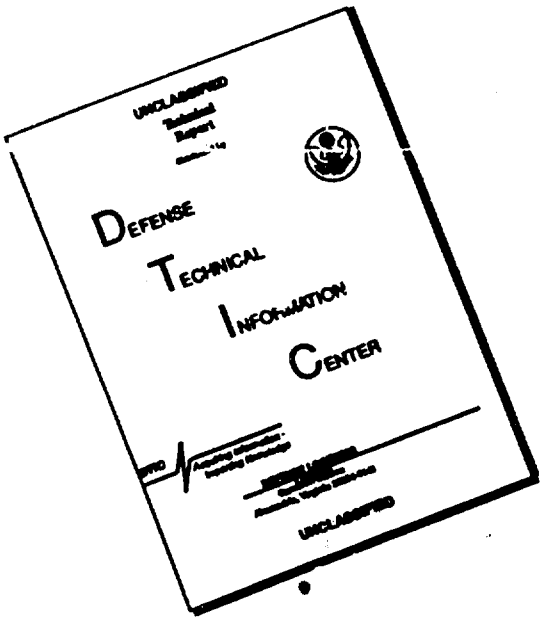
DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

AD736817

Air Force Report No.
SAMSO-TR-72-18

Aerospace Report No.
TR-0172(2311)-3

ESTIMATION OF THE COVARIANCE PARAMETERS OF NON-STATIONARY TIME-DISCRETE LINEAR SYSTEMS

Prepared by

Patrick L. Smith
Control Systems Department
Guidance and Control Systems Subdivision
Electronics Division

71 SEP 30

Engineering Science Operations
THE AEROSPACE CORPORATION

Prepared for

SPACE AND MISSILE SYSTEMS ORGANIZATION
AIR FORCE SYSTEMS COMMAND
LOS ANGELES AIR FORCE STATION
Los Angeles, California

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* The Aerospace Corporation El Segundo, California	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

ESTIMATION OF THE COVARIANCE PARAMETERS OF NON-STATIONARY TIME-DISCRETE LINEAR SYSTEMS

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*

Patrick L. Smith

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
71 SEP 30	27	8

8a. CONTRACT OR GRANT NO. F04701-71-C-0172 b. PROJECT NO.	9a. ORIGINATOR'S REPORT NUMBER(S) TR-0172(2311)-3
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* SAMSO-TR-72-18
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Space and Missile Systems Organization Air Force Systems Command Los Angeles Air Force Station

13. ABSTRACT

The Kalman filter sequentially generates the minimum variance estimate of the state of a linear dynamic system. This estimate is a function of the covariance parameters of the dynamic system model which implies that these be known a priori. Unfortunately some or all these covariance parameters are often unknown in engineering applications of the Kalman filter. In this report the maximum-likelihood estimates of the unknown covariance parameters of a time-discrete nonstationary linear system are computed from measurement residuals of a suboptimal sequential filter. Results for nonstationary linear systems are useful for nonlinear systems because most nonlinear estimation problems are solved by linearization which results in linear nonstationary plant and measurement models.

DD FORM 1473 (FACSIMILE)

UNCLASSIFIED
Security Classification

UNCLASSIFIED

Security Classification

14 KEY WORDS
Adaptive Filtering Kalman Filtering Maximum Likelihood
Distribution Statement (Continued)
Abstract (Continued)

UNCLASSIFIED

Security Classification

FOREWORD

This report is published by The Aerospace Corporation, El Segundo, California, under Air Force Contract No. F04701-71-C-0172. This report, which documents research carried out from September 1970 to September 1971, was submitted for review and approval on 9 November 1971 to Major Herbert A. Briesacher, SAMSO (SYGS).

Approved by

J. E. Farr, Director
Program 681D
Guidance and Control Systems
Subdivision

A. J. Schiewe, Director
Guidance and Control Systems
Subdivision
Electronics Division
Engineering Science Operations

Publication of this report does not constitute Air Force approval of the report's findings or conclusions. It is published only for the exchange and stimulation of ideas.

