

N 63 23805

578067  
P.12

*Technical Report No. 32-498*

*Theory of Orbit Determination—Part II  
Estimation Formulas*

*John D. Anderson*



JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA

October 1, 1963

Reproduced by  
**NATIONAL TECHNICAL  
INFORMATION SERVICE**  
US Department of Commerce  
Springfield, VA. 22151

**PRICES SUBJECT TO CHANGE**



*Technical Report No. 32-498*

*Theory of Orbit Determination—Part II  
Estimation Formulas*

*John D. Anderson*

A handwritten signature in black ink, reading "T. W. Hamilton". The signature is written in a cursive style with a horizontal line underneath it.

Thomas W. Hamilton, Chief  
Systems Analysis Section

JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA

October 1, 1963

Copyright © 1963  
Jet Propulsion Laboratory  
California Institute of Technology

Prepared Under Contract No. NAS 7-100  
National Aeronautics & Space Administration

**CONTENTS**

**I. Introduction . . . . . 1**

**II. Maximum Likelihood Estimation . . . . . 1**

**III. Solution of the Least-Squares-Estimation Formula . . . . . 4**

**IV. Schmidt-Kalman Estimation . . . . . 5**

**References . . . . . 7**

## ABSTRACT

A description of the parameter estimation formulas that are currently in use for the orbit determination of space vehicles is presented. The weighted least-squares-estimator with the inclusion of a priori parameter information is defined, and its relation to the maximum-likelihood-estimator is discussed. Also, the Schmidt-Kalman estimator is described and compared with the least-squares-procedure, showing under what conditions the two formulas are mathematically equivalent.

## I. INTRODUCTION

The practical determination of the orbit of a space vehicle from radar or optical data involves an overdetermined system. On the one hand, the orbit is defined by a number of parameters which include six orbital parameters, for example, the initial conditions in the Newtonian equations of motion, as well as several physical constants that enter into the description of the motion. On the other hand, the number of radar observations, for example, may be on the order of several thousand. The problem then is to define an estimation formula that will map these several thousand observations onto something on the order of a six dimensional space defined by the parameters of interest.

A selection of any estimation procedure can never be satisfactory if it is expected that the procedure will give the best representation of the actual orbit for all missions. It is possible to justify the selection of a particular estimation procedure only on purely statistical grounds, and then a number of idealizations are always required before an estimator can be described as optimum. Therefore, in this Report a definitive approach is taken with regard to estimation formulas, with emphasis on a description and comparison of methods. Information on the theory of estimation may be obtained from the applicable publications given in the list of References.

## II. MAXIMUM LIKELIHOOD ESTIMATION

There is in the theory of probability a formula known as Bayes's theorem which relates conditional probabilities as follows:

$$p(\mathbf{x}|\hat{\mathbf{z}}) = \frac{p(\mathbf{x})}{p(\hat{\mathbf{z}})} p(\hat{\mathbf{z}}|\mathbf{x}) \quad (1)$$

The notation  $p(A|B)$  indicates the conditional probability that  $A$  will occur given that  $B$  occurs, while  $p(A)$  indicates the unconditional probability that  $A$  will occur. In terms of orbit determination  $p(\hat{\mathbf{z}}|\mathbf{x})$  gives the probability that a data vector  $\hat{\mathbf{z}}$  occurs given that an arbitrary parameter

vector  $\mathbf{x}$  occurs. The probability of obtaining this parameter vector is given by  $p(\mathbf{x})$ , and the probability of obtaining the data vector  $\hat{\mathbf{z}}$  is  $p(\hat{\mathbf{z}})$ . If any parameter vector  $\mathbf{x}$  is equally likely, then  $p(\mathbf{x})$  will be a constant. However, it is usual to favor values of  $\mathbf{x}$  that are related to the physics of the problem so that  $p(\mathbf{x})$  is a variable. For example, let  $\mathbf{x}$  represent injection conditions of a space probe launched from the Earth. The tendency would be to assign low probabilities to values of  $\mathbf{x}$  that resulted in a launch below the surface of the Earth, for example, or to values beyond a certain height above the surface of the Earth. On the other hand, values of  $\mathbf{x}$  near the designed injection conditions might receive fairly high probability values.

