_1 \hat{X}_1 + B_1 \hat{V}_1 + W_1) \\ & + 2\hat{K}_2^T (A_2 \hat{X}_2 + B_2 \hat{V}_2 + W_2) \\ & + 2\hat{K}_3^T (\hat{X}_2 - \Phi \hat{X}_1 - \hat{Y}_m) \quad . \end{aligned} \qquad (3\text{-}14)$$

Note there are four quadratic forms to be minimized, that of the two

sets of observations, one for model errors in the secondary model, and the last for the a priori information on the parameters. The problem is to find this minimum subject to the three constraint functions arising from the linearized mathematical models. We use the Lagrange method to solve the extremal problem, and thereby adding the three sets of unknown Lagrange correlates $\hat{K}_1$, $\hat{K}_2$ and $\hat{K}_3$.

Taking the derivatives of the variation function with respect to the variates, and equating them to zero, yields:

$$\frac{\partial \phi}{\partial \hat{V}_1} = 2\hat{V}_1^T P_1 + 2\hat{K}_1^T B_1 = 0; \quad P_1 \hat{V}_1 + B_1^T \hat{K}_1 = 0 \; ; \tag{3-15}$$

$$\frac{\partial \phi}{\partial \hat{V}_2} = 2\hat{V}_2^T P_2 + 2\hat{K}_2^T B_2 = 0; \quad P_2 \hat{V}_2 + B_2^T \hat{K}_2 = 0 \; ; \tag{3-16}$$

$$\frac{\partial \phi}{\partial \hat{Y}_m} = 2\hat{Y}_m^T P_m - 2\hat{K}_3^T = 0; \quad P_m \hat{Y}_m - \hat{K}_3 = 0 \; ; \tag{3-17}$$

$$\frac{\partial \phi}{\partial \hat{X}_1} = 2\hat{X}_1^T P_x + 2\hat{K}_1^T A_1 - 2\hat{K}_3^T \Phi = 0 \; ;$$

$$P_x \hat{X}_1 + A_1^T \hat{K}_1 - \Phi^T \hat{K}_3 = 0 \; ; \tag{3-18}$$

$$\frac{\partial \phi}{\partial \hat{X}_2} = 2\hat{K}_2^T A_2 + 2\hat{K}_3^T = 0; \quad A_2^T \hat{K}_2 + \hat{K}_3 = 0 \; . \tag{3-19}$$

These five equations along with the three linearized mathematical models make-up the least squares normal equations system in the "most expanded form", namely:

$$\begin{bmatrix} P_1 & 0 & 0 & B_1^T & 0 & 0 & 0 & 0 \\ 0 & P_2 & 0 & 0 & B_2^T & 0 & 0 & 0 \\ 0 & 0 & P_m & 0 & 0 & -I & 0 & 0 \\ B_1 & 0 & 0 & 0 & 0 & 0 & A_1 & 0 \\ 0 & B_2 & 0 & 0 & 0 & 0 & 0 & A_2 \\ 0 & 0 & -I & 0 & 0 & 0 & -\Phi & I \\ 0 & 0 & 0 & A_1^T & 0 & -\Phi^T & P_x & 0 \\ 0 & 0 & 0 & 0 & A_2^T & I & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{V}_1 \\ \hat{V}_2 \\ \hat{Y}_m \\ \hat{K}_1 \\ \hat{K}_2 \\ \hat{K}_3 \\ \hat{X}_1 \\ \hat{X}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ W_1 \\ W_2 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 0 \ . \qquad (3\text{-}20)$$

3.3 The Kalman Expressions

We first obtain an expression for $\hat{X}_1$ using only the observations L_1 in F_1^* and the a priori information X_1 and P_x. The normal equation system is obtained from equation 3-20 by deleting all matrices associated with F_2 and G. The resulting system is:

$$\begin{bmatrix} P_1 & B_1^T & 0 \\ B_1 & 0 & A_1 \\ 0 & A_1^T & P_x \end{bmatrix} \begin{bmatrix} \hat{V}_1' \\ \hat{K}_1' \\ \hat{X}_1' \end{bmatrix} + \begin{bmatrix} 0 \\ W_1 \\ 0 \end{bmatrix} = 0 \ . \qquad (3\text{-}21)$$

The above is identical to the combined case with weighted parameters

---

* Since the estimate of $\hat{X}_1$ is a result of information only in F_1, we designate this partial solution by $\hat{X}_1'$.

of Section 2.4 (equation 2-21). The solution of this case is described by equations 2-74 to 2-94. The solution vector for the parameters is:

$$\hat{\bar{X}}_1' = X_1 + \hat{X}_1' \quad , \tag{3-22}$$

$$\hat{X}_1' = -[A_1^T(B_1P_1^{-1}B_1^T)^{-1} A_1+P_x]^{-1} A_1^T(B_1P_1^{-1}B_1^T)^{-1}W_1 \quad , \tag{3-23}$$

with variance-covariance matrix

$$\Sigma_{\hat{\bar{x}}_1'} = \sigma_o^2 [A_1^T (B_1P_1^{-1}B_1^T)^{-1} A_1 + P_x]^{-1} \quad . \tag{3-24}$$

In the case that the a priori information consists only of F_1, that is P_x equals zero, then the solution is given by equations 2-98 to 2-114. For the parameters we have:

$$\hat{\bar{X}}_1' = X_1^o + \hat{X}_1' \quad , \tag{3-25}$$

$$\hat{X}_1' = - [A_1^T (B_1P_1^{-1}B_1^T)^{-1} A_1]^{-1} A_1^T(B_1P_1^{-1}B_1^T)^{-1} W_1^o \quad , \tag{3-26}$$

with variance-covariance matrix

$$\Sigma_{\hat{\bar{x}}_1'} = \Sigma_{\hat{x}_1'} = \sigma_o^2 [A_1^T (B_1P_1^{-1} B_1^T)^{-1} A_1]^{-1} \tag{3-27}$$

Let us now return to the most expanded form of the normal equations (equation 3-20) to obtain a solution for $\hat{X}_2$ using all information contained in F_1, F_2 and G. First $\hat{V}_1$, $\hat{V}_2$ and $\hat{Y}_m$ are eliminated, which yields the following system:

$$\begin{bmatrix} -M_1 & 0 & 0 & A_1 & 0 \\ 0 & -M_2 & 0 & 0 & A_2 \\ 0 & 0 & -P_m^{-1} & -\Phi & I \\ A_1^T & 0 & -\Phi^T & P_x & 0 \\ 0 & A_2^T & I & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{K}_1 \\ \hat{K}_2 \\ \hat{K}_3 \\ \hat{X}_1 \\ \hat{X}_2 \end{bmatrix} + \begin{bmatrix} W_1 \\ W_2 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 0 \ , \qquad (3\text{-}28)^*$$

where the newly introduced quantity

$$M_2 = B_2 \, P_2^{-1} \, B_2^T \ , \qquad (3\text{-}29)$$

and recalling from equation 2-36 that

$$M_1 = B_1 \, P_1^{-1} \, B_1^T \ . \qquad (3\text{-}29a)$$

Next we prepare the normal equation system (equation 3-28) for the elimination of $\hat{K}_1$ and $\hat{X}_1$ by performing elementary row and column transformations; the result is

$$\left[\begin{array}{c|cccc} -M_1 & A_1 & 0 & 0 & 0 \\ \hline A_1^T & P_x & -\Phi^T & 0 & 0 \\ 0 & -\Phi & -P_m^{-1} & I & 0 \\ 0 & 0 & I & 0 & A_2^T \\ 0 & 0 & 0 & A_2 & -M_2 \end{array}\right] \left[\begin{array}{c} \hat{K}_1 \\ \hline \hat{X}_1 \\ \hat{K}_3 \\ \hat{X}_2 \\ \hat{K}_2 \end{array}\right] + \left[\begin{array}{c} W_1 \\ \hline 0 \\ 0 \\ 0 \\ W_2 \end{array}\right] = 0. \qquad (3\text{-}30)$$

Elimination of $\hat{K}_1$ yields

---

* In this equation system $\hat{X}_1$ and the other parameters can be solved for using the complete information (that is F_1, F_2 and G).

$$\left[\begin{array}{c|ccc} A_1^T M_1^{-1} A_1 \quad P_x & -\Phi^T & 0 & 0 \\ \hline -\Phi & -P_m^{-1} & I & 0 \\ 0 & I & 0 & A_2^T \\ 0 & 0 & A_2 & -M_2 \end{array}\right] \left[\begin{array}{c} \hat{X}_1 \\ \hline \hat{K}_3 \\ \hat{X}_2 \\ \hat{K}_2 \end{array}\right] + \left[\begin{array}{c} A_1^T M_1^{-1} W_1 \\ \hline 0 \\ 0 \\ W_2 \end{array}\right] = 0 \ . \tag{3-31}$$

The first equation from the above hyper-matrix equation allows one to obtain a solution for $\hat{X}_1$, that is using all information, namely

$$\hat{X}_1 = \hat{X}_1' + (N_1 + P_x)^{-1} \Phi^T \hat{K}_3 \ . \tag{3-31a}$$

Next we eliminate $\hat{X}_1$ in order to get our desired expression for $\hat{X}_2$. This yields a normal equation system containing only variates $(\hat{X}_2, \hat{K}_2, \hat{K}_3)$ related to the current set of models F_2 and G, namely

$$\left[\begin{array}{c|cc} -\left[\Phi (N_1 + P_x)^{-1} \Phi^T + P_m^{-1}\right] & I & 0 \\ \hline I & 0 & A_2^T \\ 0 & A_2 & -M_2 \end{array}\right] \left[\begin{array}{c} \hat{K}_3 \\ \hline \hat{X}_2 \\ \hat{K}_2 \end{array}\right] + \left[\begin{array}{c} \Phi (N_1 + P_x)^{-1} U_1 \\ \hline 0 \\ W_2 \end{array}\right] = 0, \tag{3-32}$$

where

$$N_1 = A_1^T M_1^{-1} A_1 \ , \tag{3-33}$$

and

$$U_1 = A_1^T M_1^{-1} W_1 \ . \tag{3-34}$$

We now interject the definition of the predicted value of the parameters at t_2 as $\hat{\bar{X}}_2' = F_3 (\hat{\bar{X}}_1', t)$.

The linear form is $\hat{\bar{X}}_2' = \Phi \hat{\bar{X}}_1' \ , \tag{3-35}$

where $\bar{X}_1'$ is the partial solution for the parameters (equation 3-22, or 3

Applying the covariance law to the above we obtain the predicted variance-covariance matrix for the parameters as

$$\Sigma_{\hat{x}_2'} = \sigma_o^2 Q_{\hat{x}_2'} = \sigma_o^2 (N_2')^{-1} = \sigma_o^2 [\Phi(N_1+P_x)^{-1} \Phi^T + P_m^{-1}] , \tag{3-36}$$

where the first term follows directly from the propagation of variances and covariances from $\hat{\bar{X}}_1'$ into $\hat{\bar{X}}_2'$ in equation 3-35, while the second term represents the contribution from the model errors. Note that this predicted value depends upon the weights on the observations L_1 (P_1 in $N_1 = A_1^T M_1^{-1} A_1$), the a priori information P_x on X_1, and the model errors represented by the weights P_m. All of these quantities appear in the hyper-matrix equation 3-32.

The aim is to obtain an expression for $\hat{X}_2$ as a function of matrices and vectors we know. To achieve this goal we eliminate $\hat{K}_3$ from equation 3-32. This results in

$$\begin{bmatrix} N_2' & A_2^T \\ A_2 & -M_2 \end{bmatrix} \begin{bmatrix} \hat{X}_2 \\ \hat{K}_2 \end{bmatrix} + \begin{bmatrix} -N_2' \hat{X}_2' \\ W_2 \end{bmatrix} = 0 , \tag{3-37}$$

where already from equation 3-36 we had

$$N_2' = [\Phi(N_1+P_x)^{-1} \Phi^T + P_m^{-1}]^{-1} , \tag{3-38}*$$

and now

$$\hat{X}_2' = -\Phi (N_1+P_x)^{-1} U_1 \tag{3-39}$$

$$\boxed{\hat{X}_2' = \Phi \hat{X}_1'} . \tag{3-40}$$

---

* N_2' is defined in terms of an inverted quantity on the right hand-side to be consistent with equation 3-36 .

Note the following relationship between the corrective quantity $\hat{X}_2'$ immediately above and the entire quantity $\hat{\bar{X}}_2'$ (equation 3-35):

$$\hat{\bar{X}}_2' = X_2^o + \hat{X}_2' = \Phi X_1 + \Phi\hat{X}_1' = \Phi(X_1+\hat{X}_1') = \Phi \hat{\bar{X}}_1' , \qquad (3\text{-}41)$$

which is equivalent to equation 3-35.

The first equation in 3-37 expresses $\hat{X}_2$ as a function of known quantities and the unknown vector of Lagrange correlates $\hat{K}_2$, namely

$$N_2' \hat{X}_2 + A_2^T \hat{K}_2 - N_2' \hat{X}_2' = 0 , \qquad (3\text{-}42)$$

and after premultiplication by $(N_2')^{-1}$ we get

$$\hat{X}_2 = \hat{X}_2' - (N_2')^{-1} A_2^T \hat{K}_2 . \qquad (3\text{-}43)$$

The final step is to obtain an expression for $\hat{K}_2$ as a function of known quantities and substitute it in the expression immediately above. From (3-37) we get the needed expression by eliminating $\hat{X}_2$, namely

$$[-M_2 - A_2(N_2')^{-1} A_2^T] \hat{K}_2 + W_2 + A_2 (N_2')^{-1} N_2' \hat{X}_2' = 0 , \qquad (3\text{-}44)$$

and

$$\hat{K}_2 = [M_2 + A_2(N_2')^{-1} A_2^T]^{-1} (W_2 + A_2 \hat{X}_2') . \qquad (3\text{-}45)$$

Substituting (3-45) into (3-43) yields

$$\boxed{\hat{X}_2 = \hat{X}_2' - (N_2')^{-1} A_2^T [M_2 + A_2(N_2')^{-1} A_2^T]^{-1} (W_2 + A_2\hat{X}_2') .} \qquad (3\text{-}46)$$

In Kalman filtering theory the expression

$$\boxed{(N_2')^{-1} A_2^T [M_2 + A_2 (N_2')^{-1} A_2^T]^{-1} = G} \qquad (3\text{-}47)$$

is called the gain matrix. Rewriting equation 3-46 we get

$$\boxed{\hat{X}_2 = \hat{X}_2' - G\,(W_2 + A_2\,\hat{X}_2')\ .} \tag{3-48}$$

The variance-covariance matrix for $\hat{X}_2'$ is obtained by first rearranging equation 3-48 so that a coefficient matrix preceeds each random variable ($\hat{X}_2'$ and W_2 in this instance) as follows:

$$\hat{X}_2 = (I - GA_2)\,\hat{X}_2' - GW_2\ . \tag{3-49}$$

Applying the covariance law to the above yields

$$Q_{\hat{x}_2} = (I - GA_2)Q_{\hat{x}_2'}\,(I - GA_2)^T + GQ_{W_2}\,G^T\ . \tag{3-50}$$

Since

$$W_2^o = F_2\,(X_2^o,\ L_2) \tag{3-51}$$

and

$$Q_{W_2^o} = \frac{\partial F_2}{\partial L_2}\bigg|_{X_2^o,L_2}\,Q_{L_2}\,\left(\frac{\partial F_2}{\partial L_2}\bigg|_{X_2^o,L_2}\right)^T \tag{3-52}$$

$$= B_2\,P_2^{-1}\,B_2^T = M_2\ . \tag{3-53}$$

Substituting equation 3-53 into 3-50 and carrying out the algebra yields

$$Q_{\hat{x}_2} = [(N_2')^{-1} - GA_2\,(N_2')^{-1}] - [(N_2')^{-1}A_2^T - GA_2(N_2')^{-1}A_2^T - GM_2]\,G^T \tag{3-54}$$

From (3-47)

$$G\,[M_2 + A_2(N_2')^{-1}\,A_2^T] = (N_2')^{-1}\,A_2^T\ , \tag{3-55}$$

$$-GM_2 - GA_2\,(N_2')^{-1}\,A_2^T + (N_2')^{-1}\,A_2^T = 0\ , \tag{3-56}$$

which is the second term in (3-54). Thus our final expression for the variance-covariance matrix of the latest estimate of the parameters is given by

$$\boxed{\Sigma_{\hat{x}_2} = \sigma_o^2 Q_{\hat{x}_2} = \sigma_o^2 [(N_2')^{-1} - GA_2(N_2')^{-1}] .} \qquad (3\text{-}57)$$

The series of expressions constituting the Kalman filtering equations are equations 3-40, 3-36, 3-47, 3-48 and 3-57.

Prediction of state vector

$$\hat{X}_2' = \Phi \hat{X}_1' ; \qquad (3\text{-}58)$$

Prediction of new variance-covariance matrix

$$\Sigma_{\hat{x}_2'} = \sigma_o^2 Q_{\hat{x}_2'} = \sigma_o^2 (N_2')^{-1} = \sigma_o^2 [\Phi(N_1+P_x)^{-1}\Phi^T + P_m^{-1}]; \qquad (3\text{-}59)$$

Computation of gain matrix

$$G = (N_2')^{-1} A_2^T [M_2 + A_2 (N_2')^{-1} A_2^T]^{-1} ; \qquad (3\text{-}60)$$

Computation of state vector

$$\hat{X}_2 = \hat{X}_2' - G (W_2 + A_2 \hat{X}_2') ; \qquad (3\text{-}61)$$

Computation of the variance-covariance matrix of state vector

$$\Sigma_{\hat{x}_2} = \sigma_o^2 Q_{\hat{x}_2} = \sigma_o^2 [(N_2')^{-1} - GA_2 (N_2')^{-1}] . \qquad (3\text{-}62)$$

Note, if we also require a solution for $\hat{X}_1$ using all data, that is F_1, F_2, and G, then we simply return to equation 3-31a.

To show the correspondence of our Kalman filtering expressions to those contained in the literature, e.g. [Kalman 1960], we make a

correspondence between the two different notations as follows:

Ours	Kalman	Ours	Kalman
$-W_2$	Z_k	$\hat{X}_2$	$X_{k/k}$
A_2	H_k	$\Sigma_{\hat{x}_2}$	$P_{k/k}$
$\hat{X}_2'$ (new)	$X_{k/k-1}$	$(N_2')^{-1}$	$P_{k/k-1}$
$-\hat{V}_2$	V_k	$\hat{K}_2$	no equivalent
P_m^{-1} (new)	Q_k	G (new)	K_k
M_2	R_k	Y_m	W_k

Let us now obtain a deeper insight into the workings of the Kalman expressions. First note that a priori information on the parameters is needed to start the process. A priori information can be of two possible types:

(i) "observed" value of X_1 with an associated weight matrix P_x from an independent determination;

(ii) a determination of X_1 via the information contained in F_1.

