Technical Report No. 32-445

Sequential Estimation of Correlated Stochastic Variables

C. G. Pfeiffer

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

July 1, 1963

Technical Report No. 32-445

Sequential Estimation of Correlated Stochastic Variables

C. G. Pfeiffer

T. W. Hamilton

T. W. Hamilton, *Chief*
Systems Analysis Section

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

July 1, 1963

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

CONTENTS

FIGURES

ABSTRACT

A technique is developed for estimating the components of the sequence of correlated random vectors $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n, \cdots\}$, when given the sequence of linearly related data vectors $\{\phi_1, \phi_2, \cdots, \phi_n, \cdots\}$. It is shown that a necessary and sufficient condition that the minimum variance estimate of $\mathbf{x}_n$ depend only upon ϕ_n and the previously computed estimate of $\mathbf{x}_{n-1}$ is that all $\mathbf{x}_i$ be "sequentially correlated." This is a condition placed upon the covariance matrix describing the otherwise unspecified physical process that generates the $\mathbf{x}_i$. Examples of sequentially correlated processes are given, and an application to the deep-space orbit determination problem is discussed.

I. SUMMARY[1]

This Report develops a technique for sequentially obtaining the minimum variance estimate of the components of the vectors $\mathbf{x}_i$ in the time-ordered sequence of correlated stochastic vectors

$$\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n, \cdots\}$$

when given observed data linearly related to the $\mathbf{x}_i$. The stochastic vectors $\mathbf{x}_i$ are the values $\mathbf{x}(t_i)$ at the discrete times

$$\{t_1, t_2, \cdots, t_n, \cdots\}$$

where $\mathbf{x}(t)$ is the output of some unspecified physical process. It is assumed that

$$\overline{\mathbf{x}_i} = 0$$

and that the correlations

$$[\overline{\mathbf{x}_i\, \mathbf{x}_i^T}] = \Lambda_i$$

and

$$[\overline{\mathbf{x}_j\, \mathbf{x}_i^T}] = P_{ji}$$

are known. The "noise" on the observations is considered to be imbedded in the $\mathbf{x}_i$, thus the observations are of the form

$$\phi_i = \phi(t_i) = A_i\, \mathbf{x}_i$$

where A_i is a known matrix.

[1]Notation: All random variables are assumed to be column vectors which are denoted by boldface type. Matrices are denoted by italicized capital letters; I is the identity matrix. The superscript T indicates the transpose of a matrix or vector, and the subscript i refers to the time t_i. The bar $(-)$ over a quantity indicates the statistical average over the ensemble of all experiments. The asterisk indicates the minimum variance estimate of the quantity $\mathbf{x}_i$, considering all data up to and including time t_i, while the $\hat{}$ indicates the estimate of the $\mathbf{x}_i$ based only upon $\mathbf{x}_{i-1}$. See the Nomenclature for definitions of symbols.

Since the data are gathered and estimates are made only at the discrete times t_i, the Gauss-Markoff theorem can be applied to yield a minimum variance estimate of any $\mathbf{x}_i$. This estimate depends, in general, upon all observed data up to and including time t_i. The situation becomes unwieldly if a large number of observations are gathered, but Kalman (Ref. 1) shows that a relatively simple iterative estimation procedure can be performed if it is postulated that each bit of observed data arises from sampling the output of a linear system excited by uncorrelated, Gaussian error sources (see Part V, the linear control process). In this case, only the statistics of the most recent estimate of the state of the system need be considered when processing each new data point, making the situation much more tractable. It is the purpose here to develop a similar iterative estimation technique without constructing such a system model, thereby demonstrating the most general type of stochastic process that can be treated in this fashion and producing estimation formulas that do not depend upon dealing with uncorrelated error sources.

Given the sequence of data vectors

$$\{\phi_1, \phi_2, \cdots, \phi_n, \cdots\}$$

where

$$\phi_i = A_i \mathbf{x}_i$$

it is shown that a necessary and sufficient condition that the minimum variance estimate of $\mathbf{x}_j$ depend only upon the observation ϕ_j and the previous estimate $\mathbf{x}_i^*$ is that all $\mathbf{x}_i$ be "sequentially correlated", that is,

$$R_{ki} = R_{kj} R_{ji} \qquad \text{for all } t_i \leq t_j \leq t_k \tag{1}$$

where

$$R_{ji} = P_{ji} \Lambda_i^{-1}$$

is defined to be the "normalized correlation" between $\mathbf{x}_j$ and $\mathbf{x}_i$. In this case, the estimate of $\mathbf{x}_j$, given the previous estimate $\mathbf{x}_i^*$ but not the observation ϕ_j, is

$$\hat{\mathbf{x}}_j = R_{ji} \mathbf{x}_i^* \tag{2}$$

Defining the errors

$$\hat{\epsilon}_j = (\hat{\mathbf{x}}_j - \mathbf{x}_j)$$

and

$$\epsilon_i = (\mathbf{x}_i^* - \mathbf{x}_i)$$

then

$$\left[\overline{\hat{\epsilon}_j \hat{\epsilon}_j^T}\right] = \Lambda_j - R_{ji} \left\{\Lambda_i - \left[\overline{\epsilon_i \epsilon_i^T}\right]\right\} R_{ji}^T \tag{3}$$

The minimum variance estimate of $\mathbf{x}_j$ then becomes

$$\mathbf{x}_j^* = \hat{\mathbf{x}}_j + W_j(\phi_j - A_j \hat{\mathbf{x}}_j) \tag{4}$$

where

$$W_j = \left[\overline{\hat{\epsilon}_j \hat{\epsilon}_j^T}\right] A_j^T \left[A_j \left[\overline{\hat{\epsilon}_j \hat{\epsilon}_j^T}\right] A_j^T\right]^{-1} \tag{5}$$

The covariance of the error in $\mathbf{x}_j^*$ is

$$\left[\overline{\epsilon_j \epsilon_j^T}\right] = \left[\overline{\hat{\epsilon}_j \hat{\epsilon}_j^T}\right] - W_j A_j \left[\overline{\hat{\epsilon}_j \hat{\epsilon}_j^T}\right] \tag{6}$$

The process is repeated at time $t_k > t_j$, and the estimation proceeds iteratively to conclusion.