In any case, if  $p(\hat{\mathbf{z}}|\mathbf{x})$  and  $p(\mathbf{x})$  are known then  $p(\mathbf{x}|\hat{\mathbf{z}})$  can be evaluated as a function of  $\mathbf{x}$ , and the probability that an arbitrary parameter vector occurs given that the data vector  $\hat{\mathbf{z}}$  occurs is known. Certainly  $\hat{\mathbf{z}}$  occurs, so the logical choice of a parameter vector  $\mathbf{x}$  is one  $\mathbf{x}^*$  which absolutely maximizes  $p(\mathbf{x}|\hat{\mathbf{z}})$ . Because  $\mathbf{x}^*$  is the vector that has the greatest probability of occurring, it is called the maximum likelihood estimate of  $\mathbf{x}$ .

A stationary value of  $p(\mathbf{x}|\hat{\mathbf{z}})$  can easily be obtained by considering the variation of Eq. (1) and by recognizing that  $p(\hat{\mathbf{z}})$ , the probability of obtaining the particular data vector  $\hat{\mathbf{z}}$ , is simply an invariant number  $P$ .

$$P\delta p(\mathbf{x}|\hat{\mathbf{z}}) = \left[ p(\hat{\mathbf{z}}|\mathbf{x}) \frac{dp(\mathbf{x})}{d\mathbf{x}} + p(\mathbf{x}) \frac{dp(\hat{\mathbf{z}}|\mathbf{x})}{d\mathbf{x}} \right] \delta\mathbf{x} \quad (2)$$

Then, because  $\delta\mathbf{x}$  is an arbitrary variation, the estimate  $\mathbf{x}^*$  of the parameter vector is the value of  $\mathbf{x}$  that results in a zero coefficient of  $\delta\mathbf{x}$ .

$$p(\hat{\mathbf{z}}|\mathbf{x}^*) \frac{dp(\mathbf{x}^*)}{d\mathbf{x}} + p(\mathbf{x}^*) \frac{dp(\hat{\mathbf{z}}|\mathbf{x}^*)}{d\mathbf{x}} = 0 \quad (3)$$

The derivative  $dp(\mathbf{x})/d\mathbf{x}$  is that of a scalar  $p$  with respect to a vector  $\mathbf{x}$  and is therefore a vector. As a matter of notation when  $\mathbf{x}^*$  is inserted in the place of the argument  $\mathbf{x}$  the meaning is that  $dp(\mathbf{x})/d\mathbf{x}$  is evaluated at  $\mathbf{x}^*$ , and similarly  $dp(\hat{\mathbf{z}}|\mathbf{x}^*)/d\mathbf{x}$  means that  $dp(\hat{\mathbf{z}}|\mathbf{x})/d\mathbf{x}$  is evaluated at  $\mathbf{x}^*$ .

In practice the probability functions  $p(\mathbf{x})$  and  $p(\hat{\mathbf{z}}|\mathbf{x})$  are not known, and in order to obtain an estimate  $\mathbf{x}^*$ , an assumption of the probability distribution is required. For a normal or Gaussian distribution on  $p(\mathbf{x})$  it is possible to write

$$p(\mathbf{x}) = \left[ \frac{Det(\tilde{\Gamma}_x^{-1})}{2\pi} \right]^{L/2} \exp \left[ -\frac{1}{2} (\mathbf{x} - \tilde{\mathbf{x}})^T \tilde{\Gamma}_x^{-1} (\mathbf{x} - \tilde{\mathbf{x}}) \right] \quad (4)$$

This is the normal distribution function for  $L$  variables  $x_1, x_2, \dots, x_L$  taken as elements of the column vector  $\mathbf{x}$  and having as a covariance matrix the  $L \times L$  matrix  $\tilde{\Gamma}_x$ . The notation  $Det(A)$  expresses the determinant of  $A$ .