These two types of information can enter both two together (equations 3-23 and 3-24); only as normal equations (F_1 in 3-25, 3-26, and 3-27); or lastly in terms of weights (X_1 with weight matrix P_x) The representation of these three possibilities in the Kalman expression equation 3-59 is respectively $N_1 \neq 0$, $P_x \neq 0$; $N_1 \neq 0$, $P_x = 0$; and $N_1 = 0$, $P_x \neq 0$.

By studying equations 3-58 and 3-59 we discover that the model noise does affect the prediction of the state vector via the shortcomings that exist in the definition of the transition matrix Φ. The effect of this model noise on the predicted variance-covariance matrix $\Sigma_{\hat{x}_2'}$ enters explicitly via the weight matrix P_m. Also contained in $\Sigma_{\hat{x}_2'}$ is the "noise" in the observations (via P) and in the starting values of the parameters (via P_x). We witness that to compute $\hat{X}_2'$ and $\Sigma_{\hat{x}_2'}$ no matrix inversions are necessary because in equation 3-59

$$\underset{uxu}{(N_2')^{-1}} = [\underset{uxu}{\Phi}\ \underset{uxu}{(N_1} + \underset{uxu}{P_x})^{-1}\ \underset{uxu}{\Phi^T} + \underset{uxu}{P_m^{-1}}] , \tag{3-63}$$

the one matrix inversion has already been carried out in the course of obtaining the solution $\hat{X}_1'$.

For the final estimates $\hat{X}_2$ and $\Sigma_{\hat{x}_2}$, one matrix inversion is required. Namely, it is the matrix of order r_2 (number of equations in F_2) encountered in the gain matrix (equation 3-60) in the term:

$$[M_2 + A_2\ (N_2')^{-1}\ A_2^T]^{-1} \tag{3-64}$$

$$= [\underset{r_2xn_2}{B_2}\ \underset{n_2xn_2}{\Gamma_2^{-1}}\ \underset{n_2xr_2}{B_2^T} + \underset{r_2xu}{A_2}\ \underset{uxu}{(N_2')^{-1}}\ \underset{uxr_2}{A_2^T}]^{-1} . \tag{3-65}$$

The matrix to be inverted is fully populated and for a parametric case is the same order as the number of new observations being added. This means that if one observation is added each stage (that is in F_2)

and F_2 fits into the parametric case, the matrix to be inverted is of order one. We will return to this subject in Sections 4 and 6 when we compare the various methods.

4. RELATIONSHIPS AMONG KALMAN, BAYES, SEQUENTIAL AND PHASE EXPRESSIONS

Having derived the Kalman filter equations, we now proceed to:

(i) show the equivalence of the Bayes and Kalman filter equations by using a matrix inversion lemma;

(ii) deduce the sequential and phase expressions respectively from the Kalman and Bayes equations by deleting the time variation in the state vector (parameters);

(iii) show the equivalence of the sequential and phase expressions by using the same inversion lemma as in (i) above;

(iv) demonstrate the equivalence of the Tienstra phase and sequential expressions;

(v) discuss the computational efficiency of the Kalman (sequential) and Bayes (phase) methods.

The theme of this section is to show that even though the Kalman (sequential) and Bayes (phase) expressions are mathematically equivalent they are not identical from the computational point of view.

4.1 Equivalence of Kalman and Bayes Expressions

First we write from Morrison [1969] the Bayes filter expressions in our notation:

Prediction of state vector

$$\hat{X}_2' = \Phi \hat{X}_1' \ ; \tag{4-1}$$

Prediction of new variance-covariance matrix

$$\Sigma_{\hat{X}_2'} = \sigma_o^2 (N_2')^{-1} = \sigma_o^2 [\Phi (N_1 + P_x)^{-1} \Phi^T + P_m^{-1}] \ ; \tag{4-2}$$

Computation of variance-covariance matrix of state vector

$$\Sigma_{\hat{X}_2} = \sigma_o^2 N_2^{-1} = \sigma_o^2 [N_2' + A_2^T M_2^{-1} A_2]^{-1} \ ; \tag{4-3}$$

Computation of gain matrix

$$G = N_2^{-1} A_2^T M_2^{-1} \ ; \tag{4-4}$$

Computation of state vector

$$\hat{X}_2 = \hat{X}_2' - G(W_2 + A_2 \hat{X}_2') \ . \tag{4-5}$$

To show the mathematical equivalence of the Kalman and Bayes expressions we invoke the following inversion lemma, contained, for example in Krakiwsky [1968], Morrison [1969], Wells [1974]:

$$[S^{-1} + T^T R^{-1} T]^{-1} = S - ST^T (R + TST^T)^{-1} TS, \tag{4-6}$$

where S and R are positive definite matrices with R being a different order than S, and that $T^T R^{-1} T$ is of the same order as S. This lemma can be corroborated simply by showing that the matrix product

$$(S^{-1} + T^T R^{-1} T) \ (S - ST^T \ (R + TST^T)^{-1} TS) \tag{4-7a}$$

reduces to an identity matrix. The above lemma is sometimes known as the Schurr identity or the inside out rule and has been used by other authors such as Tapley [1972] and Kouba [1970] to show similar equivalences.

We need yet another matrix identity, namely that proved for example in Morrison [1969] and Wells [1974], and that used for by authors such as Schwarz [1974], and Reigber and Ilk [1975]:

$$(S^{-1} + T^T R^{-1} T)^{-1} T^T R^{-1} = S T^T (R + T S T^T)^{-1}, \tag{4-7b}$$

where S and R are again of different order and are positive definite matrices, and T is conformable for multiplication with R and S.

If in the first identity we replace T by A_2, R by M_2 and S by $(N_2')^{-1}$ we obtain

$$[N_2' + A_2^T M_2^{-1} A]^{-1} = (N_2')^{-1} - (N_2')^{-1} A_2^T [M_2 + A_2 (N_2')^{-1} A_2^T]^{-1} A_2 (N_2')^{-1} \tag{4-8}$$

$$= (N_2')^{-1} - G A_2 (N_2')^{-1}, \tag{4-9}$$

where the Kalman gain matrix

$$G = (N_2')^{-1} A_2^T [M_2 + A_2 (N_2')^{-1} A_2^T]^{-1}. \tag{4-10}$$

Comparing the left-hand side of (4-8) with (4-3) of the Bayes expressions and the right-hand side of (4-9) with (3-62) of the Kalman expressions, we can conclude that the respective variance-covariance matrices are mathematically equivalent.

To show the equivalence of the two state vectors ($\hat{X}_2$) we simply prove that the two gain matrices G in (3-61) and (4-5) are mathematically equivalent. Using the second matrix identity with the substitutions S by $(N_2')^{-1}$, T by A_2, and R by M_2, we get

$$[N_2' + A_2^T M_2^{-1} A_2]^{-1} A_2^T M_2^{-1} = (N_2')^{-1} A_2^T [M_2 + A_2 (N_2')^{-1} A_2^T]^{-1}, \tag{4-11}$$

that is

$$G \text{ (Bayes)} = G \text{ (Kalman)}. \tag{4-12}$$

Thus the Kalman and Bayes expressions are mathematically equivalent.

To obtain insight into the Bayes expressions we analyse (4-1) to (4-5). Predicting the state vector $\hat{X}_2'$ and its variance-covariance matrix $\Sigma_{\hat{X}_2'}$ requires no matrix inversion since in the matrix

$$(N_2')^{-1}_{uxu} = [\ \Phi\ (\ N_1 + P_x)^{-1}_{uxu\ uxu}\ \Phi^T_{uxu} + P_m^{-1}_{uxu}\ ]\ , \qquad (4\text{-}13)$$

the single matrix inversion $(N_1 + P_x)$ was already obtained in the course of obtaining the original estimate $\hat{X}_1'$. In obtaining the variance-covariance $\Sigma_{\hat{X}_2}$, two matrix inversions in

$$[\ N_2' + A_2^T\ (\ B_2\ P_2^{-1}\ B_2^T\ )^{-1}\ A_2\ ] \qquad (4\text{-}14)$$

are necessary. First, a fully populated matrix of order r_2 is inverted (M_2). For the parametric case and uncorrelated observations, this matrix is diagonal. Then, in order to obtain N_2', the matrix in equation 4-13 of order u has to be inverted. Finally, a fully populated matrix of order u is inverted. This latter matrix is also used in computing the estimate $\hat{X}_2$ (equation 4-5).

For example, for uncorrelated observations and the parametric case, only two matrix inversions of order u are required in the Bayes expressions. This situation and others are compared with Kalman filtering and other methods in Section 6.

4.2 Sequential Expressions from the Kalman Expressions

The sequential expressions follow directly from Kalman expressions (3-58) to (3-62). In these expressions we delete all matrices and vectors pertaining to the time variation in the state

vector, namely Φ, P_m, and Y_m. We also drop the subscripts "1" and "2" of the parameters; we retain the superscript "prime" (') to indicate an estimate based on only partial data, that is only data in F_1.

The predicted state vector $\hat{X}'_2$ (equation 3-58) reduces to the initial estimate $\hat{X}'$, and is given by equation 3-23:

$$\hat{X}' = -[N_1 + P_x]^{-1} A_1^T M_1^{-1} W_1 . \tag{4-15}$$

The predicted variance-covariance matrix $\Sigma_{\hat{X}'_2}$ (equation 3-59) reduces to $\Sigma_{\hat{X}'}$, and is given by equation 3-24:

$$\Sigma_{\hat{X}'} = \sigma_o^2 [N_1 + P_x]^{-1} . \tag{4-16}$$

The computation of the gain matrix (equation 3-60) reduces to

$$G = (N_1+P_x)^{-1} A_2^T [M_2+A_2(N_1+P_x)^{-1} A_2^T]^{-1} . \tag{4-17}$$

The solution vector and variance-covariance matrix reduce to

$$\hat{X} = \hat{X}' - G(W_2 + A_2\hat{X}') \tag{4-18}$$

and

$$\Sigma_{\hat{X}} = \sigma_o^2 [(N_1+P_x)^{-1} - GA_2 (N_1+P_x)^{-1}] . \tag{4-19}$$

In the literature the sequential expressions have been derived for the case P_x equals zero. Making this simplification to the above equations we have the following expressions which are equivalent to those given in [Schmid and Schmid 1965; Wells and Krakiwsky 1971]:

$$\boxed{\hat{X}' = -N_1^{-1} U_1 ;} \tag{4-20}$$

$$\boxed{\Sigma_{\hat{X}'} = \sigma_o^2 N_1^{-1} ;} \tag{4-21}$$

$$\hat{X} = X' - N_1^{-1} A_2^T (M_2 + A_2 N_1^{-1} A_2^T)^{-1} (W_2^o + A_2 \hat{X}') \; ; \tag{4-22}$$

$$\Sigma_{\hat{X}} = \sigma_o^2 [N_1^{-1} - N_1^{-1} A_2^T (M_2 + A_2 N_1^{-1} A_2^T)^{-1} A_2 N_1^{-1}] \; . \tag{4-23}$$

The number and size of matrix inversions in the sequential expressions are identical to that of the Kalman filter expressions. (See the comments there - Section 3).

We now insert a note on the estimated variance factor $\hat{\sigma}_o^2$. We may get two estimates $(\hat{\sigma}_o^2)_1$ and $(\hat{\sigma}_o^2)_2$. They are different as the second contains all the residuals while the first contains only those of the first set of observations. Krakiwsky [1968] has derived a sequential expression for the quadratic form of the residuals which allows the first quadratic form to be up-dated without recomputing the residuals themselves; the equation is

$$(\hat{\sigma}_o^2)_2 = \frac{\hat{V}^T P \hat{V}}{\nu} = \frac{\hat{V}_1^T P_1 \hat{V}_1 + \hat{V}_2^T P_2 \hat{V}_2}{\nu_1 + \nu_2} \tag{4-24}$$

$$= \frac{(\hat{V}_1^T P_1 \hat{V}_1) + \Delta(\hat{V}^T P \hat{V})}{\nu_1 + \nu_2} \; , \tag{4-25}$$

where ν_1 is the degrees of freedom in F_1,

ν_2 is the degrees of freedom in F_2,

$\hat{V}^T P \hat{V}$ total quadratic form of all residuals,

$\hat{V}_1^T P_1 \hat{V}_1$ quadratic from of residuals in F_1 (combined solution),

$\hat{V}_2^T P_2 \hat{V}_2$ quadratic form of residuals in F_2 (combined solution),

$(\hat{V}_1^T P_1 \hat{V}_1)$ quadratic form of residuals in F_1 (solution with F_1 only),

$\Delta(\hat{V}^T P \hat{V})$ change in quadratic form of residuals due to residuals in F_2 and change in the residuals in F_1 due to F_2, and is given by

$$\Delta(V^T PV) = (W_2^T + (\hat{X}')^T A_2^T)\ (M_2 + A_2 N_1^{-1} A_2^T)\ (W_2 + A_2 \hat{X}') \ . \qquad (4\text{-}26)$$

The sequential expressions above are a result of the two sets of models

$$F_1\ (\bar{X},\ \bar{L}_1) = 0 \ , \qquad (4\text{-}27)$$

and

$$F_2\ (\bar{X},\ \bar{L}_2) = 0 \ . \qquad (4\text{-}28)$$

Expressions for more general situations have been developed in [Krakiwsky 1968; Kouba 1970], which allow for common observations and parameters between models F_1 and F_2.

4.3 Phase Expressions from the Bayes Expressions

The phase expressions follow directly from the Bayes expressions (4-1) to (4-5). We delete all quantities pertaining to the time variation in the state vector, that is Φ, P_m and Y_m vanish. We drop the subscripts "1" and "2" on the parameters, but retain the superscript "prime" (') to designate an estimate using only part of the data, that is, only in F_1.

The predicted state vector $\hat{X}'_2$ (4-1) reduces to the initial estimate $\hat{X}'$, and is given by equation 3-23:

$$\boxed{\hat{X}' = -\ [N_1 + P_x]^{-1}\ A_1^T\ M_1^{-1}\ W_1 \ .} \qquad (4\text{-}29)$$

The predicted variance-covariance matrix $\Sigma_{\hat{X}'_2}$ reduces to the initial variance-covariance matrix $\Sigma_{\hat{X}'}$ and is given by (3-24)

$$\boxed{\Sigma_{\hat{X}'} = \sigma_o^2\ [N_1 + P_x]^{-1} \ .} \qquad (4\text{-}30)$$

The variance-covariance of the state vector reduces to

$$\Sigma_X = \sigma_o^2 \left[(N_1+P_x) + A_2^T M_2^{-1} A_2\right]^{-1} = \sigma_o^2 N^{-1} , \tag{4-31}$$

and the state vector to

$$\hat{X} = \hat{X}' - N^{-1} A_2^T M_2^{-1} (W_2 + A_2 \hat{X}') . \tag{4-32}$$

We have retained the weight matrix P_x to show how a priori information on the parameters affects the solution. Phase expressions corresponding to the case P_x equals zero are:

$$\hat{X}' = -N_1^{-1} U_1 , \tag{4-33}$$

$$\Sigma_{\hat{X}'} = \Sigma_{\hat{X}'} = \sigma_o^2 N_1^{-1} , \tag{4-34}$$

$$\Sigma_{\hat{X}} = \sigma_o^2 \left[N_1 + A_2^T M_2^{-1} A_2\right]^{-1} = \sigma_o^2 N^{-1} , \tag{4-35}$$

$$X = \hat{X}' - N^{-1} A_2^T M_2^{-1} (W_2 + A_2 \hat{X}') . \tag{4-36}$$

The number and size of matrix inversions in the above are only similar to the Bayes expressions (See Section 6.4).

It is clear that the variance-covariance matrix is a result of the addition of two sets of normal equations. On the other hand, the solution vector takes on a corrective form similar to the sequential expression (4-22). This is computationally different from the straightforward case of adding normal equations. That is the solution for

$$F_1 (\bar{X}, \bar{L}_1) = 0 , \tag{4-37}$$

$$F_2 (\bar{X}, \bar{L}_2) = 0 , \tag{4-38}$$

is

$$\hat{X} = -(N_1+N_2)^{-1} (U_1 + U_2) = N^{-1} U , \tag{4-39}$$

$$\Sigma_{\hat{X}} = \sigma_o^2 \left[N_1 + N_2\right]^{-1} = \sigma_o^2 N^{-1} , \tag{4-40}$$

where N_1 and N_2 are coefficient matrices and U_1 and U_2 are the constant vectors of the least squares normal equations.

Although the form of the above expression is different from the phase expressions, the number and size of matrix inversions are identical to the phase expressions. We will, however, reserve the name "summation form of normal equations" to distinguish equation 4-39 from the phase form in equation 4-36.

4.4 Equivalence of Phase, Sequential and Summation Equations

The mathematical equivalence between the phase and sequential expressions comes via the fact that we have deduced each from the Kalman and Bayes expressions and the latter have been shown to be equivalent. How about the equivalence of the summation form of the normal equations and the sequential or phase expressions? Wells [1974] has shown the equivalence between the summation form and the sequential equations as follows.

Begin by expanding equation 4-39:

$$\hat{X} = -(N_1 + A_2^T M_2^{-1} A_2)^{-1} (U_1 + A_2^T M_2^{-1} W_2^o) \qquad (4\text{-}41)$$

$$= -(N_1 + A_2^T M_2^{-1} A_2)^{-1} U_1 - (N_1 + A_2^T M_2^{-1} A_2)^{-1} A_2^T M_2^{-1} W_2^o \; . \qquad (4\text{-}42)$$

From the first matrix identity (equation 4-6)

$$(N_1 + A_2^T M_2^{-1} A_2)^{-1} \qquad (4\text{-}43)$$

$$= N_1^{-1} - N_1^{-1} A_2^T (M_2 + A_2 N_1^{-1} A_2^T)^{-1} A_2 N_1^{-1} \; ,$$

and from the second (equation 4-7b)

$$(N_1 + A_2^T M_2^{-1} A_2)^{-1} A_2^T M_2^{-1} = N_1^{-1} A_2^T (M_2 + A_2 N_1^{-1} A_2^T)^{-1} . \quad (4\text{-}44)$$

Substituting equations 4-43 and 4-44 into 4-42 yields

$$\hat{X} = - \{N_1^{-1} - N_1^{-1} A_2^T (M_2 + A_2 N_1^{-1} A_2^T)^{-1} A_2 N_1^{-1}\} U_1 - N_1^{-1} A_2^T (M_2 + A_2 N_1^{-1} A_2^T)^{-1} W_2^o , \quad (4\text{-}45)$$

$$= -N_1^{-1} U_1 + N_1^{-1} A_2^T (M_2 + A_2 N_1^{-1} A_2^T)^{-1} A_2 N_1^{-1} U_1 - N_1^{-1} A_2^T (M_2 + A_2 N_1^{-1} A_2^T)^{-1} W_2^o , \quad (4\text{-}46)$$

$$\hat{X} = \hat{X}_1 - N_1^{-1} A_2^T (M_2 + A_2 N_1^{-1} A_2^T)^{-1} (W_2^o + A_2 \hat{X}_1) . \quad (4\text{-}47)$$

From equation 4-43

$$\Sigma_{\hat{X}} = \sigma_o^2 [N_1 + A_2^T M_2^{-1} A_2]^{-1} \quad (4\text{-}48)$$

$$= \sigma_o^2 [N_1^{-1} - N_1^{-1} A_2^T (M_2 + A_2 N_1^{-1} A_2^T)^{-1} A_2 N_1^{-1}] . \quad (4\text{-}49)$$

We thus see the mathematical equivalence of the summation form of the normal equations and the sequential expressions.