Examples of sequentially correlated processes are given, and the relation to Kalman's result is discussed.

II. THE GAUSS-MARKOFF THEOREM

The sequential estimation technique discussed in this Report is developed from the well-known Gauss-Markoff theorem (Ref. 2). In this Part, a general form of the theorem will be presented.

Suppose a data vector ϕ is observed on some experiment, and the value of some unknown random vector $\mathbf{x}$ is to be estimated from the given ϕ, where ϕ and $\mathbf{x}$ have known statistical correlation over all such experiments. It will be assumed that these quantities have zero *a priori* mean, that is, $\overline{\phi} = 0$ and $\overline{\mathbf{x}} = 0$. The specification of a linear, unbiased minimum variance estimate leads to a unique estimate of $\mathbf{x}$. Let this estimate be

$$\mathbf{x}^* = K\phi$$

where K is a matrix to be determined. Note that $\mathbf{x}^*$ is unbiased, that is,

$$\overline{\mathbf{x}}^* = 0$$

DEFINITION 1: The linear unbiased estimate $\mathbf{x}^*$ is said to be a minimum variance estimate if the error covariance matrix $[\overline{\epsilon\epsilon^T}]$ is a minimum over all experiments, where

$$\epsilon = \mathbf{x}^* - \mathbf{x}$$

This means that for any other estimate

$$\mathbf{x}_L^* = L\phi$$

with error

$$\epsilon_L = (\mathbf{x}_L^* - \mathbf{x})$$

the quadratic form associated with $\{[\overline{\epsilon_L \epsilon_L^T}] - [\overline{\epsilon\epsilon^T}]\}$ is positive semidefinite.

Theorem 1: The minimum variance estimate of $\mathbf{x}$, given ϕ, is

$$\mathbf{x}^* = [\overline{\mathbf{x}\phi^T}]\,[\overline{\phi\phi^T}]^{-1}\,\phi \qquad (7)^2$$

The error in the $\mathbf{x}^*$ has covariance

$$[\overline{\epsilon\epsilon^T}] = [\overline{\mathbf{x}\mathbf{x}^T}] - [\overline{\mathbf{x}\phi^T}]\,[\overline{\phi\phi^T}]^{-1}\,[\overline{\phi\mathbf{x}^T}] \qquad (8)$$

Proof: Suppose the *a priori* covariance matrix Λ is given, where

$$\Lambda = \begin{bmatrix} [\overline{\phi\phi^T}] & [\overline{\phi\mathbf{x}^T}] \\ [\overline{\mathbf{x}\phi^T}] & [\overline{\mathbf{x}\mathbf{x}^T}] \end{bmatrix} = \begin{bmatrix} \Lambda_\phi & P_{\phi x} \\ P_{x\phi} & \Lambda_x \end{bmatrix} \qquad (9)$$

Let

$$\mathbf{x}^* = K\phi$$

where the elements of K are to be determined. Then

$$[\overline{\epsilon\epsilon^T}] = K\Lambda_\phi K^T - KP_{\phi x} - P_{x\phi}K^T + \Lambda_x \qquad (10)$$

Let $\mathbf{z}$ be an arbitrary vector. Then from Definition 1 it follows that the quadratic form $\mathbf{z}^T\,[\overline{\epsilon\epsilon^T}]\,\mathbf{z}$ must be a minimum with respect to all elements of K. This implies

$$0 = [K\Lambda_\phi - P_{x\phi}]\,\delta K + \{[K\Lambda_\phi - P_{x\phi}]\,\delta K\}^T \qquad (11)$$

Thus

$$K = P_{x\phi}\,\Lambda_\phi^{-1} \qquad (12)$$

which will be defined as the "normalized correlation" between $\mathbf{x}$ and ϕ. Equation (8) follows from substituting Eq. (12) into Eq. (10). This completes the theorem.

Theorem 1 is often presented in a different form when discussing the process (Ref. 3)

$$\phi = A\mathbf{x} + \mathbf{n} \qquad (13)$$

where $\mathbf{x}$ is constant over any given experiment and $\mathbf{n}$ is noise on the observations that is not correlated with $\mathbf{x}$. If the *a priori* statistics over all experiments are

$$[\overline{\mathbf{x}\mathbf{x}^T}] = \Lambda_0$$

and

$$[\overline{\mathbf{n}\mathbf{n}^T}] = \Gamma$$

(Γ is usually assumed to be a diagonal matrix), then

$$[\overline{\mathbf{x}\phi^T}] = \Lambda_0 A^T \qquad (14)$$

$$[\overline{\phi\phi^T}] = A\Lambda_0 A^T + \Gamma \qquad (15)$$

From Eq. (7) and Eq. (8),

$$\mathbf{x}^* = \Lambda_0 A^T [A\Lambda_0 A^T + \Gamma]^{-1}\,\phi \qquad (16)$$

$$[\overline{\epsilon\epsilon^T}] = \Lambda_0 - \Lambda_0 A^T [A\Lambda_0 A^T + \Gamma]^{-1}\,A\Lambda_0 \qquad (17)$$