In a normal function such as  $p(\mathbf{x})$  the particular value of  $\mathbf{x}$  designated by  $\tilde{\mathbf{x}}$  is called the mean value of  $\mathbf{x}$ , which is the value of  $\mathbf{x}$  that has the highest probability of occurring. For the case where  $\mathbf{x}$  represents orbit parameters,  $\tilde{\mathbf{x}}$  represents the most likely description of the orbit without the introduction of the observation vector  $\hat{\mathbf{z}}$ , and  $\tilde{\Gamma}_x$  represents the confidence that exists in this independent determination of  $\mathbf{x}$ . For example,  $\tilde{\mathbf{x}}$  might be the designed or preflight value of  $\mathbf{x}$  before observational data are obtained, and, in fact, for this reason it is often called the a priori estimate of the orbit parameters. Similarly  $\tilde{\Gamma}_x$  might be some assigned covariance matrix on this estimate, and so correspondingly it is called the a priori covariance matrix on the parameters. The matrix  $\tilde{\Gamma}_x$  might be constructed by considering the errors on the guidance system that establish the designed trajectory, or, in other words, by converting error sources in the guidance system to errors in the injection position and velocity. Naturally such an analysis would be expected to result in correlations in the six position and velocity elements, and so  $\tilde{\Gamma}_x$  would in general be nondiagonal.

An objection to using an a priori covariance matrix based on a guidance error analysis is easily raised by considering the meaning of an error or standard deviation on a guidance component. To obtain an empirical estimate of the error on such a component it is repeatedly subjected to flight conditions and by analyzing the many samples obtained its probability function is determined. Thus the matrix  $\tilde{\Gamma}_x$  is representative of an estimation of the variances and covariances arising from an averaging of all possible flights designed to yield a certain set of injection conditions. On the other hand the estimate  $\mathbf{x}^*$  of the parameters is obtained from only one flight, namely the one associated with the observation vector  $\hat{\mathbf{z}}$ , and thus to avoid an inconsistency the a priori estimate  $\tilde{\mathbf{x}}$  and its error matrix  $\tilde{\Gamma}_x$  should be influenced by the particular flight in question. Therefore, any in-flight monitoring of the guidance system should be heavily weighted in determining  $\tilde{\mathbf{x}}$  and  $\tilde{\Gamma}_x$ . For example, if a guidance component had clearly failed the predicted a priori parameter, vector  $\tilde{\mathbf{x}}$  could be considerably different than the designed value.

Actually, the use of the a priori estimate and statistics is fairly limited in the determination of the injection conditions of a space probe. Usually a nominal injection with a very large covariance matrix  $\Gamma_x$  is assumed. Thus, in the early orbit determination, when relatively few data are available, the solution to the estimate is held within bounds.

The greatest application of the a priori information is made in later estimations when a fair orbit has been determined from independent earlier data. For example, if the data vector  $\hat{z}$  represented observations made from a spacecraft in the vicinity of a distant planet and the parameter vector was the orbital parameters of the spacecraft with respect to that planet, then the a priori estimate of these parameters would be the determination made from earlier Earth-based observations, and the covariance matrix  $\Gamma_x$  would give the error description of that determination. This will become more apparent later on, however, when the estimation formula for  $\mathbf{x}^*$  is fully developed and applied to the orbit determination problem.

For the moment it is necessary to consider an expression for  $p(\hat{z}|\mathbf{x})$  based on a hypothetical normal distribution. Because it is assumed that the data vector  $\mathbf{z}$  has a mathematical representation  $\mathbf{z} = \mathbf{z}(\mathbf{x})$ , a reasonable distribution function is

$$p(\hat{z}|\mathbf{x}) = \left[ \frac{\text{Det}(\Gamma_z^{-1})}{2\pi} \right]^{N/2} \exp \left\{ -\frac{1}{2} \left[ \hat{z} - \mathbf{z}(\mathbf{x}) \right]^T \Gamma_z^{-1} \left[ \hat{z} - \mathbf{z}(\mathbf{x}) \right] \right\} \quad (5)$$

This function  $p(\hat{z}|\mathbf{x})$  says that if one selects a particular  $\mathbf{x}$  then the probability of  $\hat{z}$  occurring can be computed, and, in fact, the implication is that of all possible values of  $\mathbf{x}$ , there is one value that maximizes this probability. The rate at which the probability of obtaining  $\hat{z}$  decreases as  $\mathbf{x}$  departs from this maximizing value is governed by the  $N \times N$  data covariance matrix  $\Gamma_z$ . If the function  $p(\mathbf{x})$  were not involved in the development then the value of  $\mathbf{x}$  required for a maximum probability  $p(\hat{z}|\mathbf{x})$  would be the appropriate estimate of  $\mathbf{x}$ . Instead, Eq. (3) is applied to  $p(\mathbf{x})$ , and  $p(\hat{z}|\mathbf{x})$  and the estimate  $\mathbf{x}^*$  differs from the value that maximizes  $p(\hat{z}|\mathbf{x})$  alone. In other words, the estimate of  $\mathbf{x}$  is influenced by its a priori value. The derivatives of the probability functions are