Herbert A. Briesacher, Major
SAMSO (SYGS)

ABSTRACT

The Kalman filter sequentially generates the minimum variance estimate of the state of a linear dynamic system. This estimate is a function of the covariance parameters of the dynamic system model which implies that these be known a priori. Unfortunately some or all these covariance parameters are often unknown in engineering applications of the Kalman filter. In this report the maximum-likelihood estimates of the unknown covariance parameters of a time-discrete nonstationary linear system are computed from measurement residuals of a suboptimal sequential filter. Results for nonstationary linear systems are useful for nonlinear systems because most nonlinear estimation problems are solved by linearization which results in linear nonstationary plant and measurement models.

CONTENTS

SECTION I

INTRODUCTION

The problem of estimating the states of dynamic systems with unknown covariance parameters has been studied by several investigators. Bucy and Follin published some significant results for linear stationary dynamic systems in 1962 [Ref. 1] not long after Kalman's paper [Ref. 2] appeared. Many navigation problems are nonstationary and the results for stationary systems do not apply. The reason is very basic: time-averages cannot be interchanged with ensemble-averages in nonstationary problems. This interchange is the crux of the methods used in stationary problems. Jazwinski [Ref. 3] and Abramson [Ref. 4] have studied nonstationary systems. Ramon Mehra published an important paper on identification of parameters in linear stationary systems [Ref. 5]. The author's results reported here are direct descendants of this work in that Mehra's formulation is generalized to nonstationary systems while the variables to be estimated are restricted to constant covariance parameters.

The method is as follows: A general class of suboptimal linear sequential filters is defined by letting the gain in the Kalman filter be an arbitrary (suboptimal) residual weighting matrix. Then the distribution of the measurement residuals generated by this filter is derived. The unknown covariance parameters are collected into a vector θ, and θ is estimated from the measurement residuals by the method of maximum likelihood. It follows immediately that the likelihood function (and hence $\hat{\theta}$) depends on the sample dispersion matrix of the measurement residuals.

The likelihood equations are too complex to solve for the estimate directly, so approximate solutions are outlined. In a special case, closed form, minimum-mean-square unbiased estimates can be derived. The example of a satellite attitude determination problem is discussed.

SECTION II

PLANT, MEASUREMENTS, AND FILTER

A set of n-dimensional vector-valued random variables $\{\mathbf{x}(i): i = 1, \ldots, N\}$ are generated by a linear stochastic difference equation called the plant

$$\mathbf{x}(i+1) = \phi(i)\,\mathbf{x}(i) + \Gamma(i)\mathbf{u}(i) \tag{1}$$

where $\{\mathbf{u}(i): i = 1, \ldots, N-1\}$ is a set of independent vector-valued random variables each of which is distributed $N_\ell(\mathbf{0}, Q)$.[†] The initial condition $\mathbf{x}(1)$ for Eq. (1) is distributed $N_n(\bar{\mathbf{x}}, M(1))$. The values of the random variables $\{\mathbf{x}(i): i = 1, \ldots, N\}$ for a particular realization are estimated sequentially from a series of m-dimensional vector-valued measurements $\{\mathbf{z}(i): i = 1, \ldots N\}$ which are related to the states by N linear relations

$$\mathbf{z}(i) = H(i)\mathbf{x}(i) + \mathbf{w}(i) \tag{2}$$

where $\{\mathbf{w}(i): i = 1, \ldots, N\}$ is a set of independent vector-valued random variables each of which is distributed $N_m(\mathbf{0}, R)$. The matrices M(1), Q, and R are positive semidefinite. In many engineering problems, elements of M(1), Q, and R are unknown. Assume the random variables $\{\mathbf{u}(i) = 1, \ldots, N \quad 1\}$ and $\{\mathbf{w}(i): i = 1, \ldots, N\}$ and $\mathbf{x}(1)$ are mutually independent. Equation (1) is often the result of linearizing a set of nonlinear equations about a nominal and so the state estimate covariance is a function of the nominal. A suboptimal filter for the states is