By a matrix identity (Ref. 4, p. 79),

$$\mathbf{x}^* = [\Lambda_0^{-1} + A^T\Gamma^{-1}A]^{-1}\,[A^T\Gamma^{-1}\,\phi] \qquad (18)$$

$$[\overline{\epsilon\epsilon^T}] = [\Lambda_0^{-1} + A^T\Gamma^{-1}A]^{-1} \qquad (19)$$

If $\Lambda_0 \to \infty$ (no *a priori* information on $\mathbf{x}$), then Eq. (18) becomes

$$\mathbf{x}^* = [A^T\Gamma^{-1}\,A]^{-1}\,[A^T\Gamma^{-1}\,\phi] \qquad (20)$$

which is the classical result. If Γ is diagonal with equal elements, Eq. (20) becomes the familiar least-squares estimate.

[2] As in Eq. (7), it will be assumed throughout this Report that the inverse of all *a priori* covariance matrices exists. There is no loss of generality in this assumption, since the condition can always be realized in theory and in practice by eliminating the redundant (perfectly correlated) variables from the problem.

III. SEQUENTIAL ESTIMATION

This Report is concerned with the estimation of certain time-varying vector random variables

$$\mathbf{x}_i = \mathbf{x}(t_i)$$

when given the sequence of observed data vectors

$$\{\phi(t_1), \phi(t_2), \cdots, \phi(t_n), \cdots\}$$

where

$$\phi_i = A_i \mathbf{x}_i$$

An iterative estimation procedure will be sought, where only the most recent estimate of the state of the system will be considered in processing each new data point. It will be shown in Part IV that this is possible if and only if the quantities to be estimated are "sequentially correlated." An intuitive derivation of the result will be presented here, and a more rigorous treatment will be given in Part IV.

The motivation for this approach is the analysis of the orbit determination process (Ref. 3) in the presence of correlated measurement errors. The problem arises as follows. Let

$$\phi_i = \left[\frac{\partial \phi_i}{\partial \mathbf{y}_i}\right] \mathbf{y}_i + \mathbf{n}_i \tag{21}$$

$$\mathbf{y}_i = \left[\frac{\partial \mathbf{y}_i}{\partial \mathbf{y}_0}\right] \mathbf{y}_0 \tag{22}$$

where $\mathbf{y}_i$ is the coordinate deviation vector at time t_i and $[\partial \mathbf{y}_i / \partial \mathbf{y}_0]$ is the state transition matrix that relates the initial condition variations $\mathbf{y}_0$ to $\mathbf{y}_i$. Equation (19) is easily put into the form of Eq. (13) by constructing the total observation vector

$$\phi^T = [\phi_1^T, \phi_2^T, \cdots, \phi_n^T, \cdots]$$

The dimension of ϕ = (number of data points/sample time) $\times$ (number of sample times), but the estimate (Eq. 19) is quite tractable if the noise on the data is uncorrelated. If this is not the case, however, it is convenient to define

$$A_i = \left[\left[\frac{\partial \phi_i}{\partial \mathbf{y}_i}\right], I\right]$$

$$\mathbf{x}_i^T = (\mathbf{y}_i^T, \mathbf{n}_i^T)$$

and thus

$$\phi_i = A_i \mathbf{x}_i$$

which is the problem formulation discussed above.

With this example in mind, consider a more general process (see Fig. 1), where observations are being made during an experiment in an effort to estimate the correlated stochastic variables

$$\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n, \cdots\}$$

The first moment of the stochastic process is given by

$$\overline{\mathbf{x}_i} = 0$$

and the second moment by

$$[\overline{\mathbf{x}_j \mathbf{x}_i^T}] = P_{ji}$$

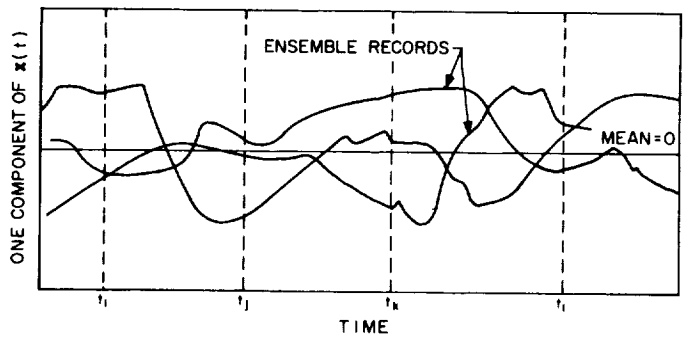

Fig. 1. The stochastic process

It will be convenient to define

$$\Lambda_i = P_{ii}$$

and

$$R_{ji} = P_{ji} \Lambda_i^{-1}$$

Suppose an estimate $\mathbf{x}_i^*$ has been obtained, and, ignoring all other information, an estimate of $\mathbf{x}_j$ ($t_j > t_i$) is to be made based on $\mathbf{x}_i^*$. From Eq. (7), this ought to be

$$\hat{\mathbf{x}}_j = R_{ji} \mathbf{x}_i^* \tag{23}$$

where R_{ji} is the "normalized" *a priori* correlation between $\mathbf{x}_i$ and $\mathbf{x}_j$. The $\hat{}$ notation indicates that new data at t_j have not been considered in estimating $\mathbf{x}_j$. The error expression is

$$\hat{\epsilon}_j = (\hat{\mathbf{x}}_j - \mathbf{x}_j) = R_{ji}(\mathbf{x}_i + \epsilon_i) - \mathbf{x}_j \tag{24}$$