In essence, the summation form of the normal equations is the foundation for the "group method" [Kruger 1905], which was developed for classical adjustments of triangulation nets. Application of the summation form of the normal equations to problems in satellite geodesy can be found in such works as [Kaula 1966; Krakiwsky and Pope 1967; Wells 1974].

4.5 Equivalence of the Tienstra Phase and Sequential Expressions

The aim of this section is to apply the ideas of Tienstra to the same situation as in the foregoing, namely for two models with common parameters but different observations. The models are:

$$F_1 (\bar{X}, \bar{L}_1) = 0 , \tag{4-50}$$

$$A_1\hat{X} + B_1\hat{V}_1 + W_1^o = 0 , \tag{4-51}$$

$$F_2 (\bar{X}, \bar{L}_2) = 0 , \tag{4-52}$$

$$A_2\hat{X} + B_2\hat{V}_2 + W_2^o = 0 . \tag{4-53}$$

Using the principles of Tienstra [1956] we first obtain a solution from F_1, namely

$$\hat{X}' = -N_1^{-1} U_1 , \tag{4-54}$$

$$\Sigma_{\hat{X}'} = \sigma_o^2 N_1^{-1} . \tag{4-55}$$

Next we use the adjusted parameters from phase one (F_1)

$$\hat{\bar{X}}' = X^o + \hat{X}' , \tag{4-56}$$

as quasi-observations with weight matrix

$$P_x = \sigma_o^2 \Sigma_{\hat{X}'}^{-1} = \sigma_o^2 (\sigma_o^2 N_1^{-1})^{-1} = N_1 \tag{4-57}$$

in F_2. Since all quantities are weighted, F_2 takes the linearized form of the condition case adjustment,

$$B^*\hat{V}^* + W^* = 0 , \tag{4-58}$$

where

$$B^* = [A_2 \mid B_2] , \tag{4-59}$$

$$\hat{V}^* = \begin{bmatrix} \delta\hat{X}' \\ \hline \hat{V}_2 \end{bmatrix} , \tag{4-60}$$

$$P^* = \begin{bmatrix} P_x & 0 \\ 0 & P_2 \end{bmatrix} = \begin{bmatrix} N_1 & 0 \\ 0 & P_2 \end{bmatrix} , \tag{4-61}$$

$$W^* = F_2 \ (\bar{X}', \ L_2 \ ) \tag{4-62}$$

$$= F_2 \ (X^o, \ L_2) + \frac{\partial F_2}{\partial \bar{X}_1} \hat{X}' \ , \tag{4-63}$$

$$= W_2^o + A_2 \hat{X}' \ . \tag{4-64}$$

The equations which describe the solution to the condition case adjustment were derived in Section 2. The equation for the residual vector is (2-132)

$$\hat{V} = - P^{-1} B^T \hat{K} \ , \tag{4-65}$$

where

$$\hat{K} = (BP^{-1} B^T)^{-1} W^o. \tag{4-66}$$

Substituting B*, V*, P* and W* into the above yields

$$\hat{K} = \{ [A_2 \ | \ B_2] \left[\begin{array}{c|c} N_1^{-1} & 0 \\ \hline 0 & P_2 \end{array}\right] \left[\begin{array}{c} A_2^T \\ \hline B_2^T \end{array}\right] \}^{-1} (W_2^o + A_2\hat{X}') \tag{4-67}$$

$$= [A_2 \ N_1^{-1} \ A_2^T + B_2 \ P_2^{-1} \ B_2^T]^{-1} \ (W_2^o + A_2 \ \hat{X}') \tag{4-68}$$

$$\hat{V}^* = \left[\begin{array}{c} \delta\hat{X}' \\ \hline \hat{V}_2 \end{array}\right] = - \left[\begin{array}{c|c} N_1^{-1} & 0 \\ \hline 0 & P_2^{-1} \end{array}\right] \left[\begin{array}{c} A_2^T \\ \hline B_2^T \end{array}\right] \ [A_2 N_1^{-1} A_2^T + M_2]^{-1} (W_2^o + A_2\hat{X}') \tag{4-69}$$

$$= - \left[\begin{array}{c} N_1^{-1} A_2^T \ [M_2 + A_2 N_1^{-1} A_2^T]^{-1} \ (W_2^o + A_2 \ \hat{X}') \\ \hline P_2^{-1} B_2^T \ [M_2 + A_2 N_1^{-1} A_2^T]^{-1} \ (W_2^o + A_2 \ \hat{X}') \end{array}\right] . \tag{4-70}$$

The adjusted (final) quantities are

$$\hat{\bar{L}} = L + \hat{V} \tag{4-71}$$

$$= \begin{bmatrix} \hat{\bar{X}} \\ \hat{\bar{L}}_2 \end{bmatrix} = \begin{bmatrix} X^o + \hat{X}' \\ L_2 \end{bmatrix} + \begin{bmatrix} \delta\hat{X}' \\ \hat{V}_2 \end{bmatrix} . \tag{4-72}$$

From the upper portion of the partitioned vector we get

$$\hat{\bar{X}} = X^o + \hat{X}' + \delta\hat{X}' \tag{4-73}$$

$$= X^o + \hat{X}' - N_1^{-1} A_2^T [M_2 + A_2 N_1^{-1} A_2^T]^{-1} (W_2^o + A_2 \hat{X}') , \tag{4-74}$$

the last two terms of which are identical to the sequential expression (4-22) for the solution vector.

Next we obtain an expression for the weight coefficient matrix. Begin recalling from (2-134) that

$$Q_{\hat{\bar{L}}} = P^{-1} - P^{-1} B^T (BP^{-1} B^T)^{-1} BP^{-1} \tag{4-75}$$

$$= \left[\begin{array}{c|c} N_1^{-1} & 0 \\ \hline 0 & P_2^{-1} \end{array}\right] - \left[\begin{array}{c|c} N_1^{-1} & 0 \\ \hline 0 & P_2^{-1} \end{array}\right] \left[\begin{array}{c} A_2^T \\ \hline B_2^T \end{array}\right] [M_2 + A_2 N_1^{-1} A_2^T]^{-1} [A_2 \; B_2] \left[\begin{array}{c|c} N_1^{-1} & 0 \\ \hline 0 & P_2^{-1} \end{array}\right]$$

$$= \left[\begin{array}{c|c} N_1^{-1} - N_1^{-1} A_2^T [M_2 + A_2 N_1^{-1} A_2^T] A_2 N_1^{-1} & -N_1^{-1} A_2^T [M_2 + A_2 N_1^{-1} A_2^T]^{-1} B_2 P_2^{-1} \\ \hline -P_2^{-1} B_2^T [M_2 + A_2 N_1^{-1} A_2^T]^{-1} A_2 N_1^{-1} & P_2^{-1} - P_2^{-1} B_2^T [M_2 + A_2 N_1^{-1} A_2^T]^{-1} B_2 P_2^{-1} \end{array}\right] . \tag{4-76}$$

Thus we have, besides the variances, also the covariance between $\hat{\bar{X}}$ and $\hat{\bar{L}}$. The upper left portion of the hyper-matrix is the weight coefficient matrix of the adjusted parameters. It is identical to the sequential expression of equation 4-23. This completes our proof.

5. COLLOCATION

In this section two different derivations of the collocation equations are given. The first derivation is fundamentally the same as the one given in Moritz [1972] and differs only in detail. One important detail is that we fully exploit the equations of the standard adjustment combined case (Section 2) in deducing the collocation equations - this considerably shortens the derivation. Secondly, weights on the parameters are included as an option - by doing so, we see the role of these weights vis-à-vis the other two sets of weights on the observations and signal, respectively. Also, one can then see the role of these three sets of weights vis-à-vis the weights on the secondary mathematical model of Kalman filtering (Section 3). With a comprehensive knowledge of these four kinds of weights (inverse of variance-covariance matrices)** one begins to discover the similarities and differences among ordinary least squares, collocation, and Kalman filtering. The main characteristic of the first derivation (Sections 5.1 to 5.4) is that the measurement error or noise* does not appear explicitly in the weighted quadratic form to be minimized nor in the constraint function, and as such an estimate for the "noise" is not computed. For this reason, it is said that the "noise" does not play a

---

* This quantity is equal to minus the observational residual in ordinary adjustments.

** σ_o^2 equal to one.

key role in collocation like in ordinary least squares, and as a consequence the two methods are fundamentally different.

In the second derivation (Section 5.5) we formulate the objectives of collocation entirely as a problem in adjustments. Here we find that the "noise" (corrections to the observations) appears explicitly, does play a key role, and is computed.

We will discover that the corresponding expressions for the parameters and signal are identical in collocation and the adjustments formulations.

The aim of this section is only to recapitulate the method of collocation in the context of a synthesis of methods. For a comprehensive treatment of the subject see [Moritz 1972; 1973a; 1973b].

5.1 Collocation Mathematical Model

We begin with the linear explicit model (equation 2-3)

$$\underset{nxl}{L} = \underset{nxu}{A}\ \underset{uxl}{\bar{X}} - \underset{nxl}{V} \quad , \tag{5-1}$$

where L is the vector of n observations, V the vector of n residuals, A the design matrix, and X the vector of u unknown parameters. We now depart slightly from our notation to be consistent with Moritz. The above is rewritten in the form,

$$\underset{nxl}{x} = \underset{nxu}{A}\ \underset{uxl}{\bar{X}} + \underset{nxl}{n} \quad , \tag{5-2}$$

where x is called the "measurement" and n its "noise".

Moritz then extends the above model to

$$\underset{nxl}{x} = \underset{nxu}{A}\ \underset{uxl}{\bar{X}} + \underset{nxl}{s'} + \underset{nxl}{n} \quad , \tag{5-3}$$

where the newly introduced quantity S' is called the "signal". Here we can perhaps interpret S' as the short-coming in the mathematical model, that is the "inability" of the model to describe completely (exactly) the actual relationship among the measurements (x) and unknown parameters (X). In other words, one could imagine that there is an "overflow" from the model into some sort of additional correction (S') to the observations. Even though this may seem plausible at first sight, it is the author's belief that is not what is intended in collocation. What is intended is to attribute the signal directly to the observable, thereby stating that the observable (measurement) has two unknown errors - the signal S' and the noise n. We can liken the noise to a measuring error, or resolution capability, and is thus internal to the instrument. On the other hand, the signal is thought of as being external to the instrument and related to the behaviour of the observable in a particular milieu - like deflections of the vertical in the gravity field, electronically measured distances in the "polluted" atmosphere or in the electron charged ionosphere; or gravity anomalies in the gravity field. An important characteristic of a signal is that it is continuous throughout the domain of a particular "milieu" or "field". One of the requirements of collocation is that the signal has known second moments (variance-covariance matrix), even though the first moments (value of the signal) remain as unknowns to be estimated.

The signal and its variance is not new to geodesists. Since the 1960's we have been calculating the variance of the noise and variance of the signal for electronically measured distances from the formula

$$\sigma_d^2 = a + b \text{ (distance)}.$$

In the above, a is the resolution of the instrument (variance of the noise) and b is a constant in parts per million which when multiplied by the distance is nothing else but the variance of the signal. The latter is a measure of the behaviour of the observed distance in the troposphere. It is rather obvious that there is no covariance in the noise but one suspects that there would be covariance between the signal components of different distances since they are measured in the troposphere. Collocation attempts to account for this correlation through a fully populated variance-covariance matrix for the signal, while in the ordinary least squares treatment, the covariances in the signal are first ignored, then the variance of the signal is combined with the variance of the noise to give one variance (σ_d^2), and finally a solution is made for only one correction (residual) for each measured distance.

In collocation, the condition imposed on the signal is that it be random with zero mean. Thus the measurement x is seen to consist of a systematic part AX, and two random parts, S' and n.

In his development, Moritz introduces the quantity

$$Z = S' + n \ , \qquad (5\text{-}4)$$

where S' denotes the signal at the observation points. S will be reserved to denote the signal at any point in general without observations. These points (p in number) are called computation points and it is at these points that the signal is said to be "predicted".

After considering the above equation, the main model becomes

$$x = A\bar{X} + Z \tag{5-5}$$

and

$$Z = x - A\bar{X} \tag{5-6}$$

represents the random part of observations after subtracting the systematic part AX.

This stage constitutes the end of formulating the collocation mathematical model and the beginning of applying the conventional least squares methodology used throughout the previous sections.

5.2 Collocation Least Squares Problem

We now state the collocation least squares problem. Determine the least squares estimate $\hat{\bar{X}}$ in equation 5-3 under the condition that

$$\hat{V}^T P \hat{V} = \text{minimum} , \tag{5-7}$$

where the residual vector, defined to have the nature of "corrections" as in previous sections,

$$\underset{1x\ (p+n)}{\hat{V}^T} = \underset{1xp}{[-\hat{S}^T} \,\vdots\, \underset{1xn}{-\hat{Z}^T}] \tag{5-8}$$

is made up of two parts - the signal at the computation points, and the random part of the observations. The weight matrix

$$P = \sigma_o^2 \begin{bmatrix} C_{ss} & C_{sx} \\ C_{xs} & C_{xx} \end{bmatrix}^{-1} , \tag{5-9}$$

where C_{ss} is the variance-covariance matrix of the signal, C_{xx} the variance-covariance matrix of the observable. C_{sx} and C_{xs} are the

covariance matrices between the signal and the observable, and where the a priori variance factor σ_o^2 is set to unity for convenience.

Remember that in collocation the observable has two random parts - the signal S' and the noise n (equation 5-4). Accordingly the covariance matrix

$$\begin{aligned} C_{xx} &= COV(x,x) = M\{Z\,Z^T\} = M\{(S' + n)\,(S'^T + n^T)\} \\ &= M\{S'\,S'^T + n\,S'^T + S'\,n^T + n\,n^T\} \\ &= M\{S'\,S'^T\} + M\{n\,S'^T\} + M\{S'\,n^T\} + M\{nn^T\} \\ &= C_{s's'} + C_{nn} = C + D\ , \end{aligned} \tag{5-10}*$$

after one assumes that the measuring error (n) has no correlation with the signal (S') at each observation point. Under this assumption

$$\begin{aligned} C_{sx} &= M\{SZ^T\} = M\{S(S' + n)^T\} = M\{SS'^T\} + M\{Sn^T\} \\ &= M\{SS'^T\} \\ &= COV\,(S,\,S')\ , \end{aligned} \tag{5-11}$$

and

$$C_{xs} = COV\,(S',\,S) \tag{5-12}$$

are pure signal covariances which describe the correlation between the signal components in the domain of the "observation" and "computation" points.

The above minimum is to be found subject to the constraint function

$$AX + BV + W = 0\ , \tag{5-13}$$

where A is the nxu design matrix of equation 5-3; X the u vector of unknown parameters of equation 5-3; V is the n+p vector of equation 5-8; and the newly introduced quantities

---

* M stands for mathematical expectation - sometimes denoted by E.

$$\underset{nx(n+p)}{B} = [\ \underset{nxp}{0}\ \vdots\ \underset{nxn}{-I}\ ] \tag{5-14}$$

is the second design matrix consisting of a null and minus identity matrix, and

$$\underset{nxl}{W} = \underset{nxl}{-x} \tag{5-15}$$

is the vector of measurements. In the case that we choose to solve for corrections to some approximate (or observed) parameters, then

$$W^\circ = AX^\circ - x\ , \tag{5-16}$$

or

$$W = AX - x\ . \tag{5-16a}$$

The purpose of introducing the null matrix in the above was to involve the signals S (to be predicted) in the equations without modifying the original mathematical model given by equation 5-5. To corroborate this, we substitute B, V and W above into equation 5-13:

$$A\bar{X} + [0\,\vdots\,-I]\begin{bmatrix} -S \\ \cdots \\ -Z \end{bmatrix} - x = 0 \tag{5-17}$$

$$A\bar{X} + Z = x \tag{5-18}$$

which is equation 5-5.

5.3 Collocation Equations

The least squares normal equations relating the unknown quantities X and S to the known quantities A, x, C_{ss}, C_{xx} and C_{sx} are obtained from the variation function

$$\phi = \hat{V}^T P \hat{V} + \hat{X}^T P_x \hat{X} + 2\hat{K}^T\ (A\hat{X} + B\hat{V} + W)\ , \tag{5-19}$$

where all quantities are defined immediately above and P_x is the weight

matrix on the parameters as before. We recognize this to be the variation function of the standard-combined case with weighted parameters. The equations corresponding to the above have been derived in detail in Section 2 (equations 2-74 to 2-94). We now specialize these equations to the collocation problem.

Solution for the Parameters

The solution for the parameters is (equation 2-76)

$$\hat{\bar{X}} = X + \hat{X} , \tag{5-20}$$

where X is the vector of weighted parameters, while the correction vector is given by (2-74)

$$\hat{X} = - [A^T (BP^{-1}B^T)^{-1} A + P_x]^{-1} A^T (BP^{-1}B^T)^{-1} W. \tag{5-21}$$

In the above

$$BP^{-1}B^T = [0 \,\vdots\, -I] \begin{bmatrix} C_{ss} & C_{sx} \\ C_{xs} & C_{xx} \end{bmatrix} \begin{bmatrix} 0 \\ \hdashline -I \end{bmatrix} = C_{xx} . \tag{5-22}$$

Using equation 5-10 and the above,

$$\boxed{\hat{X} = - [A^T (C+D)^{-1} A + P_x]^{-1} A^T (C+D)^{-1} W.} \tag{5-23}$$

Note that the variance-covariance matrix for the signal at the observation points (C) and measurement error (D), and the weight matrix for the parameters (P_x) enter as three separate pieces of information. Also note that the covariance matrix C_{xs} needed for the prediction of the signal does not affect the solution of the parameters X since it does not appear in the above equation.

For the case of weighted parameters

$$W = AX - x \ , \tag{5-24}$$

where X is the quasi-observed value of the parameters with weight matrix P_x, and x are the measurements. For the unweighted case

$$\boxed{\hat{X} = - [A^T(C+D)^{-1}A]^{-1}A^T(C+D)^{-1}W \ ,} \tag{5-25}$$

where the misclosure vector W becomes:

$$W = -x \ , \tag{5-26}$$

or

$$W = W^O = AX^O - x \tag{5-27}$$

depending upon whether one wishes to solve for the parameters themselves or corrections to some approximate value X^O.