[†] The notation $N_\ell(\mathbf{0}, Q)$ means the vector-valued random variable $\mathbf{u}(i)$ is normally distributed with mean $\mathbf{0}$ and covariance Q. The dimension of $\mathbf{u}(i)$ is ℓ.

$$\tilde{\mathbf{x}}'(i+1) = \phi(i)\,\tilde{\mathbf{x}}'(i) + \phi(i)\,K(i)\,\mathbf{v}(i) \tag{3}$$

where

$$\mathbf{v}(i) = \mathbf{z}(i) - H(i)\,\tilde{\mathbf{x}}'(i) \qquad \tilde{\mathbf{x}}'(1) = \bar{\mathbf{x}} \tag{4}$$

and where the known sequence of gains $\{K(i):\ i = 1, \ldots, N\}$ has finite elements but is otherwise arbitrary. The estimate $\tilde{\mathbf{x}}'(i)$ is the (suboptimal) one-step extrapolated estimate of $\mathbf{x}(i)$ based on the measurements $\{\mathbf{z}(i):\ j = 1, \ldots, i-1\}$. To simplify the algebra, some notation is introduced. Rewrite Eq. (3) using $\{\mathbf{z}(i):\ i = 1, \ldots, N\}$ as the input.

$$\tilde{\mathbf{x}}'(i+1) = \phi(i)\left(I - K(i)\,H(i)\right)\tilde{\mathbf{x}}'(i) + \phi(i)\,K(i)\,\mathbf{z}(i) \tag{5}$$

The fundamental matrix of Eq. (5) is

$$\Phi(i+j,\ i) = \begin{cases} T(i+j-1)\cdots T(i) & ;\quad j>1 \\ T(i) & ;\quad j=1 \\ I & ;\quad j=0 \end{cases} \tag{6}$$

where

$$T(i) \triangleq \phi(i)\left(I - K(i)\,H(i)\right) \tag{7}$$

The vector-valued random variables $\tilde{\mathbf{x}}'(i)$ and $\mathbf{v}(i)$ are linear combinations of $\mathbf{u}(i)$ and $\mathbf{w}(i)$, and so it follows that $\tilde{\mathbf{x}}'(i)$ and $\mathbf{v}(i)$ are multidimensional gaussian random variables. The distributions of $\tilde{\mathbf{x}}'(i)$ and $\mathbf{v}(i)$ are derived in Ref. 6.

$$\mathbf{v}(i) \sim N_m\left(\mathbf{0},\ B(i)\right) \qquad B(i) = H(i)\,M(i)\,H^T(i) + R \tag{8}$$

$$C(i + j,\ i) \triangleq E\left[\mathbf{v}(i + j)\ \mathbf{v}^T(i)\right] \tag{9}$$

$$C(i + j,\ i) = H(i + j)\ \Phi(i + j,\ i + 1)\ \phi(i)\left(M(i)\ H^T(i) - K(i)\ B(i)\right) \tag{10}$$

$$\tilde{\mathbf{x}}'(i) \sim N_n\left(\mathbf{x}(i),\ M(i)\right)$$

$$M(i + 1) = T(i)\ M(i)\ T^T(i) + \phi(i)\ K(i)\ RK^T(i)\ \phi^T(i) + \Gamma(i)\ Q\Gamma^T(i) \tag{11}$$

The initial condition for Eq. (11) is M(1).

Collect the unknown elements of M(1), Q, and R into an r-dimensional parameter vector θ and denote its true value by θ_o. The relationship between θ and the residuals $\{\mathbf{v}(i):\ i = 1,\ \ldots,\ N\}$ is established in the following lemma.

Lemma 2.1

If: (I1) $\{K(i):\ i = 1,\ \ldots,\ N\}$ are given.

Then: (R1) The elements of C(i + j, i) and B(i) are affine functions of the elements of θ.