If it is assumed that ϵ_i is not correlated with $\mathbf{x}_i$ or $\mathbf{x}_j$, the covariance matrix $\left[\hat{\epsilon}_j \hat{\epsilon}_j^T\right]$ is given by Eq. (3). Equations (4) through (6) now follow directly by an application of Theorem 1. In effect, this technique treats each $\hat{\mathbf{x}}_j$ as an observation of $\mathbf{x}_j$ and optimally combines the $\hat{\mathbf{x}}_j$ and ϕ_j to obtain the estimate $\mathbf{x}_j^*$.

Certain elementary special cases offer an intuitive check on this result:

1. No *a posteriori* information on $\mathbf{x}_i$ is available. Then

$$\mathbf{x}_i^* = 0$$

$$\hat{\mathbf{x}}_j = 0$$

$$\left[\overline{\epsilon_i \epsilon_i^T}\right] = \Lambda_i$$

and

$$\left[\overline{\hat{\epsilon}_j \hat{\epsilon}_j^T}\right] = \Lambda_j$$

2. No correlation exists between $\mathbf{x}_j$ and $\mathbf{x}_i$. Then

$$R_{ji} = 0$$

$$\hat{\mathbf{x}}_j = 0$$

and

$$\left[\overline{\hat{\epsilon}_j \hat{\epsilon}_j^T}\right] = \Lambda_j$$

3. Perfect correlation exists between $\mathbf{x}_j$ and $\mathbf{x}_i$ (they are the same quantities). Then

$$R_{ji} = I$$

$$\hat{\mathbf{x}}_j = \mathbf{x}_i^*$$

and

$$\left[\overline{\hat{\epsilon}_j \hat{\epsilon}_j^T}\right] = \left[\overline{\epsilon_i \epsilon_i^T}\right]$$

The conditions under which the sequential estimation technique can be applied will be developed in Part IV. It should be observed, however, that a sequential estimate can always be made in engineering applications, and the resulting error variance may be near minimum.

IV. PROOF OF THE FUNDAMENTAL THEOREM

In this Part, the results stated in the Summary will be established by proving the fundamental theorem of sequential estimation (Theorem 4). The derivation calls upon well-known "escalator methods" for iteratively inverting symmetric matrices (Ref. 4).

DEFINITION 2: The stochastic vectors, $\mathbf{x}_a$, $\mathbf{x}_b$, $\mathbf{x}_c$, defined at times t_a, t_b, t_c, respectively, are said to be sequentially correlated when

$$P_{ca} = P_{cb}\, \Lambda_b^{-1}\, P_{ba} \tag{25}$$

for $t_a \leq t_b \leq t_c$, where

$$\Lambda_b = \left[\overline{\mathbf{x}_b \mathbf{x}_b^T}\right]$$

$$P_{ca} = \left[\overline{\mathbf{x}_c \mathbf{x}_a^T}\right]$$

and P_{cb} and P_{ba} are similarly defined.

Theorem 2: The minimum variance estimate of $\mathbf{x}_c$, given $(\mathbf{x}_a, \mathbf{x}_b)$, depends only upon $\mathbf{x}_b$ if and only if these quantities are sequentially correlated. The estimate in this case is

$$\hat{\mathbf{x}}_c = R_{cb}\, \mathbf{x}_b \tag{26}$$

where

$$R_{cb} = P_{cb} \Lambda_b^{-1}$$

Proof: Let the covariance matrix describing $\mathbf{x}_a$, $\mathbf{x}_b$, $\mathbf{x}_c$ be

$$\Lambda_{abc} = \begin{bmatrix} \Lambda_a & P_{ab} & P_{ac} \\ P_{ba} & \Lambda_b & P_{bc} \\ P_{ca} & P_{cb} & \Lambda_c \end{bmatrix} \tag{27}$$

From Eq. (7),

$$\hat{\mathbf{x}}_c = \left[P_{ca} \mid P_{cb}\right] \begin{bmatrix} \Lambda_a & P_{ab} \\ P_{ba} & \Lambda_b \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix} \tag{28}$$

Equation (28) reduces to [3]

$$\hat{\mathbf{x}}_c = (P_{ca} - P_{cb}\Lambda_b^{-1}P_{ba})(B)\mathbf{x}_a + (-P_{ca} + P_{cb}\Lambda_b^{-1}P_{ba})(B)(P_{cb}\Lambda_b^{-1})\mathbf{x}_b + P_{cb}\Lambda_b^{-1}\mathbf{x}_b \quad (29)$$

where

$$B = \left[P_{ab}\Lambda_b^{-1}P_{ba} - \Lambda_a\right]^{-1} \neq 0 \quad (30)$$

(The existence of the B matrix is assumed, following Footnote 2.) The theorem thus follows directly from Eq. (29).