Solution for the Signal at the Computation Points

The expression for computing the signal at the computation points follows from the equation for the adjusted observations (equation 2-80) and residual vector (equation 2-79), namely

$$\bar{L} = L + \hat{V} \ , \tag{5-28}$$

where

$$\hat{V} = -P^{-1} B^T \hat{K} \tag{5-29}$$

and

$$\hat{K} = (BP^{-1} B^T)^{-1} (A\hat{X} + W) \ . \tag{5-30}$$

The observed value of the signal can be taken as zero as per the condition imposed upon this quantity - namely it is a random variable with zero mean. Hence the solution for the signal is given by the residual vector. Noting that

$$-P^{-1}B^T = -\begin{bmatrix} C_{ss} & C_{sx} \\ C_{xs} & C_{xx} \end{bmatrix}\begin{bmatrix} 0 \\ -I \end{bmatrix} = \begin{bmatrix} C_{sx} \\ C_{xx} \end{bmatrix}, \qquad (5\text{-}31)$$

and using equation 5-22

$$\hat{\underline{L}} = 0 + \hat{V} = \begin{bmatrix} -\hat{\underline{s}} \\ -\hat{\underline{z}} \end{bmatrix} = \begin{bmatrix} C_{sx} \\ C_{xx} \end{bmatrix} C_{xx}^{-1} (A\hat{X} + W), \qquad (5\text{-}32)$$

and finally the signal is estimated from:

$$\hat{\underline{s}} = -C_{sx} C_{xx}^{-1} (A\hat{X} + W) \qquad (5\text{-}33)$$

$$= -C_{sx} (C+D)^{-1} (A\hat{X} + W) . \qquad (5\text{-}34)$$

The collocation solution does not contain an equation for Z, however one can get from equation 5-32 the required expression:

$$\hat{\underline{z}} = -C_{xx}C_{xx}^{-1} (A\hat{X} + W) = -(A\hat{X} + W) . \qquad (5\text{-}34a)$$

The problem of spliting $\hat{\underline{z}}$ into signal $\hat{s}'$ and noise $\hat{n}$ estimates is the subject of Section 5.5.

Variance-Covariance Matrix for the Parameters

The variance-covariance matrix for the final (adjusted) parameters

$$\hat{\underline{X}} = X + \hat{X} \qquad (5\text{-}35)$$

is, using equations 2-85,

$$\Sigma_{\hat{\underline{X}}} = \sigma_o^2 Q_{\hat{\underline{X}}} , \qquad (5\text{-}36)$$

$$= \sigma_o^2 [A^T (BP^{-1}B^T)^{-1} A + P_X]^{-1} , \qquad (5\text{-}37)$$

where σ_o^2 is the usual a priori variance factor. Considering the collocation definitions of A, B (equation 5-14), and P (equation 5-9 and 5-10) we get

$$\Sigma_{\hat{X}} = \sigma_o^2 Q_{\hat{X}} \tag{5-38}$$

$$\boxed{\Sigma_{\hat{X}} = \sigma_o^2 [A^T(C+D)^{-1}A + P_x]^{-1} .} \tag{5-39}$$

In the case that no weights are applied to the parameters

$$\boxed{\Sigma_{\hat{X}} \equiv \Sigma_{\hat{X}} = \sigma_o^2 [A^T (C+D)^{-1}A]^{-1} ,} \tag{5-40}$$

which is identical to the collocation equation.

Variance-Covariance Matrix of the Signal at the Computation Points

We have seen above that minus the signal at a computation point is like an adjusted observation because an adjusted observation equals the observation plus the residual, and in collocation this equals zero (for the signal) plus, minus the signal (see equations 5-8 and 5-28). Thus the variance-covariance matrix for the signal follows from that of the adjusted observations (equation 2-36), namely

$$\Sigma_{\hat{L}} = \sigma_o^2 Q_{\hat{L}} ,$$

where

$$Q_{\hat{L}} = P^{-1} + P^{-1}B^T(BP^{-1}B^T)^{-1}AQ_{\hat{X}}A^T(BP^{-1}B^T)^{-1}BP^{-1} - P^{-1}B^T(BP^{-1}B^T)^{-1}BP^{-1} \tag{5-41}$$

and where $Q_{\hat{X}}$ is given above in equations 5-38 and 5-39. Since from adjustments

$$\hat{L} = L + \hat{V} , \tag{5-42}$$

noting L is a null vector in collocation,

thus

$$\hat{\underset{\sim}{L}} = 0 + \hat{v} = \begin{bmatrix} -\hat{s} \\ -- \\ -\hat{z} \end{bmatrix} , \tag{5-43}$$

then

$$\Sigma_{\hat{L}} = \sigma_o^2 Q_{\hat{L}} = \left[\begin{array}{c|c} \Sigma_{\hat{s}} & \Sigma_{\hat{s}\hat{z}} \\ \hline \Sigma_{\hat{z}\hat{s}} & \Sigma_{\hat{z}} \end{array}\right] . \tag{5-44}$$

It is $\Sigma_{\hat{s}}$ that we are after. We now specialize equation 5-41 as before by using the definitions of the various terms

$$\Sigma_{\hat{L}} = \left[\begin{array}{c|c} \Sigma_{\hat{s}} & \Sigma_{\hat{s}\hat{z}} \\ \hline \Sigma_{\hat{z}\hat{s}} & \Sigma_{\hat{z}} \end{array}\right]$$

$$= \left[\begin{array}{c|c} C_{ss} & C_{sx} \\ \hline C_{xs} & C_{xx} \end{array}\right] + \begin{bmatrix} C_{sx} \\ -- \\ C_{xx} \end{bmatrix} C_{xx}^{-1} A \Sigma_{\hat{X}} A^T C_{xx}^{-1} [C_{xs} \mid C_{xx}]$$

$$- \begin{bmatrix} C_{sx} \\ -- \\ C_{xx} \end{bmatrix} C_{xx}^{-1} [C_{xs} \mid C_{xx}] . \tag{5-45}$$

Picking out only the upper left portion from the above hyper-matrix equation, one obtains

$$\Sigma_{\hat{s}} = C_{ss} + C_{sx} C_{xx}^{-1} A \Sigma_{\hat{X}} A^T C_{xx}^{-1} C_{xs} - C_{sx} C_{xx}^{-1} C_{xs} \tag{5-46}$$

$$= C_{ss} + C_{sx} (C+D)^{-1} A \Sigma_{\hat{X}} A^T (C+D)^{-1} C_{xs} - C_{sx} (C+D)^{-1} C_{xs} , \tag{5-47}$$

which is identical to the collocation equation.

Let us now examine the expressions for the parameters (equation 5-25), signal equation (5-34) and their respective variance-covariance matrices (equations 5-40 and 5-47) to determine the number, form and size of matrices to be inverted. The solution and variance-covariance matrix for the parameters requires two matrix inversions in

$$[A^T(C+D)^{-1}A]^{-1} \quad . \tag{5-48}$$

The first inversion is a fully populated matrix because of the correlation among the signals at the computation points and has dimensions equal to the number of observations. In certain applications this matrix can be reduced to a band matrix and then by using special inversion tactics, like "compacting out the zero parts" the inversion is made faster, e.g. [Krakiwsky and Pope 1967; Isner 1972; Knight and Steeves 1974]. The second inversion is of a fully populated matrix of order equal to the number of parameters. Equations 5-34 and 5-47 (pertaining to the signal) require no additional inverses to that made in the determination of the parameters.

5.4 Stepwise Collocation

We have seen that collocation equations follow directly from the equations of the standard-combined case of adjustments simply by specifying the collocation forms of the design matrices A, B, and weight matrix P. On the other hand, we cannot deduce the sequential collocation equations of Moritz [1973a] from the Kalman (sequential) expressions derived in Section 3 and 4. This is because in the latter development the observations of consecutive stages are assumed to be uncorrelated.

This assumption is prohibative in collocation as the signal part of the observation has a continuous and complete correlation which must be accounted for when grouping the observations in stages. Thus one sees that from this point of view, collocation is a more general method. Below we simply state the sequential collocation equations from Moritz. Recall the collocation form for the solution of the parameters (equations 5-25 and 5-26); namely

$$\hat{X} = [A^T C_{xx}^{-1} A]^{-1} A^T C_{xx}^{-1} x \ . \tag{5-49}$$

In sequential collocation the matrices and vectors are partitioned in two, such that

$$A = \begin{bmatrix} A_1 \\ \hline A_2 \end{bmatrix} , \tag{5-50}$$

$$x = \begin{bmatrix} x_1 \\ \hline x_2 \end{bmatrix} , \tag{5-51}$$

$$C_{xx} = \left[\begin{array}{c|c} C_{11} & C_{12} \\ \hline C_{21} & C_{22} \end{array}\right] . \tag{5-52}$$

The solution for the parameters is given by

$$\boxed{\hat{X} = \hat{X}_1 + P_1^{-1} \bar{A}_2^T \bar{\bar{C}}_{22}^{-1} (x_2 - C_{21} C_{11}^{-1} x_1 - \bar{A}_2 \hat{X}_1),} \tag{5-53}$$

where

$$P_1^{-1} = [A_1^T C_{11}^{-1} A_1]^{-1} , \tag{5-54}$$

$$\bar{A}_2 = A_2 - C_{21} C_{11}^{-1} A_1 , \tag{5-55}$$

$$\bar{\bar{C}}_{22}^{-1} = [C_{22} - C_{21} C_{11}^{-1} C_{12} + \bar{A}_2 P_1^{-1} \bar{A}_2^T]^{-1} , \tag{5-56}$$

and $\hat{X}_1$ is the solution using only observations x_1, that is

$$\hat{X}_1 = (A_1^T C_{11}^{-1} A_1)^{-1} A_1^T C_{11}^{-1} x_1 . \tag{5-57}$$

The solution for a signal element in two steps, is given by

$$\boxed{\hat{S}_P = \hat{S}_1 + (C_{P2} - C_{P1} C_{11}^{-1} C_{12} - C_{P1} C_{11}^{-1} A_1 P_1^{-1} \bar{A}_2^T) \bar{C}_{22}^{-1} (x_2 - C_{21} C_{11}^{-1} x_1 - \bar{A}_2 \hat{X}_1)} , \tag{5-58}$$

where the newly introduced terms are defined as

$$C_P^T = [C_{P1} \,\vdots\, C_{P2}] , \tag{5-59}$$

and $\hat{S}_1$ is the signal computed using only observations x_1, that is

$$\hat{S}_1 = C_{P1} C_{11}^{-1} (x_1 - A_1 \hat{X}_1) . \tag{5-60}$$

C_P^T is a row vector of dimensions $1x(n_1+n_2)$ [Moritz 1973a]; here only one signal element ($\hat{S}_P$) is estimated. However, the same formula could be applied for all p signal elements (vector $\hat{S}$) by using

$$\underset{px(n_1+n_2)}{C_{sx}} = [\underset{pxn_1}{C_{s1}} \,\vdots\, \underset{pxn_2}{C_{s2}}]$$

instead of C_P.

The variance-covariance matrix for the parameters is

$$\boxed{\Sigma_{\hat{x}} = \Sigma_{\hat{x}_1} - P_1^{-1} \bar{A}_2^T \bar{C}_{22}^{-1} \bar{A}_2 P_1^{-1}} , \tag{5-61}$$

where

$$\Sigma_{\hat{X}_1} = P_1^{-1} . \tag{5-62}$$

In the above sequential collocation expressions, one sees that one matrix inversion is necessary, namely that of $\bar{C}_{22}$ which has dimensions equal to the number of observations in the second stage. It is interesting to note how the correlation among all the signal elements is accounted for even though the full matrix

$$C_{xx} = C+D = \bar{C} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \tag{5-63}$$

is not explicitly inverted. We stress that it is implicitly inverted thereby giving rise to expressions like equations 5-55 and 5-56. Thus we see how the correlation among the signal elements, which is essential for collocation, is accounted for in the sequential expressions.

5.5 Alternative Derivation of the Collocation Equations

There are several alternative ways we can derive the collocation equations. These alternatives follow from the number of ways we can arrange the parameters X, signal S and S', and the residuals (correction to the observations) V. This arrangement can take place in the form of an "association" or in terms of an "association"and a "combination" (see Table 5-1).

In the linearized mathematical model

$$AX + BV + W = 0 \ , \tag{5-64}$$

we usually identify the term AX with the parameters (weighted or not weighted) and the term BV with the observables (weighted). This leaves us with the possibility of associating the signal S and S' with either of the two terms - this leads to alternatives I and II (Table 5-1). If there are weights for the parameters as well we have alternative III - a pure condition case adjustment. Note we have not split the signal, say S to term AX and S' to term BV, since they are statistically dependent by nature.

Table 5-1. Alternatives for the Derivation of the Collocation Equations.

Alternatives	term AX	term BV
(association)		
I	X, S, S'	V (minus noise)
II*	X	V, S, S'
III		V, S, S', X
(association and combination)		
IV**	X	$Z = S' + (-V)$, S
V		Z, S, X

* choice for alternative derivation of collocation equations

** "original" collocation derivation

Alternatives IV and V arise as a result of combining the signal S' (at observation points only) with the noise (-V) into one quantity called Z, and first associating X with term AX, and Z and S with term BV, and secondly associating all quantities with term BV. Of course, the last alternative is only possible if weights for the parameters exist.

We recognize alternative IV as being the form of the "original" collocation derivation given in the previous section. We choose alternative II over the others for the alternative derivation for no other reason than that the collocation equations can be deduced in minimum space and time.

We begin our alternative derivation by assuming a non-linear implicit mathematical model (equation 2-6).

$$F(\bar{X}, \bar{L}) = 0 \ , \tag{5-65}$$

and after linearization (equation 2-16)

$$A\hat{X} + B^*\hat{V}^* + W = 0, \ P_x, \ P^* \tag{5-66}$$

where the asterisk (*) denotes a hyper matrix or vector. In the above we have the following definitions: the residual vector

$$\hat{V}^* = \begin{bmatrix} -\hat{S} \\ -\hat{S}' \\ \hat{V} \end{bmatrix} \ ; \tag{5-67}$$

the second design matrix

$$B^* = [0 \,|\, -I \,|\, -I]; \tag{5-68}$$

the misclosure vector

$$W = F\ (X, \ L) \ ; \tag{5-69}$$

---

† At this juncture we could have chosen a non-identity coefficient matrix for the signal and noise, but by not doing so we will arrive exactly at the collocation equations.

the a priori weight matrix for the parameters

$$P_x = \sigma_o^2 \Sigma_x^{-1} = \Sigma_x^{-1} , \qquad (5\text{-}70)^{\dagger\dagger}$$

where Σ_x is the a priori variance-covariance matrix for the parameters;

$$P^* = \sigma_o^2 \begin{bmatrix} \Sigma_S & \Sigma_{SS'} & 0 \\ \Sigma_{SS'} & \Sigma_{S'} & 0 \\ 0 & 0 & \Sigma_L \end{bmatrix}^{-1} = \begin{bmatrix} \begin{bmatrix} \Sigma_S & \Sigma_{SS'} \\ \Sigma_{SS'} & \Sigma_{S'} \end{bmatrix}^{-1} & 0 \\ 0 & \Sigma_L^{-1} \end{bmatrix}, \qquad (5\text{-}71)^{\dagger\dagger}$$

where Σ_L is the variance-covariance matrix of the observations. Note that X is assumed to be statistically independent from S and S'. This is a reasonable assumption as the a priori estimates for the parameters are understood to be estimates from an other outside source of information. The signal quantities are statistically dependent by nature - that is the signal is characterized by one covariance matrix in the domain defined by the observation and computation points. On the other hand, the measurement errors are assumed to be statistically independent from S and S' as they are peculiar to the measuring instrument.

We also note that in the evaluation of W, only the observed value of the observables (L) and the quasi-observed value[†] of the parameters (X) appear; the signal components do not appear as their expected values (observed values) are zero by definition.

The variation function from which the least squares estimates are deduced is

---

† or simply approximate values X^o, thus W^o

†† $\sigma_o^2 = 1$ for sake of convenience.

$$\phi = \hat{X}^T P_x \hat{X} + [\hat{S}^T \mid \hat{S}'^T] \begin{bmatrix} \Sigma_S & \Sigma_{SS'} \\ \Sigma_{S'S} & \Sigma_{S'} \end{bmatrix}^{-1} \begin{bmatrix} \hat{S} \\ \hat{S}' \end{bmatrix} + \hat{V}^T P \hat{V}$$

$$+ 2K^T (A\hat{X} + B^*V^* + W) \quad . \tag{5-72}$$

There are three quadratic forms to be minimized, the first involving the parameters, the second the two signal components, and third the observational corrections. In this formulation we see that the observational corrections do appear explicitly in the quadratic form and in the constraint function. This fact has both an advantage and a disadvantage as we will witness below.

Our model (equation 5-66) fits into the standard combined case - weighted parameters category (Section 2.4). We proceed to utilize the expressions contained therein (equations 2-74 through 2-97) to obtain the collocation equations.

Solution for the Parameters $\hat{X}$

From equation 2-74, the solution for the parameters $\hat{X}$ is given by

$$\hat{X} = -[A^T(B^*P^{*-1}B^{*T})^{-1}A + P_x]^{-1}A^T(B^*P^{*-1}B^{*T})^{-1}W. \tag{5-73}$$

We specialize the above expression to the problem at hand by use of equations 5-67 to 5-71 and get the following:

$$P^{*-1} = \begin{bmatrix} \Sigma_S & \Sigma_{SS'} & 0 \\ \Sigma_{S'S} & \Sigma_{S'} & 0 \\ 0 & 0 & \Sigma_L \end{bmatrix}; \tag{5-74}$$

$$B^*P^{*-1}B^{*T} = [0 \,|\, -I \,|\, -I] \left[\begin{array}{ccc|c} \Sigma_S & \Sigma_{SS'} & 0 & 0 \\ \Sigma_{S'S} & \Sigma_{S'} & 0 & -I \\ 0 & 0 & \Sigma_L & -I \end{array}\right] ; \tag{5-75}$$

$$(B^*P^{*-1}B^{*T})^{-1} = (\Sigma_{S'} + \Sigma_L)^{-1} . \tag{5-76}$$

Thus

$$\boxed{\hat{X} = -[A^T(\Sigma_{S'} + \Sigma_L)^{-1}A + P_x]^{-1}A^T(\Sigma_{S'} + \Sigma_L)^{-1} W ,} \tag{5-77}$$

and in collocation notation

$$\boxed{\hat{X} = -[A^T(C+D)^{-1} A + P_x]^{-1} A^T (C+D)^{-1} W ,} \tag{5-78}$$

we find that the above equation is equivalent to equation 5-23 .