Proof: The closed form for the covariance equation [Eq. (11)] is

$$M(i) = \Phi(i,\ 1)\ M(1)\ \Phi^T(i,\ 1)$$

$$+ \sum_{j=1}^{i-1} \Phi(i + 1,\ j + 1)\ \phi(j)\ K(j)\ RK^T(j)\ \phi^T(j)\ \Phi^T(i + 1,\ j + 1)$$

$$+ \sum_{j=1}^{i-1} \Phi(i + 1,\ j + 1)\ \Gamma(j)\ Q\Gamma^T(j)\ \Phi^T(i + 1,\ j + 1) \tag{12}$$

Substitute Eq. (12) into Eq. (8). It follows that the elements of B(i) are affine functions of the elements of M(1), R, and Q and hence of θ. The proof for C(i + j, i) follows in the same way.

QED

The significance of this lemma is that the partials $\frac{\partial B(i)}{\partial \theta_n}$ and $\frac{\partial C(i+j,\ i)}{\partial \theta_n}$ are independent of θ.

The Kalman filter is of course an important special case in the general class of linear sequential filters. Two important statistical properties of the Kalman filter are summarized in the following lemma.

Lemma 2.2

If: (I1) R is positive definite.

(I2) $K(i) = M(i)\, H^T(i)\, B^{-1}(i)$ (13)

Then: (R1) Tr [M(i)] is minimal. If M(i) is positive definite, |M(i)| is minimal.

(R2) $C(i + j,\ i) = 0$ for all j. (14)

Proof: See Ref. 6.

If R is singular then the pseudo inverse is used in I2 if B(i) is singular. R2 is not a sufficient condition for R1 [Ref. 6.]

SECTION III

ESTIMATION OF COVARIANCE PARAMETERS

The measurement residuals from a suboptimal filter are not statistically independent and the usual methods in parameter estimation must be extended to dependent samples in order to estimate θ.

Define a composite vector V(N) of the measurement residuals $\{v(i): i = 1, 2 \ldots, N\}$ generated by the suboptimal sequential filter. V(N) is a p-dimensional random variable distributed $N_p(0, \Sigma(N))$, where $p \triangleq Nm$ and where $\Sigma(N)$ is the following composite matrix:

$$\Sigma(N) = \begin{bmatrix} B(1) & C(1,2) & C(1,3) & \ldots. & C(1,N) \\ C(2,1) & B(2) & C(2,3) & \ldots. & C(2,N) \\ C(3,1) & C(3,2) & B(3) & \ldots. & C(3,N) \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ C(N,1) & C(N,2) & C(N,3) & \ldots. & B(N) \end{bmatrix} \tag{15}$$

If $\Sigma(N)$ is positive definite, the probability density of V(N) is

$$p(V(N)|\theta) = (\sqrt{2\pi})^{-p} |\Sigma(N)|^{-1/2} \exp\left\{-\frac{1}{2} V^T(N) \Sigma^{-1}(N) V(N)\right\} \tag{16}$$

The maximum likelihood estimate maximizes $p(V(N)|\theta)$ (or equivalently $\log p(V(N)|\theta)$) for the observed value of V(N).

$$\max_{\theta} \log p(V(N)|\theta) = \log p(V(N)|\hat{\theta}) \tag{17}$$

Define

$$J_N(\theta) \triangleq \log p(V(N)|\theta) \tag{18}$$

$$J_N(\theta) = -\frac{1}{2} \log |\Sigma(N)| - \frac{1}{2} \mathrm{Tr} \left[\Sigma^{-1}(N)\, V(N)\, V^T(N)\right] \tag{19}$$