$$\frac{dp(\mathbf{x})}{d\mathbf{x}} = p(\mathbf{x})\Gamma_x^{-1}(\mathbf{x} - \mathbf{x}) \quad (6)$$

$$\frac{dp(\hat{z}|\mathbf{x})}{d\mathbf{x}} = p(\hat{z}|\mathbf{x}) A^T \Gamma_z^{-1} [\hat{z} - \mathbf{z}(\mathbf{x})] \quad (7)$$

where  $A$  is an  $N \times L$  matrix with elements  $A_{ni} = \partial z_n / \partial x_i$ .

A substitution of Eq. (6) and (7) in Eq. (3) results in the formula for the estimate  $\mathbf{x}^*$  of the parameters

$$A^T(\mathbf{x}^*)\Gamma_z^{-1}[\hat{z} - \mathbf{z}(\mathbf{x}^*)] + \Gamma_x^{-1}(\mathbf{x} - \mathbf{x}^*) = 0 \quad (8)$$

As pointed out at the beginning of the development of the above estimation formula, no rigorous justification of its validity in the sense of being optimum or best is intended. The formula could have been written down at the outset as a definitive relation for the estimate  $\mathbf{x}^*$ , and what follows would have proceeded without any discussion of the meaning of Eq. (8). Hopefully, however, this presentation of the formula made it a little easier to accept. For a more mathematical approach there are several informative works on the theory of estimation which describe in detail the properties of estimation formulas, such as Eq. (8), and which discuss the conditions for "best" estimates (e.g., Ref. 2 and 4).

In any case, the size of the data vector  $\hat{z}$  must be considered in the establishment of a computer program to solve for  $\mathbf{x}^*$ . In particular, if the size of  $\hat{z}$  is large it is clear that the matrix  $\Gamma_z$ , which represents the assumed covariance matrix on the data, should be given in an easily invertible form. Customarily  $\Gamma_z$  is assumed diagonal so that the data are taken to be uncorrelated and the matrix  $\Gamma_z^{-1}$  is used directly in the form of a weighting matrix  $W$ . The elements of  $W$  may be simply the inverse of the squared error or standard deviation on each data sample, or if correlation is known to be present, the correlated points can be assigned smaller weights than the strictly independent data. The important point is that in constructing an orbit determination program it is agreed beforehand that the data will be weighted with a diagonal matrix  $W$ , and the numerical problem of inverting large matrices disappears. A new analytical problem is created, however, in that a weighting matrix must be chosen that is optimum in some sense and that reflects the peculiarities of the particular data obtained. The problem of finding an optimum weighting matrix is not considered here.

The other problem of including known data error descriptions in the matrix  $W$  is one that can be solved only by analyzing the relevant observational equipment in conjunction with experimental data obtained from its use.

In addition to the data weighting matrix  $W = \Gamma_z^{-1}$  there is a weighting matrix  $\Gamma_x^{-1}$  for a priori parameters.

The size of this matrix is such that it can be inverted if its determinant is significantly different from zero. In other words, orbit determination problems are concerned with the estimation of typically six orbital parameters, although a number of physical constants can be added as well. However, it is always feasible to limit the number of parameters to some reasonable number. Therefore,  $\tilde{\Gamma}_x$  is assumed to be the actual nondiagonal covariance

matrix on the a priori parameters  $\tilde{\mathbf{x}}$  and the estimation formula for  $\mathbf{x}^*$  is written

$$A^T(\mathbf{x}^*)W[\dot{\mathbf{z}} - \mathbf{z}(\mathbf{x}^*)] + \tilde{\Gamma}_x^{-1}(\tilde{\mathbf{x}} - \mathbf{x}^*) = 0 \quad (9)$$

Formula (9) defines the weighted least-squares estimation procedure with the inclusion of a priori parameter information.