Theorem 3: A necessary and sufficient condition that the minimum variance estimate of $\mathbf{x}_{n+1}$, given the time-ordered sequence

$$\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$$

depend only on $\mathbf{x}_n$ is that $\mathbf{x}_i$, $\mathbf{x}_n$, and $\mathbf{x}_{n+1}$ be sequentially correlated for all $i \leq n$. The estimate in this case is

$$\hat{\mathbf{x}}_{n+1} = R_{n+1,n}\mathbf{x}_n \quad (31)$$

Proof: Let

$$\mathbf{x}_a^T = \left[\mathbf{x}_1^T, \mathbf{x}_2^T, \cdots, \mathbf{x}_{n-1}^T\right]$$

$$\mathbf{x}_b = \mathbf{x}_n$$

and

$$\mathbf{x}_c = \mathbf{x}_{n+1}$$

Then

$$P_{ca} = \left[\overline{\mathbf{x}_c \mathbf{x}_a^T}\right] = \left[P_{n+1,1}, P_{n+1,2}, \cdots, P_{n+1,n-1}\right] \quad (32)$$

$$P_{cb} = \left[\overline{\mathbf{x}_c \mathbf{x}_b^T}\right] = P_{n+1,n} \quad (33)$$

$$P_{ba} = \left[\overline{\mathbf{x}_b \mathbf{x}_a^T}\right] = \left[P_{n,1}, P_{n,2}, \cdots P_{n,n-1}\right] \quad (34)$$

$$\Lambda_b = \left[\overline{\mathbf{x}_n \mathbf{x}_n^T}\right] = \Lambda_n \quad (35)$$

The proof now follows from Theorem 2 and the definition of sequential correlation.

A sequentially correlated process is a generalized Markoff process, as the following corrollary points out:

Corollary: A Gaussian stochastic process is Markovian if and only if the correlation is sequential.

[3]From Ref. 3, p. 78.

Proof: A stochastic process is said to be Markovian if the conditional probability of a future state depends only on the present state but not on the past history of the process. The process is said to be Gaussian if the $\mathbf{x}_i$ are Gaussian. The minimum variance estimate completely determines the conditional distribution of the estimate if the random variables considered are chosen from Gaussian distributions, since the estimate then is Gaussian and the mean and variance are specified. This proves the corollary.

The difficulty remains that the $\mathbf{x}_i$, $i = 1, \cdots, n$, are usually not given; only certain observations and the estimates of the $\mathbf{x}_i$ that follow from them are available. Theorem 3 can be applied if the ϕ_i or the $\mathbf{x}_i^*$ are sequentially correlated, but a more useful result would place a condition on the parameters to be estimated. Theorem 4 is the fundamental theorem which yields this result for the case of observations that are linearly related to the parameters.

Theorem 4: (The fundamental theorem of sequential estimation.) Let the observations (ϕ^n) taken up to and including time t_n be linearly related to the $\{\mathbf{x}_1 \cdots \mathbf{x}_n\}$, i.e.,

$$\phi^n = A^n\mathbf{z}^n = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ & & \cdots\cdots & \\ 0 & 0 & \cdots & A_n \end{bmatrix}\mathbf{z}^n \quad (36)$$

where

$$\mathbf{z}^{nT} = \left[\mathbf{x}_1^T, \mathbf{x}_2^T, \cdots, \mathbf{x}_n^T\right]$$

and A^n is a matrix with nonrandom elements. Then the minimum variance estimate of $\mathbf{x}_{n+1}$, given the minimum variance estimates

$$\{\mathbf{x}_1^*, \mathbf{x}_2^*, \cdots, \mathbf{x}_n^*\}$$

is

$$\hat{\mathbf{x}}_{n+1} = R_{n+1,n}\mathbf{x}_n^* \quad (37)$$

if and only if $\mathbf{x}_{n+1}$, $\mathbf{x}_n$, and $\mathbf{x}_i$ are sequentially correlated for $i \leq n$. The covariance of the error in $\hat{\mathbf{x}}_{n+1}$ is

$$\left[\overline{\hat{\epsilon}_{n+1}\hat{\epsilon}_{n+1}^T}\right] = \Lambda_{n+1} - R_{n+1,n}\left\{\Lambda_n - \left[\overline{\epsilon_n \epsilon_n^T}\right]\right\}R_{n+1,n}^T \quad (38)$$

Proof: Let $\mathbf{x}_{n+1}$ be the estimate based on data up to, but not including, time t_{n+1}. Then

$$\begin{aligned}\hat{\mathbf{x}}_{n+1} &= \left[\overline{\mathbf{x}_{n+1}\phi^{nT}}\right]\left[\overline{\phi^n \phi^{nT}}\right]^{-1}\phi^n \\ &= \left[\overline{\mathbf{x}_{n+1}\mathbf{z}^{nT}}\right]\left[\overline{\mathbf{z}^n \mathbf{z}^{nT}}\right]^{-1}\left[\overline{\mathbf{z}^n \mathbf{z}^{nT}}\right]\left[A^{nT}\right]\left[\overline{\phi^n \phi^{nT}}\right]^{-1}\phi^n \\ &= \left[\overline{\mathbf{x}_{n+1}\mathbf{z}^{nT}}\right]\left[\overline{\mathbf{z}^n \mathbf{z}^{nT}}\right]^{-1}\mathbf{z}^{n*} \end{aligned} \quad (39)$$

But Eq. (39) is identical in form with the estimate based on the true $\mathbf{z}_n$; thus Eq. (37) follows directly from Theorems 2 and 3.