Solution for the Signal $\hat{\hat{S}}$

From equation 2-79

$$\hat{V}^* = -P^{*-1} B^{*T} \hat{K} , \tag{5-79}$$

where

$$\hat{K} = (B^*P^{*-1} B^{*T})^{-1} (A\hat{X} + W) . \tag{5-80}$$

Specializing the above as before we get

$$-P^{*-1}B^{*T} = -\left[\begin{array}{cc|c} \Sigma_S & \Sigma_{SS'} & 0 \\ \Sigma_{S'S} & \Sigma_{S'} & 0 \\ \hline 0 & 0 & \Sigma_L \end{array}\right]\left[\begin{array}{c} 0 \\ -I \\ -I \end{array}\right] = \left[\begin{array}{c} \Sigma_{SS'} \\ \Sigma_{S'} \\ \Sigma_L \end{array}\right] . \tag{5-81}$$

Substituting the above, along with equation 5-76, into equations 5-79 and 5-80 yields

$$\hat{\bar{L}} = 0 + \hat{V}^* = \begin{bmatrix} -\hat{\bar{S}} \\ \hline -\hat{\bar{S}}' \\ \hline \hat{V} \end{bmatrix} = \begin{bmatrix} \Sigma_{SS'} \\ \hline \Sigma_{S'} \\ \hline \Sigma_{L} \end{bmatrix} (\Sigma_{S'}+\Sigma_{L})^{-1}(A\hat{X}+W) \quad . \tag{5-82}$$

From the first row of the above hyper-matrix equation we get an expression for the signal at the computation points, namely

$$\boxed{\hat{\bar{S}} = -\Sigma_{SS'} (\Sigma_{S'} + \Sigma_{L})^{-1}(A\hat{X} + W) \ ,} \tag{5-83}$$

and in collocation notation

$$\boxed{\hat{\bar{S}} = -C_{SX} (C+D)^{-1} (A\hat{X} + W) \ ,} \tag{5-84}$$

which is identical to equation 5-34.

Solution for the Signal $\hat{\bar{S}}'$

From the second row of equation (5-82) we get an expression for the signal at the observation points, namely

$$\boxed{\hat{\bar{S}}' = -\Sigma_{S'} (\Sigma_{S'}+\Sigma_{L})^{-1} (A\hat{X}+W) \ ,} \tag{5-85}$$

and in collocation notation

$$\boxed{\hat{\bar{S}}' = -C (C + D)^{-1} (A\hat{X} + W) \ .} \tag{5-86}$$

There is no equivalent expression in the "original" collocation derivation (See equation 5-34a).

Solution for the Observational Correction $\hat{V}$

From the third row of equation (5-82) we get an expression for the observational correction, namely

$$\boxed{\hat{V} = \Sigma_{L} (\Sigma_{S'} + \Sigma_{L})^{-1} (A\hat{X} + W) \ ,} \tag{5-87}$$

and in collocation notation

$$\boxed{\hat{V} = D(C + D)^{-1} (A\hat{X} + W) .} \tag{5-88}$$

There is no equivalent expression in the "original" collocation derivation (see equation 5-34a). We can deduce an expression for the combined quantity $\hat{\bar{Z}}$ (equation 5-34a) by addition of equations 5-88 and 5-86, namely

$$\boxed{\hat{\bar{Z}} = \hat{S}' + (-\hat{V}) = -(C+D)(C+D)^{-1}(A\hat{X}+W) = -(A\hat{X}+W) .} \tag{5-89}$$

Variance-Covariance Matrix for the Parameters

The variance-covariance matrix for the final (adjusted) parameters

$$\hat{\bar{X}} = X + \hat{X} \tag{5-90}$$

is, using equation 2-85,

$$\Sigma_{\hat{\bar{X}}} = \sigma_o^2 \, Q_{\hat{\bar{X}}} \tag{5-91}$$

$$= \sigma_o^2 \, [A^T(B^*P^{*-1}B^{*T})^{-1}A + P_x]^{-1} . \tag{5-92}$$

After considering the definitions of B* and P* we get

$$\boxed{\Sigma_{\hat{\bar{X}}} = \sigma_o^2 \, [A^T(\Sigma_{S'} + \Sigma_L)^{-1}A + P_x]^{-1} .} \tag{5-93}$$

After deleting weights on parameters and changing notation we get

$$\boxed{\Sigma_{\hat{\bar{X}}} = \sigma_o^2 \, [A^T(C+D)^{-1}A]^{-1} ,} \tag{5-94}$$

which is identical to the collocation expression (equation 5-40).

Variance-Covariance Matrix for the Signal of the Computation Points

The variance-covariance matrix for the signal components follows from the corresponding expression for the adjusted observables. The arguments why this is so have already been given in Section 5.3. From equation 2-86, the weight coefficient matrix for the adjusted observables is

$$Q_{\hat{\underline{L}}^*} = P^{*-1} + P^{*-1}B^{*T}\,(B^*P^{*-1}B^{*T})^{-1}A\,Q_{\hat{\underline{X}}}\,A^T(B^*P^{*-1}B^{*T})^{-1}B^*P^{*-1}$$
$$- P^{*-1}B^{*T}\,(B^*P^{*-1}B^{*T})^{-1}\,B^*P^{*-1}\,, \qquad (5\text{-}95)$$

where $Q_{\hat{\underline{X}}}$ is defined in equation 5-91. Since

$$\hat{\underline{L}}^* = L^* + \hat{V}^* = \begin{bmatrix} 0 \\ 0 \\ L \end{bmatrix} + \begin{bmatrix} -\bar{S} \\ -\bar{S}' \\ \hat{V} \end{bmatrix} \qquad (5\text{-}96)$$

$$Q_{\hat{\underline{L}}^*} = \begin{bmatrix} Q_{\hat{\underline{S}}} & Q_{\hat{\underline{S}}\hat{\underline{S}}'} & Q_{\hat{\underline{S}}\hat{\underline{L}}} \\ Q_{\hat{\underline{S}}'\hat{\underline{S}}} & Q_{\hat{\underline{S}}'} & Q_{\bar{S}'\hat{\underline{L}}} \\ Q_{\hat{\underline{L}}\hat{\underline{S}}} & Q_{\hat{\underline{L}}\hat{\underline{S}}'} & Q_{\hat{\underline{L}}} \end{bmatrix} \qquad (5\text{-}97)$$

and from equation 2-93

$$\Sigma_{\hat{\underline{L}}} = \sigma_o^2 Q_{\hat{\underline{L}}}\,. \qquad (5\text{-}98)$$

The three diagonal terms are the variance-covariance matrices for the signal at the computation points, the signal at the observation points, and the adjusted observations, respectively. The off-diagonal terms are the covariances between them.

Let us now specialize equation 5-95 as before by using the definitions of the various terms.

$$\Sigma_{\hat{\underline{L}}^*} = \begin{bmatrix} \Sigma_{\hat{\underline{S}}} & \Sigma_{\hat{\underline{S}}\hat{\underline{S}}'} & \Sigma_{\hat{\underline{S}}\hat{\underline{L}}} \\ \Sigma_{\hat{\underline{S}}'\hat{\underline{S}}} & \Sigma_{\hat{\underline{S}}'} & \Sigma_{\hat{\underline{S}}'\hat{\underline{L}}} \\ \Sigma_{\hat{\underline{L}}\hat{\underline{S}}} & \Sigma_{\hat{\underline{L}}\hat{\underline{S}}'} & \Sigma_{\hat{\underline{L}}} \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_S & \Sigma_{SS'} & 0 \\ \Sigma_{S'S} & \Sigma_{S'} & 0 \\ 0 & 0 & \Sigma_L \end{bmatrix} + \begin{bmatrix} \Sigma_{SS'} \\ \Sigma_{S'} \\ \Sigma_L \end{bmatrix} (\Sigma_{S'}+\Sigma_L)^{-1} A\Sigma_{\hat{\underline{X}}} A^T (\Sigma_{S'}+\Sigma_L)^{-1} \; [\Sigma_{S'S} \Sigma_{S'} \Sigma_L]$$

$$- \begin{bmatrix} \Sigma_{SS'} \\ \Sigma_{S'} \\ \Sigma_L \end{bmatrix} (\Sigma_{S'} + \Sigma_L)^{-1} \; [\Sigma_{S'S} \; \Sigma_{S'} \; \Sigma_L] \; . \qquad (5\text{-}99)$$

From the (one, one) position of the hyper matrix equation we obtain the variance-covariance matrix for the signal of the computation points, namely

$$\boxed{\Sigma_{\hat{\underline{S}}} = \Sigma_S + \Sigma_{SS'}(\Sigma_{S'}+\Sigma_L)^{-1} A\Sigma_{\hat{\underline{X}}} A^T (\Sigma_{S'}+\Sigma_L)^{-1} \Sigma_{S'S} - \Sigma_{SS'}(\Sigma_{S'}+\Sigma_L)^{-1}\Sigma_{S'S}} \qquad (5\text{-}100)$$

and in collocation notation

$$\boxed{\Sigma_{\hat{\underline{S}}} = C_{ss} + C_{sx}(C+D)^{-1} A\Sigma_{\hat{\underline{X}}} A^T (C+D)^{-1} C_{xs} - C_{sx}(C+D)^{-1} C_{xs} \; ,} \qquad (5\text{-}101)$$

which is identical to the "original" collocation expression (equation 5-47).

Variance-Covariance Matrix for the Signal $\hat{\underline{S}}'$ at the Observation Points

From the (two, two) position of the above hyper matrix equation we get the desired expression for $\hat{\underline{S}}'$, namely

$$\boxed{\Sigma_{\hat{\underline{S}}'} = \Sigma_{S'} + \Sigma_{S'}(\Sigma_{S'}+\Sigma_L)^{-1} A\Sigma_{\hat{\underline{X}}} A^T (\Sigma_{S'}+\Sigma_L)^{-1} \Sigma_{S'} - \Sigma_{S'}(\Sigma_{S'}+\Sigma_L)^{-1}\Sigma_{S'} \; ,} \qquad (5\text{-}102)$$

which has no equivalent in the "original" collocation formulation.

Variance-Covariance Matrix for the Adjusted Observation $\hat{\bar{L}}$

From the (three, three) position of the above hyper-matrix equation we get the desired expression for $\hat{\bar{L}}$, namely

$$\boxed{\Sigma_{\hat{\bar{L}}} = \Sigma_L + \Sigma_L(\Sigma_{S'}+\Sigma_L)^{-1}A\Sigma_{\hat{\bar{X}}}A^T(\Sigma_{S'}+\Sigma_L)^{-1}\Sigma_L - \Sigma_L(\Sigma_{S'}+\Sigma_L)^{-1}\Sigma_L,} \quad (5\text{-}103)$$

which has no equivalent in the "original" collocation formulation.

Variance-Covariance Matrix for the Observational Corrections $\hat{V}$

From equations 2-87 and 2-94

$$\begin{aligned}\Sigma_{\hat{V}} &= \sigma_o^2 Q_{\hat{V}}, \quad (\sigma_o^2 = 1)\\ &= -P^{-1}B^T(BP^{-1}B^T)^{-1}A\Sigma_{\hat{\bar{X}}}A^T(BP^{-1}B^T)^{-1}B\,P^{-1}\\ &\quad + P^{-1}B^T(BP^{-1}B^T)^{-1}BP^{-1}.\end{aligned} \quad (5\text{-}104)$$

Specializing by using the definitions of B and P^{-1}, we get

$$\boxed{\Sigma_{\hat{V}} = -\Sigma_L(\Sigma_{S'}+\Sigma_L)^{-1}A\Sigma_{\hat{\bar{X}}}A^T(\Sigma_{S'}+\Sigma_L)^{-1}\Sigma_L + \Sigma_L(\Sigma_{S'}+\Sigma_L)^{-1}\Sigma_L.} \quad (5\text{-}105)$$

Covariance Between Signal $\hat{\bar{S}}$ and Signal $\hat{\bar{S}}'$

From the (one, two) position of the hyper-matrix equation we get

$$\boxed{\Sigma_{\hat{\bar{S}}\hat{\bar{S}}'} = \Sigma_{SS'} + \Sigma_{SS'}(\Sigma_{S'}+\Sigma_L)^{-1}A\Sigma_{\hat{\bar{X}}}A^T(\Sigma_{S'}+\Sigma_L)^{-1}\Sigma_{S'} - \Sigma_{SS'}(\Sigma_{S'}+\Sigma_L)^{-1}\Sigma_{S'},} \quad (5\text{-}106)$$

and in collocation notation

$$\Sigma_{\hat{S}\hat{S}'} = C_{sx} + C_{sx}(C+D)^{-1} A\Sigma_{\hat{X}}A^T (C+D)^{-1}C - C_{sx}(C+D)^{-1} C \ . \qquad (5\text{-}107)$$

Covariance Between $\hat{\bar{S}}$ and $\hat{\bar{L}}$.

The covariance matrix between the predicted signal $\hat{S}$ and the adjusted observation $\hat{\bar{L}}$ follows from the (one, three) position, namely

$$\Sigma_{\hat{S}\hat{L}} = \Sigma_{SS'}(\Sigma_{S'}+\Sigma_L)^{-1}A\Sigma_{\hat{X}}A^T(\Sigma_{S'}+\Sigma_L)^{-1}\Sigma_L - \Sigma_{SS'}(\Sigma_{S'}+\Sigma_L)^{-1}\Sigma_L \ . \qquad (5\text{-}108)$$

Covariance Between $\hat{\bar{S}}'$ and $\hat{\bar{L}}$

The covariance matrix between the signal $\hat{S}'$ at the observation points and the adjusted observations $\hat{\bar{L}}$ follows from the (two, three) position, namely

$$\Sigma_{\hat{S}'\hat{L}} = \Sigma_{S'}(\Sigma_{S'}+\Sigma_L)^{-1} A\Sigma_{\hat{X}}A^T(\Sigma_{S'}+\Sigma_L)^{-1}\Sigma_L - \Sigma_{S'}(\Sigma_{S'}+\Sigma_L)^{-1}\Sigma_L. \qquad (5\text{-}109)$$

Closing Remarks Regarding the Alternative Derivation

This alternative derivation of the collocation equations has one basic difference from the previous derivation - that is it explicitly involves the observational corrections. These corrections appear in the minimum and constraint function of the variation function (equation 5-72), are solved for in equation 5-87, and have a variance-covariance matrix given by equation 5-105.

When working with errorless data, simply delete all quantities related to the observational corrections $\hat{V}$, that is Σ_L, $\hat{V}$, $\Sigma_{\hat{V}}$, $\hat{\bar{L}}$, and covariance matrices involving $\hat{\bar{L}}$ vanish. This does not affect the expressions.

6. ANALYSIS OF METHODS

In this section we analyze the standard cases of adjustment, the step-by-step procedures, and the method of collocation. We begin by recalling the various methods and discuss their derivations from a general point of view (Section 6.1). This is followed by a detailed comparison of the characteristics of, and assumptions underlying the methods (Section 6.2). Then the methods are labelled according to terms used in the literature (Section 6.3). Next the methods are compared from the computational point of view (Section 6.4). We close by discussing the methods derived herein along with related topics (Section 6.5).

6.1 Methods

In this section we classify the methods and trace the flow of their derivations. The so-called methods derived in the foregoing sections fall into the following three main groups (Figure 6-1): the standard cases of adjustments; the step-by-step procedures; and the collocation approach. All these methods are rigorous estimation procedures - no approximate methods are treated herein.

We first derived the equations for the standard cases of adjustments (Section 2). They serve two purposes. Firstly, in their derivation we demonstrated a methodology which was used later in the

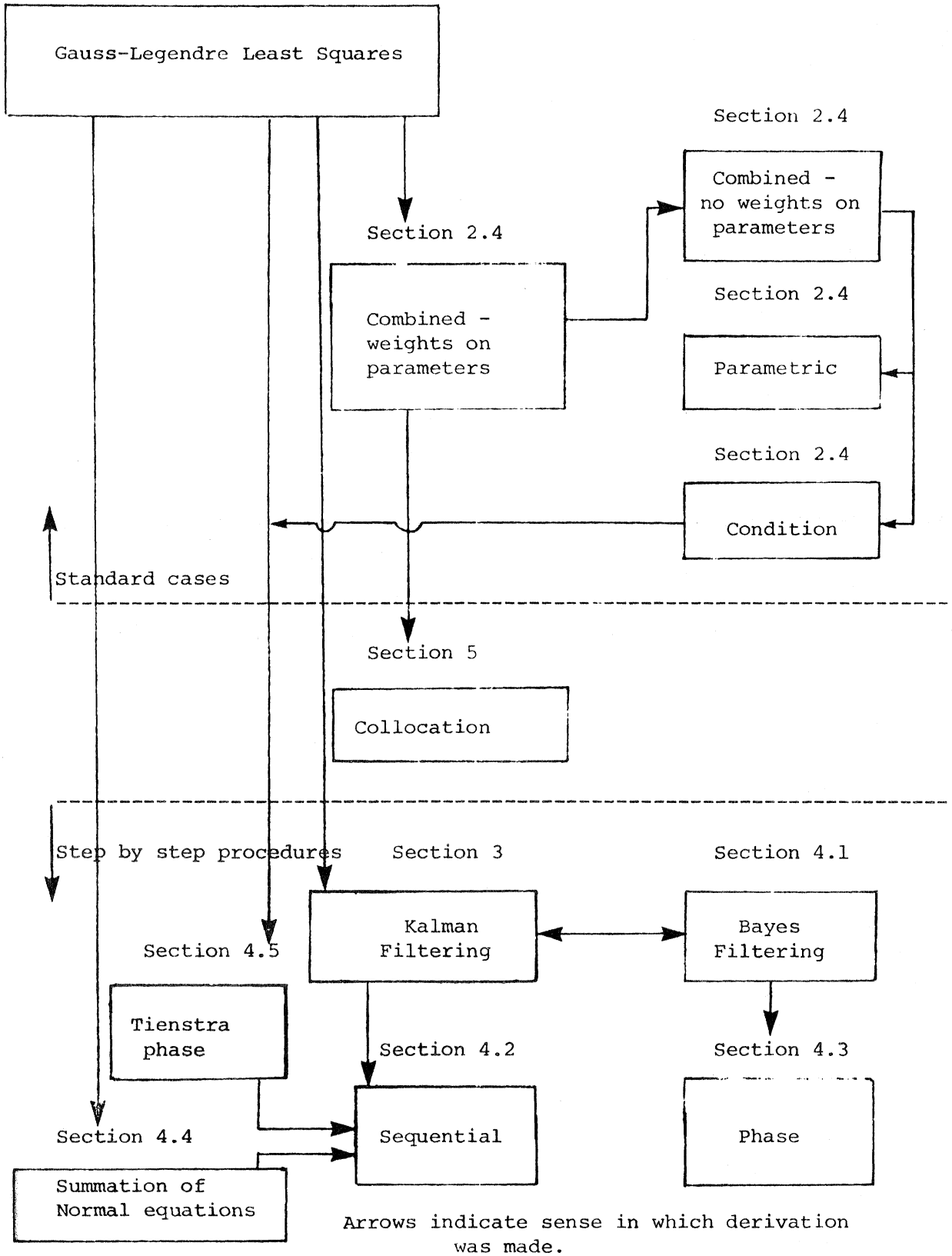


FIGURE 6-1. Trace of the Flow of Derivations

derivation of the Kalman filter equations. Secondly, the equations of the combined and condition cases respectively, lead to the collocation and Tienstra phase equations. We do not wish to imply here that collocation is a special case of the combined adjustment. We stress that it was only the collocation equations themselves that were deduced from the standard-combined case adjustment equations and not the basic mathematical model itself.