### III. SOLUTION OF THE LEAST-SQUARES-ESTIMATION FORMULA

A system of  $L$  nonlinear algebraic equations in  $L$  unknowns  $(x_1^*, x_2^*, \dots, x_n^*)$ , is represented by Eq. (9) and the solution of this system for the parameter estimate  $\mathbf{x}^*$  naturally follows an iterative scheme. The procedure followed here applies the Newton-Raphson iteration formula for  $L$  variables. If a solution to a system of equations  $\mathbf{F}(\mathbf{x}) = 0$  is sought where  $\mathbf{F}$  and  $\mathbf{x}$  are both  $L$  vectors then the iteration formula is

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - \left(\frac{\partial \mathbf{F}}{\partial \mathbf{x}}\right)^{-1} \mathbf{F}(\mathbf{x}^{(n)}) \quad (10)$$

where  $\mathbf{x}^{(n)}$  is the  $n$ th approximation to the solution  $\mathbf{x}$  and  $\mathbf{x}^{(n+1)}$  is the improved  $(n+1)$ st approximation. The matrix  $\partial \mathbf{F} / \partial \mathbf{x}$  is normally evaluated at  $\mathbf{x}^{(n)}$  although the procedure can be modified by evaluating  $\partial \mathbf{F} / \partial \mathbf{x}$  only once at the initial approximation to  $\mathbf{x}$ . (It should be noted that the derivative of a vector with respect to a vector is introduced as a notational convenience only.)

A discussion of the convergence properties of Eq. (10) is fairly involved if the dimension of the system is greater than unity, although for the special case where  $x$  is a scalar it can be shown quite easily that the iteration formula is second order. In other words, the error in the  $(n+1)$ st approximation to the solution is proportional to the square of the error in the  $n$ th approximation, and if the initial guess to  $x$  is good enough, convergence is assured. In practice, this sort of statement applies to the  $L$  dimensional system of equations as well, and for this reason the Newton-Raphson formula is used in the orbit determination problem. The convergence properties of

Eq. (10) are not discussed here, although any standard numerical analysis text can be consulted for information on convergence.

Now the function  $\mathbf{F}(\mathbf{x}^*)$  is simply the left-hand side of Eq. (9), and all that is required to apply formula (10) is the matrix  $\partial \mathbf{F} / \partial \mathbf{x}$  or, equivalently, the derivative of Eq. (9) with respect to the variable  $\mathbf{x}$ . In other words, the function  $\mathbf{F}(\mathbf{x})$  is

$$\mathbf{F}(\mathbf{x}) = A^T(\mathbf{x})W[\dot{\mathbf{z}} - \mathbf{z}(\mathbf{x})] + \tilde{\Gamma}_x^{-1}(\tilde{\mathbf{x}} - \mathbf{x}) \quad (11)$$

and

$$\frac{\partial \mathbf{F}}{\partial \mathbf{x}} = -A^T(\mathbf{x})WA(\mathbf{x}) - \tilde{\Gamma}_x^{-1} \quad (12)$$

The variation of  $A(\mathbf{x})$  with  $\mathbf{x}$  has been neglected in Eq. (12). Its inclusion in  $\partial \mathbf{F} / \partial \mathbf{x}$  results in considerable complication because of the way it enters and because it requires the second partial derivatives of the data vector  $\mathbf{z}$  with respect to the parameters  $\mathbf{x}$ . Experience has shown that the iteration formula, which results from a neglect of the variation in  $A(\mathbf{x})$ , is satisfactory in providing rapid convergence to the solution  $\mathbf{x}^*$  of Eq. (9). Therefore, the contribution of the second partial derivatives is only noted here before proceeding to the usual estimation formula. Designate the second derivatives of the observation vector  $\mathbf{z}(\mathbf{x})$  with respect to the  $r$  and  $s$  components of  $\mathbf{x}$  by the vector  $\mathbf{d}_{rs}$ .

$$\mathbf{d}_{rs} = \frac{\partial^2 \mathbf{z}}{\partial x_r \partial x_s} \quad (13)$$