Let

$$\hat{\boldsymbol{\epsilon}}_{n+1} = \hat{\mathbf{x}}_{n+1} - \mathbf{x}_{n+1} \tag{40}$$

Then, from Theorem 1,

$$\left[\overline{\hat{\boldsymbol{\epsilon}}_{n+1}\hat{\boldsymbol{\epsilon}}_{n+1}^{T}}\right] = \Lambda_{n+1} - R_{n+1,n}\left[\overline{\mathbf{x}_n^{*}\mathbf{x}_n^{*T}}\right] R_{n+1,n}^{T} \tag{41}$$

But

$$\left[\overline{\mathbf{x}_n^{*}\mathbf{x}_n^{*T}}\right] = \left[\overline{\mathbf{x}_n\boldsymbol{\phi}^{nT}}\right]\left[\overline{\boldsymbol{\phi}^{n}\boldsymbol{\phi}^{nT}}\right]^{-1}\left[\overline{\boldsymbol{\phi}^{n}\mathbf{x}_n^{T}}\right] = \Lambda_n - \left[\overline{\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n^{T}}\right] \tag{42}$$

where

$$\boldsymbol{\epsilon}_n = \mathbf{x}_n^{*} - \mathbf{x}_n \tag{43}$$

Equation (38) follows from Eq. (41) and (42). This completes the theorem.

Theorem 4 is the foundation for the sequential estimation procedure described in the Summary, for now Theorem 1 can be applied to improve the estimate $\mathbf{x}_{n+1}$ by the incorporation of the new data gathered at time t_{n+1}. Note that the $\mathbf{x}_{n+1}^{*}$ so obtained is a minimum variance estimate for the data set $\{\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{n+1}\}$, but not for a data set that includes the later measurements $\{\boldsymbol{\phi}_{n+2}, \boldsymbol{\phi}_{n+3}, \cdots\}$. In most applications, however, it is sufficient to obtain a best estimate of only the most recent state of the system.

V. EXAMPLES OF SEQUENTIALLY CORRELATED PROCESSES

A sequentially correlated process can be defined by specifying the functional form of the normalized correlation matrices, or by describing the physical system which generates the stochastic vectors $\mathbf{x}(t_i)$. Examples are given below:

1. *The general functional form.* Let

$$R_{ji} = F(t_j)\,F^{-1}(t_i)$$

where F is an arbitrary non-singular matrix. It can be shown that R_{ji} always has this functional form for a continuous sequentially correlated process.

2. *Perfectly correlated variables.* Let $\mathbf{x}(t)$ be constant over any one experiment; then

$$R_{ji} = I$$

3. *Uncorrelated variables.* Let $\mathbf{x}_i$ and $\mathbf{x}_j$ be uncorrelated; then

$$R_{ji} = \text{the matrix delta function}$$

(Equation 1 holds for $t_i \leq t_j \leq t_k$.)

4. *Simple exponential correlation.* Let each component of $\mathbf{x}(t)$ be exponentially correlated, and let the cross-correlations be zero; then

$$R_{ji} = \begin{bmatrix} \exp\{\alpha(t_j - t_i)\} & 0 & \cdots & 0 \\ 0 & \exp\{\beta(t_j - t_i)\} & \cdots & 0 \\ & \cdots\cdots\cdots & & \\ 0 & 0 & \cdots & \exp\{\omega(t_j - t_i)\} \end{bmatrix}$$

5. *Orthogonal correlation.* Let $\mathbf{x}(t)$ be orthogonally correlated in two dimensions; then

$$R_{ji} = \begin{bmatrix} \cos(t_j - t_i) & -\sin(t_j - t_i) \\ \sin(t_j - t_i) & \cos(t_j - t_i) \end{bmatrix}$$

It can be shown that a necessary and sufficient condition that $R_{ki} = R_{kj}\, R_{ji}$ for all t_i, t_j, t_k (not just $t_i \leq t_j \leq t_k$) is that R_{ji} be orthogonal.

6. *The stationary process.* Let E be a constant matrix, and let

$$R_{ji} = \exp[(t_j - t_i)E] = I + (t_j - t_i)E + \cdots + \frac{(t_j - t_i)^n E^n}{n!} + \cdots$$

It can be shown that a necessary and sufficient condition that a continuous sequentially correlated process be stationary is that R_{ji} be of this form (Ref. 5).

7. *General exponential correlation.* Let the matrix $C(t)$ have the commutative property $C(t_j)\, C(t_i) = C(t_i)\, C(t_j)$, and take

$$R_{ji} = \exp[C(t_j) - C(t_i)]$$

(See Ref. 5 for a discussion of the matrix exponential.)

8. *Scintillation noise.* (See Fig. 2.) The scalar signal $s(t)$ is multiplied by the random gain $g(t)$ to obtain $\phi(t) = g(t)\, s(t)$, where $\overline{g(t)} = 1$, $\overline{s(t)} = 0$, and $\overline{[s(t_i)\, g(t_j)]} = 0$. Let $n(t) = g(t) - 1$, and

$$\mathbf{x}(t) = \begin{bmatrix} s(t) \\ s(t)n(t) \end{bmatrix}$$

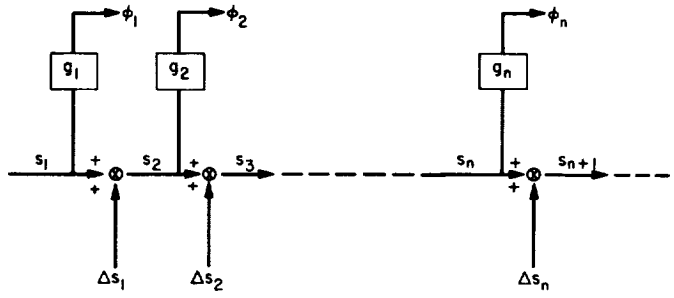

Fig. 2. The sequentially correlated scintillation noise

Then

$$\phi(t) = s(t) + s(t)n(t) = A\mathbf{x}(t)$$

where $A = [1, 1]$. Thus $\mathbf{x}(t)$ is a sequentially correlated process if both $s(t)$ and $n(t)$ are sequentially correlated.