The second group of equations were derived for the step-by-step procedures (Sections 3 and 4). We began by deriving the Kalman filter equations from basic least squares using the standard methodology. Then by applying two matrix identities we were able to show the equivalence of the Bayes and Kalman filter equations. We stress that although the equations are mathematically equivalent they are not identical from the computational point of view. We return to this very important aspect in Section 6.3 below.

The sequential and phase expressions we derived respectively from the Kalman and Bayes filter expressions simply by deleting matrices and vectors pertaining to the time variation in the parameters. The mathematical equivalence of the sequential and phase expressions is a logical consequence, since the Kalman and Bayes expressions were already shown to be equivalent, but again we stress they are not identical from the computational point of view.

Finally, two other step by step procedures were formulated, the Tienstra phase and addition of normal equations. The Tienstra phase equations were derived in two steps: first the basic mathematical model was fomulated using his principle; and then the equations themselves

were obtained directly from the equations of the standard-condition adjustment case. The sequential and Tienstra phase equations were found to be identical, thus we did not have to prove their mathematical equivalence.

The summation form of the normal equations was formulated as a direct consequence of matrix partitioning. The equation for the variance-covariance matrix of the parameters was found to be identical to the corresponding phase expression. The corresponding expressions for the summation form of normal equations and the sequential approach were found to be mathematically equivalent, different in form, and different from the computational point of view. The mathematical equivalence of the two corresponding expressions for the parameters in the sequential and phase approaches follows as a logical consequence -- they are different in form and different from the computational point of view.

The third group of equations derived were those for the collocation method. Once the basic collocation mathematical model was formulated, the collocation equations followed directly from the standard-combined case of adjustments.

In this sense, collocation is identified by the unique manner in which the elements of the basic mathematical model are specified and not by the methodology used in deriving the equations for the various solution vectors and covariance matrices. Also, we have seen (Section 5.5) how to obtain the collocation equations by formulating the objectives of collocation purely in adjustment terms. This is why we show in Figure 6-1 a direct connection of collocation with ordinary least squares.

We found however, that it was not possible to derive the stepwise collocation equations from any of the standard adjustment cases, nor was it possible from the Kalman or Bayes equations. This is because, in collocation the signal part of the observations is correlated between consecutive steps, while in the other step by step procedures the observations (and model noise in the case of Kalman and Bayes filtering) are assumed to be uncorrelated between consecutive steps. Also, there are basic differences between stepwise collocation and the other step by step procedures which do not make the derivation of one from the other possible. We describe these differences immediately below.

6.2 Characteristics and Assumptions

The characteristics of the methods derived in the foregoing are given in Table 6-1. Listed in the seven columns are: the symbolic forms of the mathematical models (F and G) representing the functional relationships between the observables (L and x), parameters (X) and signal components (S and S'); the four possible quantities to be weighted - observations (P), parameters (P_x), model G (P_m) and signal components ; the weighed quadratic form to be minimized; and the indication whether the parameters (X) are taken to vary with time. Given in the rows heading each group is a series of "labels", consisting of a combination of the terms-adjustment, filtering, smoothing and prediction. First we discuss the characteristics and then in Section 6.3 the labelling.

TABLE 6-1. Characteristics of Derived Methods

	Model	Weights on Observations	Weights on Parameters	Weights on Model (G)	Weights on Signals	Quadratic form	Time Varying Parameters
Standard Cases (Adjustment)							
Condition	$F(L) = 0$	P	--	--	--	residual V	--
Parametric	$L = F(X)$	P	--	--	--	"	--
Combined	$F(X, L) = 0$	P	--	--	--	"	--
Combined with weights	$F(X, L) = 0$	P	P_x	--	--	residuals V parameters X	--
Step by Step (Adjustment and Filtering)							
Sequential	$F_1(X, L_1) = 0$ $F_2(X, L_2) = 0$	P_1, P_2	--	--	--	residuals V_1, V_2	--
Phase	"	"	--	--	--	"	--
Summation Normal Equations	"	"	--	--	--	"	--
Tienstra Phase	"	"	--	--	--	"	--
Step by Step (Adjustment, Filtering, Prediction of Parameters, Smoothing)							
Kalman Bayes	$F_1(X, L_1) = 0$ $F_2(X, L_2) = 0$ $G(X(t)) = 0$	P_1, P_2	P_x(option)	P_m	--	residuals V_1, V_2 parameters X model errors Y_m	yes
Collocation (Adjustment, Prediction, Smoothing, Filtering)							
original	$F(X, S, S', x) = 0$	with signal S'	P_x(option)	--	for S and S'	signal S, combined quantity $Z=S+n$(noise)	--
in adjustment terms	$F(X, S, S', L) = 0$	$P = \Sigma_L^{-1}$	P_x(option)	--	for S and S'	residual V signals S and S'	--
stepwise collocation	$F_1(X, S, S_1', x_1)=0$ $F_2(X, S, S_2', x_2)=0$	with signals S_1', S_2'	--	--	for S, S_1', S_2'	signal S, combined quantity $Z_1=S_1'+n_1$ $Z_2=S_2'+n_2$	--

Models

The methods contain different types of mathematical models as follows. The condition case contains only observables. The parametric method has observables expressed explicitly as a function of the parameters. In the combined case the observables and the parameters are implicitly related.

The sequential, phase, summation of normal equations, and Tienstra phase methods have two sets of models with common parameters and different observables. The only reason for this specific arrangement of parameters and observables was to be consistent with the Kalman filtering formulation.

The Kalman and Bayes filtering formulation has two groups of models: one implicit form (combined case) in the first step; and one implicit form along with a second model giving the time variation in the parameters in the second step.

The original formulation of collocation has an implicit model relating the parameters X, two signal components S and S', and the observations (x). The noise n is combined with the signal S'. Expressions are given for X and S but not for S' and n. The formulation of collocation in adjustment terms also uses the implicit type of model. All three quantities S, S' and residuals V (-n) are treated explicitly and expressions given for their computation along with X. Stepwise collocation employs an implicit model partitioned into two parts according to the observations x_1 and x_2; the parameters X and signal S to be computed remain common to the two steps. Note the signal at the observation points (S'_1, S'_2) are partitioned along with the noise (n_1, n_2).

Both developments of collocation contained herein utilized a minus identity matrix in the second design matrix, namely equations 5-14 and 5-68, e.g.

$$B^* = [0 \,\vdots\, -I \,\vdots\, -I] \ ,$$

where the null, and two minus identity matrices correspond respectively to $\hat{\bar{S}}$, $\hat{\bar{S}}'$ and $\hat{V}$. As well, one could have non-identity matrices such as in

$$B^* = [0 \,\vdots\, B_S \,\vdots\, B] \ .$$

The latter would yield more general expressions with the same analogy holding that exists between the combined and parametric cases. Kouba [1975] has used collocation, without interpolation, in this more general form; Schwartz [1975] has done so as well but without the parameters X.

We note that by including a null matrix in the first position of the hypermatrix B* we simply build-in an interpolation process for the signal S (at the computation points). These signal components do not enter into the solutions of $\hat{X}$, $\hat{\bar{S}}'$, and $\hat{V}$. $\hat{\bar{S}}$ is computed via its covariance with S' (See equations 5-83 and 5-84), namely

$$\hat{\bar{S}} = -\Sigma_{SS'}(\Sigma_{S'}+\Sigma_L)^{-1}(A\hat{X}+W) \ ,$$

$$\hat{\bar{S}} = -C_{SX}(C + D)^{-1}(A\hat{X}+W) \ .$$

There is no theoretical limit on the number of signal quantities $\hat{\bar{S}}$ that can be computed. This is why collocation has become known in some scientific circles as "the method that can compute an infinite number of parameters". We, of course should not confuse these "interpolated parameters" $\hat{\bar{S}}$ with the parameters $\hat{X}$.

One other aspect of collocation worth mentioning is the possibility of processing errorless data, that is discrete data procured

from some computational procedure as contrasted to data from a measurement procedure. Errorless data will have no noise component but will have a signal component due to its discrete distribution. This topic is sometimes known as least squares approximation [Vaníček and Wells 1972]. This latter usage of collocation is probably the rationale in the original derivation of collocation for not explicitly involving the noise but combining it with the signal S'. Even if we include the noise n and signal S' as separate-explicit quantities, as in the alternative derivation (Section 5.5), we still can use the equations for processing errorless data by simply deleting all quantities and expressions dealing with the observed quantities.

Weights on Observations

The weight matrix of the observations plays a role in all methods. In the standard cases the weight matrix (P) is usually defined as the inverse of the variance-covariance matrix, Σ_L^{-1} ($\sigma_o^2 = 1$). In most applications Σ_L has a diagonal form but may be on occasions a banded or fully populated matrix. In the first two groups of step by step methods (sequential through Bayes), the observations in different stages are assumed to be statistically independent. In the collocation equations we find that Σ_L (D in collocation notation) is combined with the variance-covariance matrix of the signal S' (denoted $\Sigma_{S'}$ or C). Because by nature, signal quantities are statistically dependent, C is a fully populated matrix, as is C plus D and its inverse. Thus, we witness in collocation that a fully populated matrix of order equal to the number of observations needs to be inverted. Even in stepwise collocation where the observed data is partitioned into groups, the

signal quantities in respective groups are statistically dependent and the correlation is indeed taken into account by the stepwise collocation equations.

Weights on Parameters

The weighting of parameters was included in the combined case and carried through our derivation of the Kalman and collocation equations for two reasons. One was to involve these weights explicitly so that one can see where in the expressions they enter vis-a-vis the other weights such as on the signal, observations, and model. Secondly, in the Kalman filtering equations a priori information is needed to start-off the filter process. This information can be from two sources: from a solution of F_1 or from some independent solution not related to the present process, or both. Including weights on the parameters in collocation makes the expressions more general. If no information on the parameters X is available the weight matrix is simply a null matrix.

Weights on the Model

The weighting of a model arises only in Kalman and Bayes filtering. It is assumed that the model (G) describing the time variation in the parameters is not exact. These model errors are assumed to have expectation zero (zero mean) and second moments given by a variance-covariance matrix containing no correlation between consecutive steps. One should note that model errors are assumed to be

present only in the secondary model (G) and not in the primary model (F).

Weights on the Signal

The weighting of the signal is intrinsic to collocation. One can loosely interpret the signal as a model error by imagining that we are unable to select the appropriate parameters X and design matrix A to describe the signal component of the observable. We then say that this is a shortcoming in the primary model. But as it turns out, in collocation it is hypothesized that the signal is better described by a variance-covariance matrix rather than by some analytical parameterization. Thus we see weighting the signal in collocation as being something quite different from say the problem or parameterization in models for the purpose of eliminating systematic effects in observables. Collocation is then a powerful tool in processing observed data containing a known signal component of zero mean and a certain variance-covariance matrix. Note the signal component occurs in the primary model (F) contrasted with the Kalman filtering model error which occurs in the secondary model (G).

At this point in our analysis, we may introduce the argument that Kalman filtering is identical to collocation because all one needs to do to show their equivalence is to substitute the time variation in the parameters described in the secondary model (G), along with the weights on the model (P_m), into the primary model F, thereby yielding one model F as in collocation where the parameters X and weight matrix P_m are interpreted as the signal S' and weight matrix for the signal.

By doing this one "looses" the parameters X. Also how does one introduce the signal S? Thus it is difficult to see how collocation is equivalent to Kalman filtering. We return to this discussion in another context in Section 6.5.

Quadratic Forms

Next we discuss the quadratic form of each method. This is the most fundamental quantity for it is through its minimization that one obtains least squares estimates for the parameters and other quantities. We have seen in the foregoing developments that our quadratic forms consist of two parts, a vector and a weight matrix.

In the quadratic forms for the standard cases of adjustment the vector is defined in terms of residual vector $\hat{V}$(corrections to the observations) and weight matrix P of the measurements, that is

$$\hat{V}^T P \hat{V} = \text{minimum} . \qquad (6\text{-}1)$$

The exception is the case where the parameters are weighted thus giving rise to a second quadratic form where the vector is defined in terms of the corrections to the parameters and weight matrix corresponding to the a priori information on the parameters, that is

$$\hat{V}^T P \hat{V} + \hat{X}^T P_X \hat{X} = \text{minimum} . \qquad (6\text{-}2)$$

In the step by step procedures of sequential, phase, summation of normal, and Tienstra phase, two quadratic forms are minimized. Both are defined in terms of a vector of residuals and weight matrix for the measurements. The fundamental assumption is that there is no

correlation between the measurements of two consecutive steps, thus allowing the quadratic form to be broken in two, that is

$$\hat{V}_1^T P_1 \hat{V}_1 + \hat{V}_2^T P_2 \hat{V}_2 = \text{minimum} . \qquad (6\text{-}3)$$

Thus we see that the standard cases of adjustments do not differ fundamentally from these step by step procedures. All are mathematically equivalent and rigorous, the only difference is in computational efficiency (see Section 6.4).

The step-by-step procedures of Kalman and Bayes filtering have the following four quadratic forms:

$$\hat{V}_1^T P_1 \hat{V}_1 + \hat{V}_2^T P_2 \hat{V}_2 + \hat{X}^T P_x \hat{X} + \hat{Y}_m^T P_m \hat{Y}_m = \text{minimum}, \qquad (6\text{-}4)$$

where the first three pertain to the measurement corrections and parameters (an option), while the fourth is defined in terms of the vector of errors in the model describing the time variation in the parameters and the corresponding model error weight matrix. It is assumed that the model errors are statistically independent between consecutive stages. Thus we witness that these two step by step procedures are fundamentally different and more general than the other methods mentioned immediately above.

The quadratic form for the original collocation is fundamentally different from all of the above. Disregarding weights on the parameters, the quadratic form is

$$\hat{V}^T P \hat{V} = \text{minimum} \qquad (6\text{-}5)$$

$$\left[\hat{S}^T \,\vdots\, \hat{Z}^T\right] \begin{bmatrix} C_{ss} & C_{sx} \\ C_{xs} & C_{xx} \end{bmatrix}^{-1} \begin{bmatrix} \hat{S} \\ \cdots \\ \hat{Z} \end{bmatrix} = \text{minimum} ,$$

where $\hat{S}$ is the vector of signal components to be estimated at the computation points and $\hat{Z}$ are "centred observations" defined by (equation 5-4)

$$\hat{Z} = x - A\hat{X} , \qquad (6\text{-}6)$$

or the combined quantity

$$\hat{Z} = \hat{S}' + \hat{n} . \qquad (6\text{-}6a)$$

The variance-covariance matrices C_{ss} is for the signal, C_{xx} for the measurements (containing two components S' and n), and C_{sx} and C_{xs} are covariances between the signal S and measurements. Since the signal and measurement errors (noise) are assumed to be statistically independent, these covariances originate purely from a signal source. For collocation to make sense, these signal covariances are always present. Thus we see the sharp difference between collocation and the other methods, for in the step by step procedures above, the observations (and model errors) were assumed to be statistically independent between stages. It is because of the presence of this covariance, that the stepwise collocation equations could not be deduced from the standard cases or the Kalman filtering expressions, thus another reason for saying that collocation is a more general method. On the other hand, the Kalman filter equations cannot be deduced from the stepwise collocation expressions which do not have, in their present form, the time variation aspect in the parameters.

With regards to the signal quantity S, we witness that even though it enters into the quadratic form, it does not enter into the constraint function as seen below (equation 5-13):

$$A\hat{X} + B\hat{V} + W = 0 \tag{6-7}$$

$$A\hat{X} + [0 \quad -I]\begin{bmatrix} -\hat{S} \\ -- \\ -\hat{Z} \end{bmatrix} - x = 0 \tag{6-8}$$

$$A\hat{X} + \hat{Z} = x \ , \tag{6-9}$$

which is the original collocation model. This fact is also true for the alternative derivation of collocation. On the other hand, however, the signal S is "involved" in the constraint function and its determination is made via the signal variance-covariances entered into the computation as already mentioned earlier.

The quadratic form used in the alternative derivation (equation 5-72) is (after deleting weights on parameters)

$$[\hat{S}^T \mid \hat{S}'^T] \begin{bmatrix} \Sigma_S & \Sigma_{SS'} \\ \Sigma_{S'S} & \Sigma_{S'} \end{bmatrix}^{-1} \begin{bmatrix} \hat{S} \\ -- \\ \hat{S}' \end{bmatrix} + \hat{V}^T P \hat{V} = \text{minimum}, \tag{6-10}$$

where we see that the signal quantities $\hat{S}$, $\hat{S}'$, and corrections to the observations $\hat{V}$ enter as separate quantities. Even though the above quadratic form looks somewhat different from that of the original collocation (equation 6-5), the same set of equations are obtained for the parameters $\hat{X}$ and signal $\hat{S}$.

To demonstrate that collocation is indeed different from ordinary least square methods, we compare the respective equations giving the solution for the parameters. For ordinary least squares we chose the equation of the parametric case (B = -I) with weighted parameters (equation 2-74)

$$\hat{X} = -[A^T PA + P_x]^{-1} A^T PW \qquad (6\text{-}11)$$

and that of collocation (equation 5-23)

$$\hat{X} = -[A^T (C+D)^{-1} A + P_x]^{-1} A^T (C+D)^{-1} W. \qquad (6\text{-}12)$$

We chose the cases with weights on the parameters so one can see where all the various weights enter into the solutions. First we see that the weight matrix P_x on the parameters enters identically. This is not surprising since the collocation equations were deduced from the standard equations. Secondly, we see that the weights on the observations (P) enter alone in ordinary least squares, while the variance-covariance of the purely measurement errors (D) enters together with the variance-covariance of the signal (C) in collocation. This latter fact is what gives rise to the two different solutions for the parameters (and the other quantities as well). For example Rapp [1973] and Reigber and Ilk [1975] have used both ordinary least squares and collocation methods in practise and have found significant differences in the respective solutions.