A necessary condition for $\hat{\theta}$ to be a relative maximum of $J_N(\theta)$ is

$$\nabla_{\theta} J_N(\hat{\theta}) = 0 \tag{20}$$

Define the score (a classical term) of θ

$$S_N(\theta) \triangleq \nabla_{\theta} J_N(\theta) = \begin{bmatrix} \frac{\partial J_N(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial J_N(\theta)}{\partial \theta_r} \end{bmatrix} \tag{21}$$

and the following notation for the partial derivatives of $\Sigma(N)$,

$$W(N;i) \triangleq \frac{\partial \Sigma(N)}{\partial \theta_i} \tag{22}$$

By Lemma 2.1, W(N;i) is independent of θ for all i. Hence it follows that

$$\frac{\partial J_N(\hat{\theta})}{\partial \theta_j} = -\frac{1}{2}\ \mathrm{Tr}\left[\Sigma^{-1}(N)\ W(N;j)\left(I - \Sigma^{-1}(N)\ V(N)\ V^T(N)\right)\right] = 0 \qquad (23)$$
$$j = i, \ldots, r$$

In general, Eq. (23) must be solved numerically. The average score $A_N(\theta)$ is defined by

$$A_N(\theta) \triangleq E\left[S_N(\theta)\right]$$

The variance of the score is Fisher's information matrix $\mathcal{J}_N(\theta)$ and is given by

$$\mathcal{J}_N(\theta) \triangleq E\left[S_N(\theta)\ S_N^T(\theta)\right]$$

The expectations which define $A_N(\theta)$ and $\mathcal{J}_N(\theta)$ are taken over the probability density function $p(V(N)|\theta_0)$.

The Cramer-Rao inequality is a lower bound on the mean square error of an arbitrary estimator $\tilde{\theta}$ [Ref. 7]. If $\tilde{\theta}$ is an unbiased estimator of θ, then

$$\mathrm{cov}\left[\hat{\theta}\right] \geq \mathcal{J}_N^{-1}(\theta_0) \qquad (24)$$

If the equality in Eq. (24) holds for a particular unbiased estimate $\hat{\theta}$, then $\hat{\theta}$ is the minimum-mean-square unbiased estimate of θ_0.

The information matrix is a measure of the sensitivity of the distribution of the residuals with respect to variations in the unknown parameter θ. At one extreme, if there is a one-to-one correspondence between each sequence of residuals and each value of θ, then the sensitivity or information is a maximum. At the other extreme, if the distribution of the residuals is

independent of θ, then the information about θ is a minimum. An expression for the information contained in the residuals is derived next.

Lemma 3.1

If: (I1) $\Sigma(N)$ is a positive definite for $\theta = \theta_0$.

Then: (R1) $A_N(\theta_0) = 0$.

(R2) The (i, j) element of $\mathcal{I}_N(\theta_0)$ equals

$$- E\left[\frac{\partial^2 J_N(\theta)}{\partial\theta_i\,\partial\theta_j}\right] \text{ evaluated at } \theta = \theta_0.$$

$$\text{(R3)}\quad E\left[\frac{\partial^2 J_N(\theta)}{\partial\theta_i\,\partial\theta_j}\right] = -\frac{1}{2}\,\mathrm{Tr}\left[\Sigma^{-1}(N)\,W(n;j)\,\Sigma^{-1}(N)\,W(N;i)\right] \qquad (25)$$

Proof: R1 and R2 are standard results which follow immediately from the definition of expectation [Ref. 7]. R3 follows from Eq. (19).

As an example, suppose there is just a single scalar covariance parameter, θ. From Lemma 2.1, it follows that $\Sigma(N)$ can be written

$$\Sigma(N) = W(N)\,\theta \qquad (26)$$

A closed form maximum likelihood estimate of θ can be obtained in this particular case. Drop the index N for clarity. From Eq. (23)

$$S(\hat{\theta}) = -\frac{1}{2}\,\mathrm{Tr}\left[\Sigma^{-1}\,W(I - \Sigma^{-1}\,VV^T)\right] = 0 \qquad (27)$$

But since $\Sigma = W\hat{\theta}$ where $\hat{\theta} > 0$, it follows that

$$S(\hat{\theta}) = -\frac{1}{2\hat{\theta}} \operatorname{Tr}\left[I - \frac{1}{\hat{\theta}} W^{-1} VV^T\right] = 0 \tag{28}$$

$$\hat{\theta} = \frac{1}{N} V^T W^{-1} V \tag{29}$$

The following lemma is an application of the Cramer-Rao inequality which shows that the maximum likelihood estimate in this case is also the minimum-mean-square unbiased estimator for θ.