Then to include  $\mathbf{d}_{rs}$  in the iteration formula each  $r,s$  element  $(\partial \mathbf{F} / \partial \mathbf{x})_{rs}$  of the matrix  $\partial \mathbf{F} / \partial \mathbf{x}$  is modified so that

$$\left(\frac{\partial \mathbf{F}}{\partial \mathbf{x}}\right)_{rs} = - [A^T(\mathbf{x})WA(\mathbf{x}) + \Gamma_x^{-1}]_{rs} + \mathbf{d}_{rs}^T W [\hat{\mathbf{z}} - \mathbf{z}(\mathbf{x})] \quad (14)$$

An interesting variation of formula (15) has been used by Schmidt, Smith and others in the determination of orbits by a continual updating of the estimate of the orbit. In other words, at some observation time  $t$  the a priori estimate  $\mathbf{x}$  of the orbit, as obtained from previous data, is improved by adding the observations taken at time  $t$ . An immediate application of this method (also known as the method of sequential estimation) is in the navigation of a manned space vehicle where the orbit is periodically re-established by taking one or more observations from the vehicle and by processing the data through a small on board computer.

The form of the estimator was established by Kalman with the aid of linear statistical filter theory, and, therefore, to derive his formula from the least-squares formulation it is necessary to linearize formula (15) by setting the estimate  $\mathbf{x}^*$  of  $\mathbf{x}$  equal to  $\mathbf{x}^{(n+1)}$  and the a priori estimate  $\mathbf{x}$  of  $\mathbf{x}$  equal to  $\mathbf{x}^{(n)}$ . This is justified if the a priori estimate of the orbit is sufficiently close to the new estimate  $\mathbf{x}^*$ , or, equivalently, if the new data are sufficiently consistent with the past data so that only one iteration is required in formula (15). If this is not the case, if, for example, a relatively uncertain midcourse maneuver were applied just before the new observation time, then the estimate of the orbit would deviate from the least-squares estimate that satisfies Eq. (9). Further, because  $\mathbf{x}^{(n)}$  is assumed equal to  $\mathbf{x}$  the second term in the brackets of Eq. (15) is always zero, and a repeated application of the Schmidt-Kalman estimator would not result in convergence to the solution of Eq. (9). However, with these reservations in mind the linear form of formula (15) is written

$$\delta \mathbf{x} = \mathbf{x}^* - \mathbf{x} = (A^T W A + \Gamma_x^{-1})^{-1} A^T W \delta \mathbf{z} \quad (16)$$

The final iteration formula thus assumes that the second term in Eq. (14) is small and there results the standard relation

$$\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)} = [A^T(\mathbf{x}^{(n)})WA(\mathbf{x}^{(n)}) + \Gamma_x^{-1}]^{-1} \left\{ A^T(\mathbf{x}^{(n)})W [\hat{\mathbf{z}} - \mathbf{z}(\mathbf{x}^{(n)})] + \Gamma_x^{-1}(\mathbf{x} - \mathbf{x}^{(n)}) \right\} \quad (15)$$

#### IV. SCHMIDT-KALMAN ESTIMATION

where, of course, the terms on the right-hand side of the formula are evaluated at  $\mathbf{x} = \mathbf{x}$  and the residual vector  $\delta \mathbf{z}$  is

$$\delta \mathbf{z} = \hat{\mathbf{z}} - \mathbf{z}(\mathbf{x}) \quad (17)$$

Note that for one observation and six parameters both  $\delta \mathbf{z}$  and  $W$  are scalars,  $\Gamma_x$  is a six by six matrix and  $A$  is a six-dimensional row vector. The Schmidt-Kalman form of the correction  $\delta \mathbf{x}$  is obtained by writing the matrix  $(A^T W A + \Gamma_x^{-1})^{-1}$  in a different form.

Consider the matrix

$$M = \begin{pmatrix} \Gamma_x^{-1} A^T \\ A - W^{-1} \end{pmatrix} \quad (18)$$

If the dimension of  $\Gamma_x$  is  $L \times L$  while that of  $W$  is  $N \times N$ , then the matrix  $A$  is of dimension  $N \times L$  while that of  $M$  is  $(L + N) \times (L + N)$ . If an inverse to  $M$  exists then it can be written

$$M^{-1} = \begin{pmatrix} X_1 & Y^T \\ Y & X_2 \end{pmatrix} \quad (19)$$

where the dimensions of  $X_1$ ,  $X_2$ , and  $Y$  are respectively  $L \times L$ ,  $N \times N$  and  $N \times L$ .