9. *The linear control process.*[4] (See Fig. 3.) Let $\mathbf{x}(t)$ be derived from the process shown below, where

$$[\overline{\mathbf{n}_k \mathbf{x}_i^T}] = [\overline{\mathbf{n}_k \mathbf{x}_j^T}]\, [\overline{\mathbf{x}_j \mathbf{x}_j^T}]^{-1}\, [\overline{\mathbf{x}_j \mathbf{x}_i^T}]$$

It can be shown that this is the most general noise model which will yield a sequentially correlated linear control process.

[4]This model is discussed in Part VI.

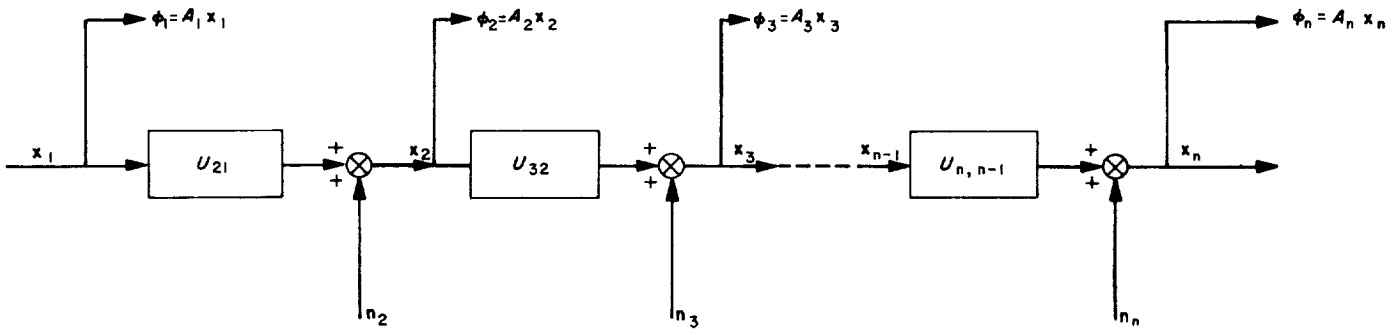

Fig. 3. The sequentially correlated linear system

VI. DISCUSSION OF KALMAN'S MODEL

The linear control process of Part V is a generalization of the model discussed by Kalman in Ref. 1, where he postulated

$$\left[\overline{\mathbf{n}_i\mathbf{n}_j^T}\right] = 0$$

for all i, j. It is easily verified that in this case Eq. (3) reduces to

$$\left[\overline{\hat{\boldsymbol{\epsilon}}_j\hat{\boldsymbol{\epsilon}}_j^T}\right] = U_{ji}\left[\overline{\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^T}\right]U_{ji}^T + \left[\overline{\mathbf{n}_j\mathbf{n}_j^T}\right]$$

Note that a sequentially correlated process sampled at the discrete times t_i can be put into Kalman's form by simply defining

$$U_{ji} = R_{ji}$$

and

$$\left[\overline{\mathbf{n}_j\mathbf{n}_j^T}\right] = \left[\Lambda_j - R_{ji}\Lambda_i R_{ji}^T\right]$$

This replaces the true physical process with an artificial one that has random discontinuities at the selected sample points t_i. The covariance matrix of the "noise" on the artificial process has the diagonal form

$$\left[\overline{\mathbf{n}\mathbf{n}^T}\right] = \begin{bmatrix} \left[\overline{\mathbf{n}_1\mathbf{n}_1^T}\right] & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \left[\overline{\mathbf{n}_2\mathbf{n}_2^T}\right] & \cdots & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \left[\overline{\mathbf{n}_i\mathbf{n}_i^T}\right] & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

$$\underbrace{\quad}_{t_1} \quad \underbrace{\quad}_{t_2} \qquad \underbrace{\quad}_{t_i}$$

With this construction in mind, the fundamental theorem of sequential estimation demonstrates the most general stochastic process that can be analyzed by Kalman's method. It provides a simple means for constructing his equivalent linear control process.

VII. CONCLUSION

An estimation procedure has been presented that is quite practical for obtaining numerical answers to many physical problems, and that provides a convenient analytic framework for exploiting theoretical problems in control theory. The only description of the system needed is the mean and covariance matrix of the time-varying vector $\mathbf{x}(t_i)$. The assumptions required are (1) the sequential correlation condition on the covariance matrix of the $\mathbf{x}_i$ and (2) the linear relationship between the observations ϕ_i and the state vector $\mathbf{x}_i$. Since the statistical description of a physical system is often not well known, it is many times possible simply to postulate sequential correlation and be assured of obtaining a reasonable estimate. The extension of the result to include nonlinear relationships between ϕ_i and $\mathbf{x}_i$ is not apparent. However, such a generalization might not be very useful since the validity of a linear and unbiased estimate would be called into question.

One of the limitations of this estimation technique is that it depends so strongly on the *a priori* correlations, which are often not well known. The estimation can be thought of as a weighting process, where each new data point is weighed against the previous (*a priori*) information. This can lead to an erroneous result if the *a priori* statistics are not correct.