Lemma 3.2

If: (I1) $\theta > 0$

(I2) W is positive definite.

Then: (R1) $\hat{\theta} = \frac{1}{N} V^T W^{-1} V$ (30)

is the minimum-mean-square unbiased estimate of θ.

Proof: (P1) From Eq. (29),

$$E[\hat{\theta}] = \frac{1}{N} \operatorname{Tr}\left[W^{-1} W\theta_0\right] = \theta_0 \tag{31}$$

(P2) $$\operatorname{cov}[\hat{\theta}] = \operatorname{var}[\hat{\theta}] - \theta_0^2 \tag{32}$$

$$E\left[\hat{\theta}^2\right] = \frac{1}{N^2} E\left[(V^T W^{-1} V)^2\right] \tag{33}$$

$$E[\hat{\theta}^2] = \frac{1}{N^2}\left\{2\ \mathrm{Tr}\left[(W^{-1}\Sigma)^2\right] + \left(\mathrm{Tr}\left[(W^{-1}\Sigma)\right]\right)^2\right\} \tag{34}$$

$$E[\hat{\theta}^2] = \frac{1}{N^2}\left\{2N\theta^2 + N^2\theta^2\right\} \tag{35}$$

See Ref. 6 for the details of this last step.

$$\mathrm{cov}[\hat{\theta}^2] = \frac{2}{N}\,\theta^2 \tag{36}$$

(P3) Fisher's Information is

$$\mathcal{J}(\theta) = \frac{1}{2}\ \mathrm{Tr}\left[(\Sigma^{-1}W)^2\right] = \frac{N}{2}\cdot\frac{1}{\theta^2} \tag{37}$$

R1 follows from the Cramer-Rao inequality.

QED

In general, of course, the maximum likelihood estimate does not have any optimal properties for finite N.

SECTION IV

APPROXIMATE MAXIMUM LIKELIHOOD ESTIMATES

In general, closed form solutions of Eq. (23) do not exist, hence Eq. (23) must be solved numerically for $\hat{\theta}$. For large p, storing and inverting $\Sigma(N)$ is impractical. A feasible approximate solution to Eq. (23) can be computed if the elements of K(i) are reasonably near the elements of $M(i)\, H^T(i)\, B^{-1}(i)$ in value and the eigenvalues of $\Phi(i + j, i)$ decrease to zero for increasing j for all i. These conditions can usually be fulfilled in practice. $\Sigma(N)$ may then be approximated by a band matrix. For example, define the following band matrix by replacing $C(i + j, i)$ with zeros for $j = 3, \ldots, N - i$ for $i = 1, \ldots, N$.

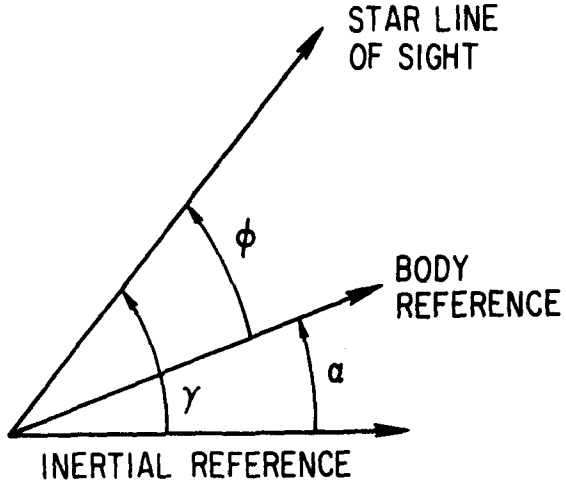
$$\Sigma_2(N) = \begin{bmatrix} B(1) & C(1,2) & 0 & \cdots & 0 \\ C(2,1) & B(2) & C(2,3) & \cdots & 0 \\ 0 & C(3,2) & B(3) & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & B(N) \end{bmatrix} \tag{38}$$

Band matrices retaining higher order lag correlations can be defined in the same way. Special algorithms can efficiently compute the vector $\Sigma_b^{-1}(N)\, V(N)$ and the matrix $\Sigma_b^{-1}(N)\, W(N;i)$ where $\Sigma_b(N)$ is a band matrix. To reduce the computer storage requirements, W(N;i) may be approximated by a band matrix as well. An iterative technique for solving Eq. (23) is described in Ref. 7.