A point worth reiterating vis-a-vis the comparison of original collocation, the alternative derivation of collocation, and ordinary least squares is that even though we were able to formulate collocation as a straightforward problem in adjustments, thereby explicitly involving the corrections to the observations in the quadratic form and constraint function and even solving for them, we still deduced the same set of equations for the parameters $\hat{X}$ and $\hat{S}$; but this equivalent set of collocation equations is different from the corresponding expression for the parameters X in ordinary least squares as demonstrated immediately above.

6.3 Labelling of Methods

We are now in a position to label the above methods by terms such as "adjustment", "filtering", "smoothing", and "prediction" often used in the literature. First we define them and then apply them. We deal with two situations (Figure 6-2) - the parameters not varying with time ($X \neq X(t)$) and in the case they do ($X = X(t)$).

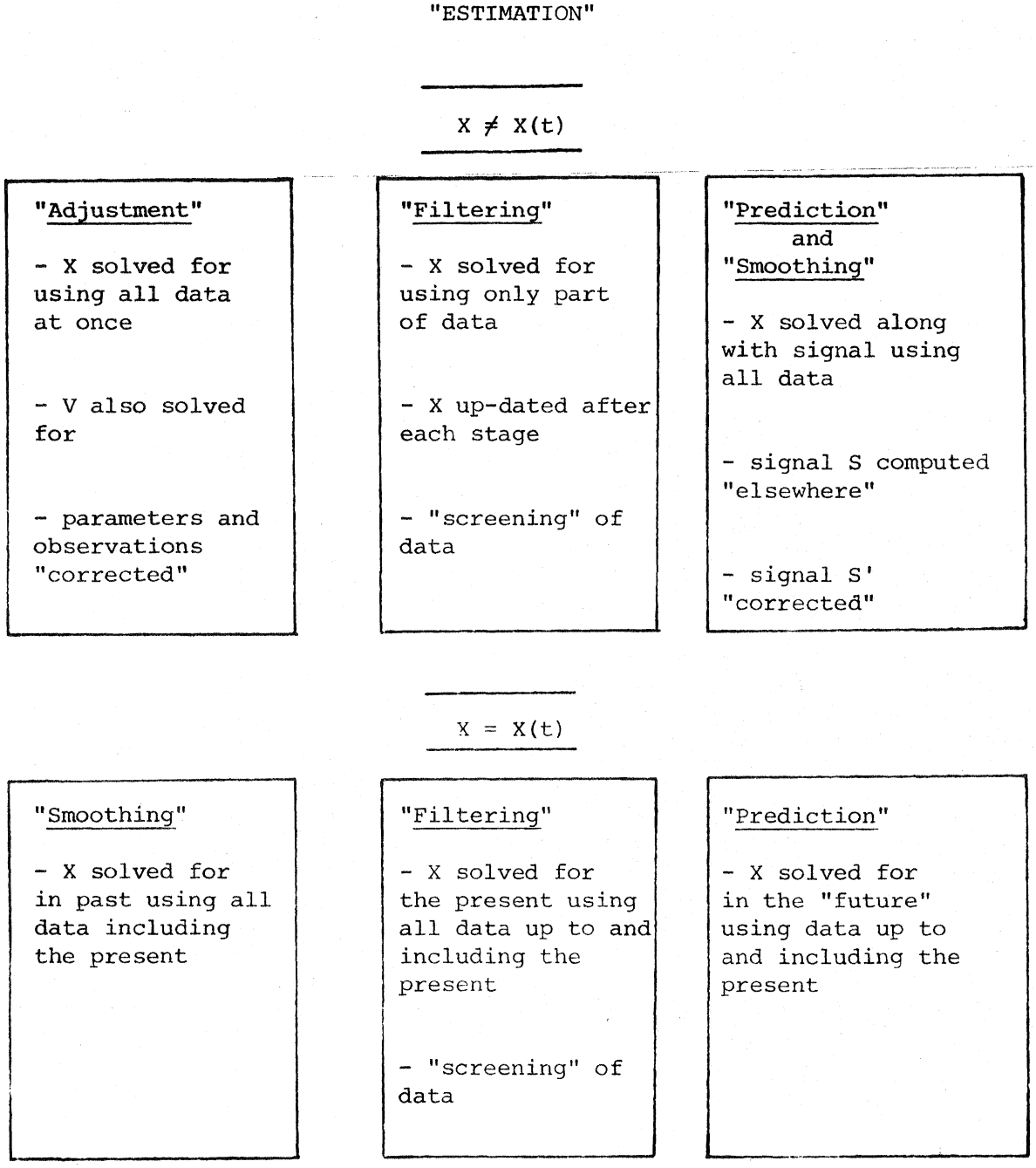
An adjustment is said to take place when an estimate of the parameters and the observational residuals is made from a least squares solution using all the observations at once. Then the approximate values (or quasi-observed values) of the parameters and the observed value of the observables are corrected (adjusted).

Filtering is said to take place when an estimate of the parameters ($X \neq X(t)$) is made using only part of all the available data in any given stage of the estimation, and then this estimate is up-dated as additional data is added in the subsequent steps of the estimation process. By estimation in steps, it is possible to screen (filter) observed data as it is added.

Prediction takes place in collocation ($X \neq X(t)$) when an estimate of the signal at computation points is made using all or part of the data. The signal S' at the observation points is said to be smoothed after they have been corrected.

Smoothing is said to take place in Kalman filtering when an estimate for the parameters ($X = X(t)$) is made for sometime in the past using all observed data including that for the present; that is work backwards and up date X_1 due to F_2 because X_1 is a result of F_1 only.

Figure 6-2. Terms: Adjustment, Filtering, Prediction, and Smoothing



Filtering is said to take place when an estimate for the parameters ($X = X(t)$) is made for the present using all observed data in the past, up to, and including the present. The data of the "present" is allowed to affect the solution of the "past", thus allowing screening (filtering) as for the case ($X \neq X(t)$).

Prediction is said to take place when an estimate for the parameters ($X = X(t)$), is made for the "future" using all data up to and including the present. This is possible because of the existence of the secondary model which describes the variation of the parameters with time.

We have just witnessed the striking difference in the meanings of the terms prediction and smoothing in Kalman filtering and collocation, showing yet another difference between these two methods. Further to this discussion we understand that in Kalman filtering the prediction essentially takes place at the observation points where the future observations will be made, while in collocation, prediction is made at computation points where the signal is sought. In Kalman filtering we could, however, predict the parameters X at points where there will be no observations and thus have no benefit from the new observations.

These interpolated values of the parameters begin to look like the signal quantities S, while the parameters that coincide with the observation points seem to look like the signal S'. In attempting to strengthen this argument we could point out that the variance-covariance matrix for the model errors Σ_m corresponds to the variance-covariance matrix for the signal $\Sigma_{S'}$ ($C_{S'S}$ or C in collocation notation). If the parameters X in Kalman filtering would take on this interpretation then where are the parameters X themselves that would correspond to the parameters X in collocation? We must not forget in one case the parameters X of the mathematical model are being predicted and in the other it is the signal S, a quantity related to the observables, that is being predicted. Also, one process is static - the other is dynamic. One can go on endlessly predicting signals in collocation. In Kalman filtering, prediction of the parameters only has sense if new observed data is forthcoming, for the lack of data stops the filter. One sees some loose analogies between Kalman filtering and collocation but it is difficult to see they are equivalent.

From the foregoing developments and discussions the following labelling is possible (see Table 6-1):

(i) the standard cases are adjustments (of observations and parameters);

(ii) the step by step procedures of sequential, phase, summation of normal equations, and Tienstra phase are adjustments (of observations and parameters) and filtering (of observations);

(iii) the step by step procedures of Kalman and Bayes are adjustments (of parameters), prediction and smoothing (of parameters), and filtering parameters (due to G) and observations;

(iv) collocation is an adjustment (of observations and parameters) and a smoothing and prediction (of signal);

(v) stepwise collocation, in addition to the characteristics of collocation, is filtering of observables (with two random parts, the signal and the noise).

6.4 Comparison from the Computational Point of View

We choose to compare the methods from the computational point of view by examining the number, form and size of matrices to be inverted in the respective methods. It is recognized that this is not an entirely complete analysis as one should also consider the number of multiplications, subtractions, and additions. Nevertheless, this will give us some indication of the computational efficiency of each approach. We list below the expressions giving the solution for the parameters for the various methods for easy reference.

Condition case:

The correlates are computed from equation 2-130, namely

$$\hat{K} = \underbrace{(B \quad P^{-1} \quad B^T)^{-1}}_{rxr} W. \qquad (6\text{-}13)$$
$$\quad rxn \quad nxn \quad nxr \quad rxl$$

Parametric case (equation 2-115):

$$\hat{X} = -\underbrace{[A^T \quad P \quad A]^{-1}}_{uxu} A^T \quad P \quad W \quad . \qquad (6\text{-}14)$$
$$\quad uxn \quad nxn \quad nxu \quad uxn \quad nxn \quad nxl$$

Combined case with weighted parameters (equation 2-74):

$$\hat{X} = -\underbrace{[A^T \underbrace{(B \quad P^{-1} \quad B^T)^{-1}}_{rxr} A + P_x]^{-1}}_{uxu} A^T \quad (B \quad P^{-1} \quad B^T)^{-1} W. \qquad (6\text{-}15)$$
$$\quad uxr \quad rxn \quad nxn \quad nxr \quad rxu \quad uxu \quad uxr \quad rxn \quad nxn \quad nxr \quad rxl$$

Sequential (equation 4-22) and Tienstra Phase (equation 4-74):

$$\hat{X} = \hat{X}' - N_1^{-1} A_2^T [M_2 + A_2 N_1^{-1} A_2^T]^{-1} (W_2 + A_2 \hat{X}')$$

$$= \hat{X}' - N_1^{-1} A_2^T [B_2 \quad P_2^{-1} \quad B_2^T + A_2 \quad N_1^{-1} \quad A_2^T]^{-1} (W_2 + A_2 \quad \hat{X}') ,$$
$$uxl \quad uxu \quad uxr_2 \quad r_2xn \quad nxn \quad nxr_2 \quad r_2xu \quad uxu \quad uxr_2 \quad r_2xl \quad r_2xu \quad uxl$$

(6-16)

where the initial solution (equation 4-20) is

$$\hat{X}' = -N_1^{-1} U_1 = -\underbrace{[A_1^T \underbrace{(B_1 P_1^{-1} B_1^T)^{-1}}_{r_1 x r_1} A_1]^{-1}}_{uxu} U_1 . \qquad (6\text{-}16a)$$

Phase (equation 4-36):

$$\hat{X} = \hat{X}' - N^{-1} A_2^T M_2^{-1} (W_2 + A_2 \hat{X}') ,$$

$$= \hat{X}' - \underbrace{[N_1 + A_2^T \underbrace{(B_2 P_2^{-1} B_2^T)^{-1}}_{r_2 x r_2} A_2]^{-1}}_{uxu} A_2^T \underbrace{(B_2 P_2^{-1} B_2^T)^{-1}}_{r_2 x r_2} (W_2 + A_2 \hat{X}') , \qquad (6\text{-}17)$$

where

$$\hat{X}' = -N_1^{-1} U_1 \tag{6-17a}$$

as before.

Summation of normal equations (equation 4-39)

$$\hat{X} = -(N_1 + N_2)^{-1} (U_1 + U_2)$$

$$= -[A_1^T (B_1 \underbrace{P_1^{-1}}_{r_1 x r_1} B_1^T)^{-1} A_1 + A_2^T (B_2 \underbrace{P_2^{-1}}_{r_2 x r_2} B_2^T)^{-1} A_2]^{-1} (U_1 + U_2). \tag{6-18}$$

Kalman filtering (equations 3-60 and 3-61):

$$\hat{X}_2 = \hat{X}_2' - (N_2')^{-1} A_2^T [M_2 + A_2 (N_2')^{-1} A_2^T]^{-1} (W_2 + A_2 \hat{X}_2') , \tag{6-19}$$

where

$$(N_2')^{-1} = [\Phi (N_1 + P_x)^{-1} \Phi^T + P_m^{-1}] , \tag{6-20}$$

$$[M_2 + A_2 (N_2')^{-1} A_2^T]^{-1} = [B_2 \underbrace{P_2^{-1}}_{r_2 x r_2} B_2^T + A_2 (N_2')^{-1} A_2^T]^{-1} , \tag{6-21}$$

and where from equations 3-58 and 3-23

$$\hat{X}_2' = \Phi \hat{X}_1' , \tag{6-21a}$$

$$\hat{X}_1' = -(N_1 + P_x)^{-1} U_1 = -[A_1^T \underbrace{(B_1 P_1^{-1} B_1^T)}_{r_1 x r_1} A_1 + P_x]^{-1} U_1 . \tag{6-21b}$$

Bayes filtering (equations 4-4 and 4-5):

$$\hat{X}_2 = \hat{X}_2' - N_2^{-1} A_2^T M_2^{-1} (W_2 + A_2 \hat{X}_2') , \tag{6-22}$$

$$\underset{uxu}{N_2^{-1}} = \underbrace{[(\Phi(N_1 + P_x)^{-1} \Phi^T + P_m^{-1})^{-1}}_{uxu} + \underbrace{A_2^T (B_2 P_2^{-1} B_2^T)^{-1} A_2}_{r_2 x r_2}]^{-1} , \tag{6-23}$$

$$M_2^{-1} = \underbrace{(B_2 \quad P_2^{-1} \quad B_2^T)^{-1}}_{r_2 x r_2} , \tag{6-24}$$
$$r_2 x n_2 \quad n_2 x n_2 \quad n_2 x r_2$$

and where $\hat{X}_2'$ is obtained as previously (Kalman filtering).

Collocation (equation 5-23):

$$\underset{uxl}{\hat{X}} = - \underbrace{[\underbrace{A^T (C + D)^{-1} A}_{nxn} + P_x]^{-1}}_{uxu} \underbrace{A^T (C+D)^{-1}}_{uxn} W. \tag{6-25}$$
$$nxu \quad uxu$$

Stepwise collocation (equation 5-54):

$$\hat{X} = \hat{X}_1 + P_1^{-1} \bar{A}_2^T \bar{C}_{22}^{-1} (x_2 - C_{21} C_{11}^{-1} x_1 - \bar{A}_2 \hat{X}_1) , \tag{6-26}$$

where

$$\underset{uxu}{P_1^{-1}} = (\underset{uxn_1}{A_1^T} \quad \underset{n_1 x n_1}{C_{11}^{-1}} \quad \underset{n_1 xu}{A_1})^{-1} , \tag{6-27}$$

$$\underset{n_2 x n_2}{\bar{C}_{22}^{-1}} = [\underset{n_2 x n_2}{C_{22}} - \underset{n_2 x n_1}{C_{21}} \underset{n_1 x n_1}{C_{11}^{-1}} \underset{n_1 x n_2}{C_{12}} + \underset{n_2 xu}{\bar{A}_2} \underset{uxu}{P_1^{-1}} \underset{uxn_2}{\bar{A}_2^T}]^{-1} , \tag{6-28}$$

and where from equations 5-57 and 5-54

$$\hat{X}_1 = P_1^{-1} A_1^T C_{11}^{-1} x_1 \quad . \tag{6-28a}$$

The number and size of the matrices to be inverted in the various methods indicated in their respective solution vectors are given in Table 6-2. The choice between sizes r or n follows from whether one has a combined (implicit) or parametric (explicit) type of mathematical model. The asterisk (*) on the n_1 or n_2 indicates that an inversion is not necessary if the observations are uncorrelated within the stage. We of course have the assumption that the observations are uncorrelated between stages; the only exception to this is in collocation where the signal is correlated both between different stages and within a stage. To assist in our analysis we articulate the efficiencies of some methods for specific situations in Table 6-3.

Tables 6-2 and 6-3 speak for themselves; we only emphasize some of the important comparisons. For the first situation in Table 6-3 (I - combined case) we see that the sequential or Tienstra phase techniques require one less inversion than the phase or summation techniques. The Kalman filter is identical to the sequential technique. The Bayes technique requires two more inversions than the Kalman technique.

The second situation (II - parametric case with correlated observations) makes no difference in the comparison of the methods - the number of observations (n_1 and n_2) simply replace the number of equations.

The third situation (III - parametric case uncorrelated observations) makes a difference in the comparison as several inversions of size n_1 and n_2 are not required. All techniques require one inversion in the initial solution. In the corrective term, the sequential or

Table 6-2. Number and Order of Matrices to be Inverted

	Initial Solution			Corrective Term		
	1st	2nd	3rd	1st	2nd	3rd
Condition	r					
Parametric	n*	u				
Combined	r	u				
Sequential Tienstra Phase	r_1 or n_1^*	u		r_2 or n_2		
Phase	r_1 or n_1^*	u		r_2 or n_2^*	u	
Summation	r_1 or n_1^*	r_2 or n_2^*	u			
Kalman	r_1 or n_1^*	u		r_2 or n_2		
Bayes	r_1 or n_1^*	u		r_2 or n_2^*	u	u
Collocation	n	u				
Stepwise Collocation	n_1	u		n_2		

r - number of equations; n - number of observations;
u - number of parameters;
* this inversion not needed if observations are uncorrelated.

Table 6-3. Efficiencies of Methods for Specific Situations

	Initial Solution			Corrective Term		
	1	2	3	1	2	3
I. Combined Case - correlated or uncorrelated observations within stage						
Sequential and Tienstra Phase	r_1	u		r_2		
Phase	r_1	u		r_2	u	
Summation	r_1	u	r_2 ¦ u			
Kalman	r_1	u		r_2		
Bayes	r_1	u		r_2	u	u
II. Parametric Case - correlated observations within stage						
Sequential and Tienstra Phase	n_1	u		n_2		
Phase	n_1	u		n_2	u	
Summation	n_1	u	n_2 ¦ u			
Kalman	n_1	u		n_2		
Bayes	n_1	u		n_2	u	u
III. Parametric Case - uncorrelated observations within stage						
Sequential and Tienstra Phase	--	u		n_2		
Phase	--	u		--	u	
Summation	--	u	-- ¦ u			
Kalman	--	u		n_2		
Bayes	--	u		--	u	u

Tienstra phase require only one inversion of size n_2 while in the phase and summation forms only one inversion of the size u is needed. Thus the most efficient method then depends upon the relative sizes of the u and n_2 matrices. In comparing the Kalman and Bayes techniques we see that our comparison hinges upon which is a greater effort - one inversion of size n_2 or two of size u each.

In the case that the number of new observations (n_2) or equations (r_2) is less than the number of parameters (u), it is only the sequential or Tienstra phase techniques that are more efficient than redoing a simultaneous solution over again the required number of times. In fact, the phase or summation approaches are identical to a repeated simultaneous solution from a computational point of view as the tables indicate the same number and size of matrices are to be inverted.