Now

$$M M^{-1} = I_{L+N} \quad (20)$$

where  $I_n$  indicates the  $n \times n$  unit matrix. Forming the matrix products involved in  $M M^{-1}$  yields

$$\Gamma_x^{-1} X_1 + A^T Y = I_L \quad (21)$$

$$A Y^T - W^{-1} X_2 = I_N \quad (22)$$

$$\Gamma_x^{-1} Y^T + A^T X_2 = 0 \quad (23)$$

$$AX_1 - W^{-1}Y = 0 \quad (24)$$

From Eq. (24), if the inverse of  $W^{-1}$  exists, then

$$Y = WAX_1 \quad (25)$$

and if the inverse of  $\Gamma_x^{-1}$  also exists, then Eq. (21) and (25) yield

$$(A^TWA + \Gamma_x^{-1}) X_1 = I_L \quad (26)$$

or for  $(A^TWA + \Gamma_x^{-1})$  nonsingular

$$X_1 = (A^TWA + \Gamma_x^{-1})^{-1} \quad (27)$$

a second expression for  $X_1$  can be obtained from Eq. (21), (22) and (23).

From Eq. (23)

$$Y^T = -\Gamma_x A^T X_2 \quad (28)$$

and  $X_2$  can be found from Eq. (22).

$$X_2 = -(W^{-1} + A\Gamma_x A^T)^{-1} \quad (29)$$

Substitute  $X_2$  in Eq. (28) to obtain

$$Y = (W^{-1} + A\Gamma_x A^T)^{-1} A\Gamma_x \quad (30)$$

where  $\Gamma_x$ ,  $W$  and  $A\Gamma_x A^T$  are all symmetric matrices.

Now solve Eq. (21) for  $X_1$

$$X_1 = \Gamma_x - \Gamma_x A^T (W^{-1} + A\Gamma_x A^T)^{-1} A\Gamma_x \quad (31)$$

and simply equate the two expressions for  $X_1$ .

$$(A^TWA + \Gamma_x^{-1})^{-1} = \Gamma_x - \Gamma_x A^T (W^{-1} + A\Gamma_x A^T)^{-1} A\Gamma_x \quad (32)$$

or define

$$\Delta^* = \Gamma_x A^T (W^{-1} + A\Gamma_x A^T)^{-1} \quad (33)$$

so that

$$(A^TWA + \Gamma_x^{-1})^{-1} = \Gamma_x - \Delta^* A\Gamma_x \quad (34)$$

then the estimator  $\delta \mathbf{x}$  from formula (16) can be written

$$\delta \mathbf{x} = (\Gamma_x - \Delta^* A\Gamma_x) A^T W \delta \mathbf{z} \quad (35)$$

Schmidt and Kalman call  $(\Gamma_x - \Delta^* A\Gamma_x)$  the covariance matrix  $\Gamma_x$  on the parameters after the data  $\hat{\mathbf{z}}$  are added.

$$\Gamma_x = (A^TWA + \Gamma_x^{-1})^{-1} = \Gamma_x - \Delta^* A\Gamma_x \quad (36)$$

It will be shown in a later part of this sequence of reports that  $\Gamma_x$  is the covariance matrix only under the assumption

that there are no systematic errors in the residual vector  $\delta \mathbf{z}$ . However, for the moment it will be assumed that Eq. (36) is valid and write

$$\delta \mathbf{x} = \Gamma_x A^T W \delta \mathbf{z} \quad (37)$$

The expression  $\Gamma_x A^T W$  can be simplified. Set

$$\Delta = \Gamma_x A^T W \quad (38)$$

Then

$$\Delta = [\Gamma_x - \Gamma_x A^T (W^{-1} + A\Gamma_x A^T)^{-1} A\Gamma_x] A^T W \quad (39)$$

or

$$\Delta = \Gamma_x A^T [W - (W^{-1} + A\Gamma_x A^T)^{-1} A\Gamma_x A^T W] \quad (40)$$

Further

$$\Delta = \Gamma_x A^T (W^{-1} + A\Gamma_x A^T)^{-1} [(W^{-1} + A\Gamma_x A^T) W - A\Gamma_x A^T W] \quad (41)$$

and

$$\Delta = \Gamma_x A^T (W^{-1} + A\Gamma_x A^T)^{-1} \quad (42)$$

so that

$$\Delta = \Delta^* \quad (43)$$

and the estimation formula becomes

$$\delta \mathbf{x} = \Delta^* \delta \mathbf{z} \quad (44)$$

The procedure for continually updating an orbit can be expressed as a closed loop process which flows according to the following steps.