SECTION V

EXAMPLE

A single axis analog of a satellite in a planar orbit is examined (see the sketch below). The problem is to estimate the attitude of the satellite with respect to an inertial reference line.



A star mapper attached to the body measures ϕ at time t_i.

$$\phi_m(t_i) = \phi(t_i) + w(i) = \gamma(t_i) - \alpha(t_i) + w(i) \tag{39}$$

w(i) is the measurement error and Y is determined from star charts. A gyro continuously monitors the angular rate ω

$$\omega_m = \omega + n \tag{40}$$

where n is the random drift rate. The attitude error is e, where

$$\alpha = \alpha_o + e \qquad \dot{\alpha}_o = \omega_m \tag{41}$$

$$\dot{e} = \dot{\alpha} - \dot{\alpha}_o = \omega - \omega_m = n \tag{42}$$

Integrate $\dot{e}$ between star sightings.

$$e(t_{i+1}) = e(t_i) + u(i) \tag{43}$$

$$u'(i) = \int_{t_i}^{t_{i+1}} n(t)\, dt \tag{44}$$

The angle u'(i) is the drift error of the gyro between the star sighting time and is accurately modeled by a gaussian random variable with the following statistics:

$$E[u'(i)] = 0 \tag{45}$$

$$E[u'(i)\, u'(j)] = (t_{i+1} - t_i)\, q\, \delta_{ij} \tag{46}$$

Subtract the nominal attitude predicted from the angle measured by the star mapper.

$$z(i) \triangleq \phi_m(t_i) - \alpha_o(t_i) - \gamma(t_i) = e(t_i) + w(i) \tag{47}$$

Assume that w(i) is a gaussian random variable with the following statistics:

$$E\,[w(i)] = 0 \tag{48}$$

$$E\,[w(i)\ w(j)] = r\ \delta_{ij} \tag{49}$$

Define the following variables:

$$x(i) \triangleq e(t_i) \tag{50}$$

$$\Gamma(i) \triangleq \sqrt{t_{i+1} - t_i} \tag{51}$$

$$u(i) \triangleq \frac{u'(i)}{\Gamma(i)} \tag{52}$$

Hence Eqs. (43) and (47) become

$$x(i+1) = x(i) + \Gamma(i)\ u(i) \tag{53}$$

$$z(i) = x(i) + w(i) \tag{54}$$

$$E\,[x(1)] = 0 \qquad E\,[u(i)] = 0 \tag{55}$$

$$E\left[x^2(1)\right] = m(1) \qquad E\,[u(i+j)\ u(i)] = q \tag{56}$$

A constant gain suboptimal filter [Eq. (3)] for x(i) is

$$\tilde{x}'(i+1) = \tilde{x}'(i) + k\left(z(i) - \tilde{x}'(i)\right) \tag{57}$$

The sequential filter given by Eq. (57) can be implemented on-board the satellite and the estimates and star sighting times relayed back to earth. It is desired to estimate q and r from these data in order to estimate m(i) (which is an affine function of q and r) and to detect a failure in the gyro or star mapper [Ref. 8]. After substituting, one finds

$$m(i+1) = \bar{k}^2\, m(i) + k^2\, r + \Gamma^2(i)\, q \tag{53}$$

$$m(i) \triangleq \operatorname{cov}[\tilde{x}'(i)] \qquad \bar{k} \triangleq 1 - k \tag{59}$$

$$c(i+j,\, i) = \bar{k}^j\, m(i) - \bar{k}^{j-1}\, kr \tag{60}$$

Define the unknown parameter vector.