NOMENCLATURE

T	superscript indicating the transpose of a matrix or vector
—	indicates the statistical average of a quantity over the ensemble of all experiments

Scalars

t	time
a, b, ···, n, ···	subscripts referring to times t_a, t_b, ···. The times are ordered according to the alphabetic sequence, thus $t_a < t_b < \cdots$.
x, y, z, ϕ	subscripts on a covariance matrix indicating to which random variables it applies
δ	denotes a variation in the indicated quantity
L	a special subscript referring to a non-optimum (not minimum variance) estimate
$\alpha, \beta, \gamma, \cdots$	the exponential decay constants (example 4)
s	a signal to be detected (example 8)
g	the gain multiplying s (example 8)

Vectors

$\mathbf{x}$	a column vector with random, time-varying components
$\mathbf{x}_i$	the value of $\mathbf{x}$ at the time t_i, thus $\mathbf{x}_i = \mathbf{x}(t_i)$
$\boldsymbol{\phi}_i$	an observation (data) taken at time t_i, thus $\boldsymbol{\phi}_i = \boldsymbol{\phi}(t_i)$
$\mathbf{x}_i^*$	the minimum variance estimate of $\mathbf{x}_i$ based on all observations up to and including time t_i
$\boldsymbol{\epsilon}_i$	the error in $\mathbf{x}_i^*$, thus $\boldsymbol{\epsilon}_i = \mathbf{x}_i^* - \mathbf{x}_i$
$\hat{\mathbf{x}}_i$	the minimum variance estimate of $\mathbf{x}_i$ based on all observations prior to, but not including, time t_i
$\hat{\boldsymbol{\epsilon}}_i$	the error in $\hat{\mathbf{x}}_i$, thus $\hat{\boldsymbol{\epsilon}}_i = \hat{\mathbf{x}}_i - \mathbf{x}_i$
$\mathbf{z}$	a dummy vector introduced to develop the proof of the Gauss-Markoff theorem
$\mathbf{z}^n$	a total random vector formed from all $\mathbf{x}_i$ up to and including time t_n, thus $\mathbf{z}^{nT} = \{\mathbf{x}_1^T, \mathbf{x}_2^T, \cdots, \mathbf{x}_n^T\}$
$\boldsymbol{\phi}^n$	a total observation vector formed from all $\boldsymbol{\phi}_i$ up to and including time t_n,thus $\boldsymbol{\phi}^{nT} = \{\boldsymbol{\phi}_1^T, \boldsymbol{\phi}_2^T, \cdots, \boldsymbol{\phi}_n^T\}$
$\mathbf{n}$	noise on an observation
$\mathbf{y}_i$	the coordinate deviation vector at time t_i, which is to be estimated by the orbit determination process

Matrices

I	the identity matrix
Λ_x	the covariance matrix describing the second moment of the distribution of the random vector $\mathbf{x}$
Λ_ϕ	the covariance matrix similarly defined for the random vector ϕ
Λ_i	the covariance matrix similarly defined for the random vector $\mathbf{x}_i$
Λ_0	the covariance matrix similarly defined for the unknown parameter vector $\mathbf{x}$ (Eq. 14, 15)

NOMENCLATURE (Cont'd)

P_{ji} the covariance matrix describing the linear correlation between the random vectors $\mathbf{x}_j$ and $\mathbf{x}_i$

R_{ji} the normalized correlation between $\mathbf{x}_j$ and $\mathbf{x}_i$, given by $R_{ji} = P_{ji}\Lambda_i^{-1}$

W_j the matrix which optimally weights each each new data point at the time t_j

A a matrix which relates observations to unknown parameters in the system (Eq. 13)

Γ the covariance matrix of noise on the observations (Eq. 15)

A_i a matrix which relates the observations ϕ_i to the random vectors $\mathbf{x}_i$

A^n a matrix formed from all A_i matrices up to and including time t_n (Eq. 36)

B a combination of matrices that appears in the proof of Theorem 2

C a matrix introduced in example 7

K, L dummy matrices introduced in the proof of the Gauss-Markoff theorem

E a constant matrix (example 6)

E^n indicates E multiplied by itself n times

U_{ji} the state transition matrix relating $\mathbf{x}_i$ to $\mathbf{x}_j$ in the linear control process (example 9)

F an arbitrary non-singular matrix introduced in example 1

REFERENCES

1. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *American Society of Mechanical Engineers Transactions*, Series D, Vol. 82, No. 1, March 1960, pp. 35-45.
2. Scheffe, H., *The Analysis of Variance*, John Wiley and Sons, New York, 1959.
3. Shapiro, I. I., *The Prediction of Ballistic Missile Trajectories from Radar Observations*, McGraw-Hill Book Company, New York, 1958.
4. Householder, A. S., *Principles of Numerical Analysis*, McGraw-Hill Book Company, New York, 1953.
5. Bellman, R., *Introduction to Matrix Analysis*, McGraw-Hill Book Company, New York, 1960.
6. Smith, G. L., Schmidt, S. F., and McGee, L. A., *Application of Statistical Filter Theory to the Optimal Estimation of Position and Velocity Onboard a Circumlunar Vehicle*, NASA Technical Report R-135, Washington D.C., 1962.

ACKNOWLEDGMENTS

The author wishes to thank the members of the Systems Analysis Section of the Jet Propulsion Laboratory who have aided and stimulated the development of this analysis. In particular, the author is indebted to C. Solloway, who provided valuable consultations and who developed Examples 1 and 5 of Part V. The motivation for this work was the treatment of correlated measurement errors in the orbit determination problem, as described in Part III. The iterative approach was suggested by the work of R. E. Kalman (Ref. 1), as applied by S. F. Schmidt (Ref. 6).