*Step 1.* Given an estimate of the orbit  $\mathbf{x}^*$  at time  $t$  integrate the equations of motion of the vehicle to a time  $t + \Delta t$ , the next observation time. This yields an a priori estimate  $\mathbf{x}$  of the orbital parameters at time  $t + \Delta t$ , as well as the residual vector  $\delta \mathbf{z} = \mathbf{z} - \mathbf{z}(\mathbf{x})$ .

*Step 2.* Map the covariance matrix  $\Gamma_x$  at time  $t$  to the new time  $t + \Delta t$  to yield the a priori covariance matrix  $\Gamma_x$  at time  $t + \Delta t$ . Again if systematic errors are neglected

$$\Gamma_x = U \Gamma_x U^T$$

where  $U$  is a matrix that relates differential increments in the orbital parameters referenced to time  $t + \Delta t$  to parameters referenced to  $t$ . The  $r, s$  element of  $U$  is

$$U_{rs} = \frac{\partial x_r(t + \Delta t)}{\partial x_s(t)} \quad r, s = 1, 2, \dots, L$$

For example,  $\mathbf{x}(t + \Delta t)$  represents the Cartesian components of position and velocity at time  $t + \Delta t$ , while  $\mathbf{x}(t)$

represents the position and velocity components at time  $t$ . In this case  $U$  is a six by six matrix.

*Step 3.* Compute the matrix  $A$ . Because the parameters  $\mathbf{x}$  are referenced to the observation time  $t + \Delta t$ , the  $A$  matrix simply contains the partial derivatives of the observation vector with respect to the parameters at time  $t + \Delta t$ .

$$A_{rs} = \left( \frac{\partial z_r}{\partial x_s} \right)_{t + \Delta t} \quad \begin{array}{l} r = 1, 2, \dots, N \\ s = 1, 2, \dots, L \end{array}$$

The number  $N$  is greater than or equal to unity but less than some number which represents the upper limit for a feasible numerical  $N \times N$  matrix inversion in step 4.

*Step 4.* Compute the weighting function

$$\Delta^* = \tilde{\Gamma}_x A^T (W^{-1} + A \tilde{\Gamma}_x A^T)^{-1}$$

*Step 5.* Compute an improved estimate of the orbit at time  $t + \Delta t$

$$\mathbf{x}^* = \tilde{\mathbf{x}} + \Delta^* \delta \mathbf{z}$$

*Step 6.* Compute the new covariance matrix for the estimated parameters  $\mathbf{x}^*$ .

$$\Gamma_x = \tilde{\Gamma}_x - \Delta^* A \tilde{\Gamma}_x$$

*Step 7.* Set  $t + \Delta t$  equal to  $t$  and go back to *step 1*.

## REFERENCES

1. Brouwer, D., and Clemence, D. M., *Methods of Celestial Mechanics*, Academic Press, New York, 1961, Chapter VIII.
2. Cramer, H., *Mathematical Methods of Statistics*, Princeton University Press, 1947, Chapters 32-34.
3. Kalman, R. E., *New Methods and Results in Linear Prediction and Filtering Theory*, RIAS Inc., Baltimore, Md., Technical Report 61-1, 1960.
4. Kendall, M. G., *The Advanced Theory of Statistics*, Vol. 2, New York, 1961, Chapters 17-19.
5. McLean, J. D., Schmidt, S. F., and McGee, L. A., *Optimal Filtering and Linear Prediction Applied to a Midcourse Navigation System for the Circumlunar Mission*, Technical Manual D-1208, NASA, Washington, D. C., 1962.
6. Shapiro, I. I., *The Prediction of Ballistic Missile Trajectories from Radar Observations*, McGraw-Hill Book Co., New York, 1957.
7. Smith, G. L., Schmidt, S. F., and McGee, L. A., *Application of Statistical Filter Theory to the Optimal Estimation of Position and Velocity on Board a Circumlunar Vehicle*, Technical Report R-135, NASA, Washington, D. C., 1962.