For collocation the problem of inverting a matrix of order equal to the number of observables (signals) is present because of the correlated signal. We see that stepwise collocation is formulated in an analogous way to the sequential and Tienstra phase adjustments as the corrective term involves the inversion of a matrix of order equal to the number of new observations added. This may seem curious as the two cases are significantly different from the point of view of the correlation in the observables (signal components) from one step to the next. However, one can be assured that the full variance-covariance matrix of the signal is being implicitly inverted in the case of collocation, and thereby increasing the number of multiplications, additions and subtraction operations over that

present in the Kalman (sequential) expressions. The advantage* of stepwise collocation over ordinary collocation may not be as great as that for sequential or Tienstra phase methods over the non step by step methods because of the correlation in the signal. Stepwise collocation is very useful when one cannot invert the variance-covariance matrix of the signal all at once because of limited computer storage. Also it is useful for screening of data.

6.5 The Derived Methods and Related Topics

In this section we discuss miscellaneous topics which are marginally related to the methods developed herein, yet are close enough to be of importance. In Figure 6-3, we illustrate some of these related topics along with the methods already developed. Variations to the least squares method are grouped according to: cases; weighting; combinations; partitioning; time and model errors; and collocation. We have already discussed the cases of adjustments in sufficient detail. As far as variations arising from weighting is concerned, there are two interesting aspects, one dealing with change in weights on the observations and the other concerning weighting of a subset of parameters.

In Feddeev and Feddeeva [1963] one can find expressions which give the correction to the parameters due to making a change in the weights on the parameters without recomputing the entire problem over again. One can be assured that changing the scale of the weight matrices (P and P_x) will not change the solution, but changing the individual variances will have an effect. The neglection of covariance in the observables has been studied by, for example,

* from the computational point of view

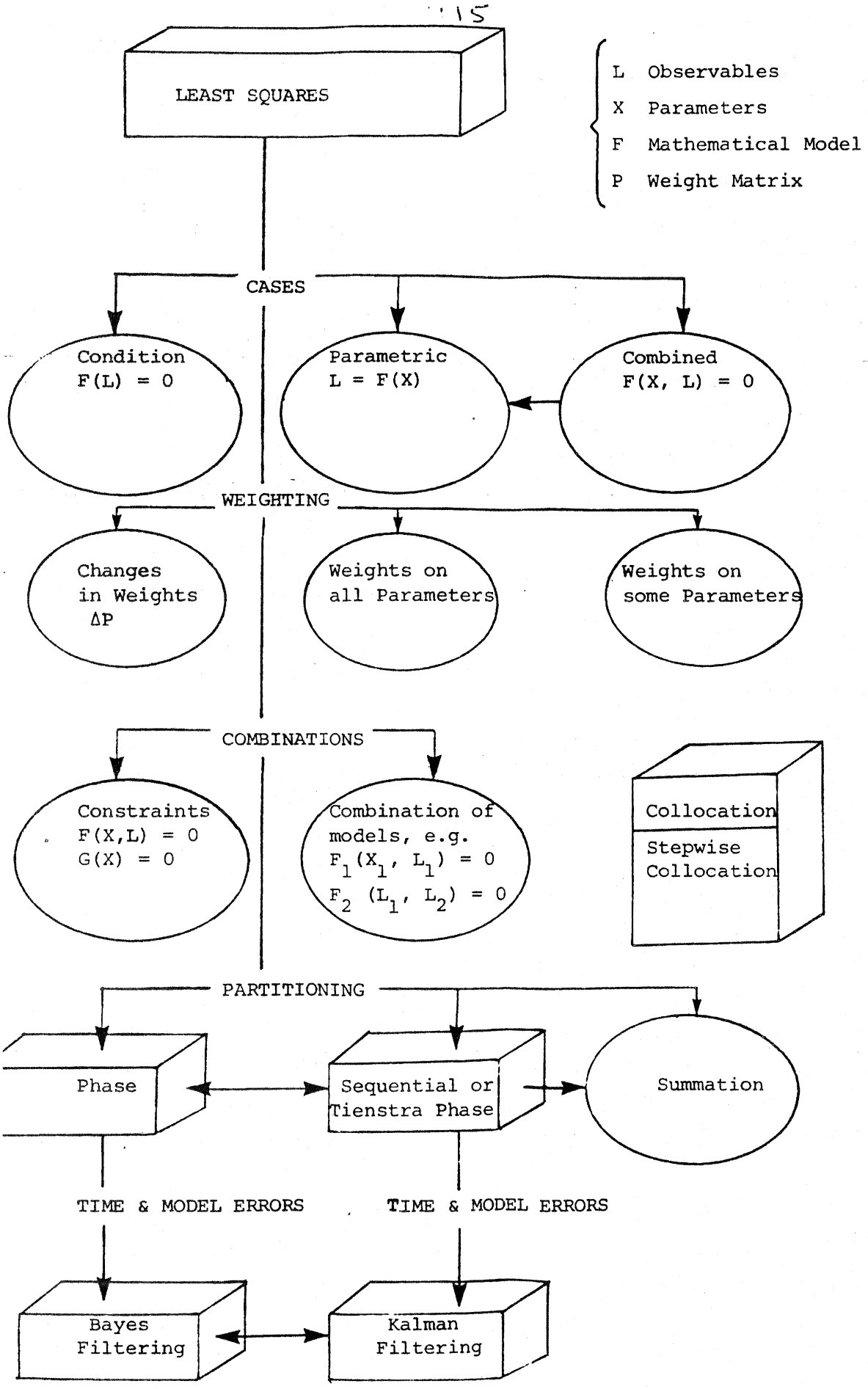


Figure 6-3 : The Derived Methods and Related Topics

Magnuss and Meguire [1962] in general terms, and by Schwarz [1969] for a specific problem in satellite geodesy.

Blaha [1974] has studied the problem of applying weights to none, all, or a subset of the parameters X; further, he considered another set of parameters $\tilde{X}$ with an associated variance-covariance matrix $\Sigma_{\tilde{x}} \cdot \Sigma_{\tilde{x}}$ was used exclusively for the purpose of error propagation into other parameters $\bar{X}$ without changing the values of $\tilde{X}$. This study differs fundamentally from our developments herein as each time the parameters had a variance-covariance matrix, they were included in the quadratic form and corrections made to them. Blaha's work has application in the densification of geodetic networks where the coordinates of super-control points are held fixed and their variances and covariances are propagated into the densified points. In our scheme, one can imagine this being achieved by the following combination of mathematical model technique:

$$\text{(adjustment model) } F\ (X,\ L) = 0,\ P,\ P_x \begin{cases} [0] \\ \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_x^{-1} \end{bmatrix} \\ \Sigma_x^{-1} \end{cases}$$

(error propagation model) $F\ (X,\ L,\ \tilde{X}) = 0$.

Reigber and Ilk [1975] have found that the collocation equations can be derived from the combination of mathematical models:

$$F_1(X,\ L) = 0,\ P$$

$$F_2(X) = 0,\ P_x \ .$$

They show that the equation for the solution vector X using the weighted constraint approach is identical to the equation for the signal in collocation, when in collocation the parameters (X) are not present.

The combination of mathematical model technique has yet another interesting application, this time in terrestrial geodesy. Thomson [1970] has showed that a geodetic network adjusted by the condition method can be rigorously combined with a second network adjusted by the parametric method by use of the combined case model of adjustments. This means that one does not have to redo the adjustment of the two networks from the beginning, but can simply take the design matrices and misclosure vectors from the two respective adjustments and affectuate a rigorous-combined adjustment. In symbolic form the above is represented by the following two mathematical models:

$$F_1 (X, L_1) = 0, P_1 ,$$

$$F_2 (L_1, L_2) = 0, P_1, P_2.$$

A basic assumption in collocation was that the expected value of the signal be zero. In some applications this assumption is easily met while in other applications it is not. For this reason we may find it useful to employ least squares approximation [e.g. Vaníček and Wells 1972] to overcome this rather limiting prerequisite on the signal component.

When considering the task of introducing a priori knowledge on the parameters X, one enters into whole new area called Bayesian

estimation. We have only touched upon the subject herein from the least squares adjustment point of view. The reader may wish to refer to a more exhaustive treatment of this subject [e.g. Bossler 1972].

7. SUMMARY

We group our findings under three categories: derivations; equivalences; and usage.

Derivations

1. The standard cases of adjustment serve as a basis from which one can derive directly the Tienstra phase equations (using condition case equations - Section 4.5) and the collocation equations (using the combined case - Section 5);
2. The Kalman filter equations can be derived from basic least squares using the conventional methodology used in deriving the standard adjustment cases (Section 3);
3. The sequential equations follow directly from the Kalman filtering equation (Section 4.2);
4. The phase equations follow directly from the Bayes filter equations (Section 4.3);
5. The stepwise collocation equations could not be deduced from the equations of the standard methods nor from the Kalman filter equations because of the correlation in the signal (Section 5.4);

Equivalences

6. The Bayes and Kalman filter equations are shown to be mathematically equivalent but different from the computational point of view (Sections 4.1 and 6.4);
7. The Tienstra phase equations are mathematically equivalent to, and identical with, the sequential equations (Section 4.5);
8. The Tienstra phase, sequential, phase and summation form of normal equations are all mathematically equivalent (Section 4) but in certain cases are different from the computational point of view (Section 6.4);
9. The Kalman (sequential, Tienstra phase) equations are better suited to the situation when a fewer number of observations, relative to the number of parameters, are added in the new stage, while the Bayes (phase) equations are better suited to the situation when more observations, relative to the original number of parameters, are added per new stage(for parametric case, uncorrelated observations) (Section 6.4).

Usage

10. When observations have only measurement errors and only one solution is needed, then solution by the standard cases of adjustments is adequate (Section 2-4);
11. When observables have both signal (correlated) and measurement error components, then a solution by the collocation method is advisable (Section 6.2).

12. When observables have only a measurement error component which is uncorrelated between stages, and several solutions are required, then the sequential, phase methods, and summation equations are most suitable (Sections 4.2, 4.3, 4.4);
13. When we have (12) immediately above and the parameters vary with time, then the Kalman and Bayes methods are most suitable (Sections 3 and 4.1);
14. When the observations contain signal and measurement errors and several solutions are required, then stepwise collocation may be used but due to the correlation in the signal, savings from the computation point of view are not that great because of the many operations that are necessary in carrying the signal correlation through to the final result (Section 6.4). Stepwise collocation is useful for screening of data and when a direct inversion of the variance-covariance matrix of the signal is not possible due to the lack of sufficient computer storage.

Closing Remarks

In this synthesis, a concerted effort was expended to discover the similarities and differences among the various methods by deriving them under the same cover using a common methodology. In those cases where equivalences were proved mathematically, the record is clear. However, for cases such as between Kalman filtering and stepwise collocation where no proof of equivalence was possible herein, reasons were given (Section 6.2) why the author only believes these methods are different. The author would appreciate learning from the reader of any proofs of equivalences not contained in this work. Furthermore, as the author did not concentrate on proving lack of equivalence, the knowledge of the existence of any such proofs would also be appreciated.

REFERENCES

Bjerhammer, A. (1973). General Model for Linear Filtering. A paper of the Royal Institute of Technology, Division of Geodesy, Stockholm.

Blaha, G. (1974). "Influence de l'incertitude des parametres Tenus Fixes dans uns Compensation sur la Propagation des Variances-Covariances, Bulletin Geodesique No. 113.

Bossler, John D. (1972). Bayesian Inference in Geodesy. Ph.D. Dissertation Dept. Geodetic Science, The Ohio State University, Columbus.

Tail, J.P. (1972). Application du Filtre de Kalman a la Navigation. Geophysical Prospecting, 20.

Fedeev and Feddeeva (1963). Computational Methods of Linear Algebra. W.H. Freeman and Company, San Francisco.

Gauss, K.F. (1963). Theory of Motion of the Heavenly Bodies. Dover, New York.

Grant, S. and E.J. Krakiwsky (1974). An Alternative Derivation of the Kalman Filter Equations. Unpublished Paper, Department of Surveying Engineering, The University of New Brunswick, Fredericton.

Hamilton, W.C. (1964). Statistics in Physical Science. Ronald, New York.

Isner, J. (1972). Determination of Surface Densities from a Combination of Gravimetry and Satellite Altimetry. Reports of the Department of Geodetic Science No. 186, The Ohio State University, Columbus.

Kaula, W.M. (1966). Theory of Satellite Geodesy. Blaisdell, Waltham.

Kalman, R.E. (1960). A New Approach to Linear Filtering and Prediction. Journal of Basic Engineering, ASME, 82D.

Kouba, J. (1970). Generalized Sequential Least Squares Expressions and Matlan Programming. M.Sc. Thesis, Dept. of Surveying Engineering, University of New Brunswick, Fredericton.

Kouba, J. (1972). Principle Property and Least Squares Adjustment. The Canadian Surveyor, 26, Ottawa.

Kouba, J. (1975). Program GEODOP. Geodetic Survey of Canada, Ottawa.

Knight, W. and P. Steeves (1974). Partial Solution of the Variance-Covariance Matrix for Geodetic Networks. The Canadian Surveyor, 28, No. 5.

Krakiwsky, E.J. and A.J. Pope (1967). Least Squares Adjustment of Satellite Observations for Simultaneous Directions and Ranges. Reports of the Department of Geodetic Science, No. 86, The Ohio State University, Columbus.

Krakiwsky, E.J. (1968). Sequential Least Squares Adjustment of Satellite Triangulation and Trilateration in Combination with Terrestrial Data. Reports of the Department of Geodetic Science, No. 114, The Ohio State University, Columbus,

Krarup, T. (1969). A Contribution to the Mathematical Foundation of Physical Geodesy. Publication No. 44 of the Danish Geodetic Institute, Copenhagen.

Krarup, T. (1970). The Method of Least Squares Collocation. Studia Geophysica, NO. 2, Roc. 14.

Krüger, L. (1905). Über die Ausgleichung von bedingten Beobachtungen in zwei Gruppen, Postdam.

Lambeck, (1963). The Adjustment of Triangulation with Particular Reference to Subdivision into Phases. B.Sc. Thesis, University of New South Wales, Sydney.

Magnuss, T.A. and T.B. McGuire (1962). Comparison of Least Squares and Minimum Variance Estimates of Regression Parameters. Annals of Mathematical Statistics 33, No. 2.

Moritz, H. (1972). Advanced Least Squares Methods. Reports of the Department of Geodetic Science, No. 75, The Ohio State University, Columbus.

Moritz, H. (1973). Determination of the Gravity Field by Collocation. Bolletino Di Geodesia E Science Affini, Anno XXXII.

Moritz, H. (1973a). Stepwise and Sequential Collocation. Reports of the Department of Geodetic Science, No. 203, The Ohio State University, Columbus.

Morrison, N. (1969). Introduction to Sequential Smoothing and Prediction. McGraw-Hill Book Company, New York.

Nassar, M. (1972). Least Squares Adjustment of Doppler Satellite Observations. Department of Surveying Engineering Technical Report No. 21, The University of New Brunswick, Fredericton.

Ott, N. and H.G. Meder (1972). The Kalman Filter as a Prediction Error Filter. Geophysical Prospecting 20.

Rapp, R. (1974). Numerical Results from the Combination of Gravimetric and Satellite Data Using the Principles of Least Squares Collocation. Reports of the Department of Geodetic Science, No. 200, The Ohio State University, Columbus.

Reigber, C. and K.H. Ilk (1975). Vergleich von Resonanzparameterbestimmungen mittels Ausgleichung und Kollokation. Mitteilung Nr. 119 des Instituts für Astronomische und Physikalische Geodäsie der Technischen, Universität, München.

Schwarz, C.R. (1967). Optimal Filter and Least Squares Estimation. U.S. Army Map Service Geodetic Memorandum, No. 1610, Washington D.C.

Schwarz, C.R. (1969). The Use of Short Arc Orbital Constraints in the Adjustment of Geodetic Satellite Data. Reports of the Department of Geodetic Science, No. 118, The Ohio State University, Columbus.

Schwarz, K.P. (1973). Even Zonal Harmonics from Satellite Observations by Collocation. Paper presented to the Second International Symposium, The Use of Artificial Satellites for Geodesy and Geodynamics, Athens.

Schwarz, K.P. (1974). Zonale Kugel Funktion Koeffizienten aus Satellitedaten durch Kollocation.

Schwarz, K.P. (1975). "Formulas and Fortran Programs to Determine Zonal Coefficients from Satellite Observations by Collocation". Reports of the Department of Geodetic Science No 222, The Ohio State University, Columbus.

Schmid, H.H. and E. Schmid (1965). A Generalized Least Squares Solution for a Hybrid Measuring System. The Canadian Surveyor, XIX, No. 1, Ottawa.

Tapley, B.D. (1972). Statistical Orbit Determination Theory. Proceedings of the Nato Advanced Study Institute in Dynamical Astronomy, held Cortina d'Ampezzo, Italy, August 9-21.

Tapley, B.D. and B.E. Schutz (1974). A Comparison of Orbit Determination Methods for Geodetic Satellites. A Report of the Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin.

Thompson, E.H. (1969). An Introduction to the Algebra of Matrices with Some Applications. The University of Toronto Press.

Thomson, D.B. (1971). Analysis of New Brunswick First Order Geodetic Networks. M.Sc. Thesis, Department of Surveying Engineering, The University of New Brunswick, Fredericton.

Tienstra, J.M. (1956). Theory of the Adjustment of Normally Distributed Observations, Argus, Amsterdam.

Tobey, W.M. (1930). The Differential Adjustment of Normally Distributed Observations. Geodetic Survey of Canada Pub. 27, Ottawa.

Tscherning, C.C. (1974). A Fortran IV Program for the Determination of the Anomalous Potential Using Stepwise Least Squares Collocation. "Reports of the Department of Geodetic Science No. 212, The Ohio State University, Columbus.

Tscherning, C.C. (1975). Determination of Datum Shift Parameters Using Least Squares Collocation. A Paper Presented at the Sixth Symposium on Mathematical Geodesy, Siena, Italy.

Uotila, U.A. (1967). Introduction to Adjustment Computations with Matrices. Lecture Notes, Department of Geodetic Science, The Ohio State University, Columbus.

Wells, D.E. and E.J. Krakiwsky (1971). The Method of Least Squares. Department of Surveying Engineering Lecture Notes No. 18, University of New Brunswick, Fredericton.

Vaníček, P. and D.E. Wells (1972). Least Squares Approximation and Related Topics. Department of Surveying Engineering Lecture Notes No. 22, The University of New Brunswick, Fredericton.

Wells, D.E. (1974). Doppler Satellite Control. Department of Surveying Engineering Technical Report No. 29, The University of New Brunswick, Fredericton.

Wiener, N. (1949). The Extrapolation, Interpolation, and Smoothing of Stationary Time Series, Wiley, New York.

Wolf, H. (1968). Ausgleichungsrechnung nach der Methode der Kleinsten Quadrate. Dümmler, Bonn.

Ying Chung-Chi (1972). On Phased Adjustment. Survey Review, 20.