$$\theta = \begin{pmatrix} q \\ r \end{pmatrix} \tag{61}$$

Denote $\sigma(i,j)$ as the i, j element of $\Sigma(N)$ and $w_\ell(i,j)$ as the i, j element of $W(N;\ell)$, $\ell = 1, 2$. One finds from Eqs. (60) and (58) the following expressions:

$$w_\ell(i+j,\, i) = \begin{cases} \bar{k}^j\, \dfrac{\partial m(i)}{\partial \theta_\ell} + \bar{k}^{j-1} k\, \delta_{\ell 2}; & j = 1, \ldots, N - i \\ \dfrac{\partial m(i)}{\partial \theta_\ell} + \delta_{\ell 2} \quad ; & j = 0 \end{cases} \tag{62}$$

where

$$\frac{\partial m(i+1)}{\partial \theta_\ell} = \bar{k}^2 \frac{\partial m(i)}{\partial \theta_\ell} + k^2 \delta_{\ell 2} + \Gamma^2(i) \delta_{\ell 1} \tag{63}$$

$$\frac{\partial m(1)}{\partial \theta_\ell} = 0 \tag{64}$$

N may range up to 25 sightings for a single orbit.

The degree of correlation $\{v(i): i = 1, \ldots, N\}$ depends on k. For insight, it is convenient to let $c(i) = 1$ for the present, so that the sequence $\{v(i): i = 1, \ldots, N\}$ is asymptotically stationary. The correlation coefficients of $\{v(i): i = 1, \ldots, N\}$ are computed from Eq. (58).

$$m = \bar{k}^2 m + k^2 r + q \qquad m = \lim_{i \to \infty} m(i) \tag{65}$$

$$m = \frac{k^2 r + q}{1 - \bar{k}^2} \tag{66}$$

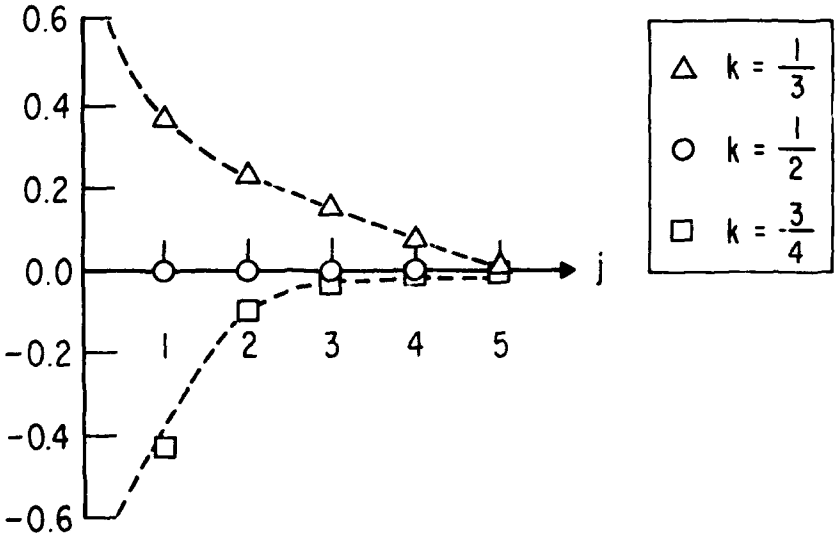
$$c(j) = \bar{k}^j m - \bar{k}^{j-1} kr \qquad c(j) = \lim_{i \to \infty} c(i + j, i) \tag{67}$$

$$\rho(j) = \frac{c(j)}{m + r} \tag{68}$$

As a numerical example, let $q = 45$ arc sec^2 and $r = 90$ arc sec^2. Then, one finds from Eq. (66), for various values of k,

$$
\begin{aligned}
k &= 1/3 \qquad m = 99 \text{ arc sec}^2 \\
k &= 1/2 \qquad m = 90 \text{ arc sec}^2 \\
k &= 3/4 \qquad m = 102 \text{ arc sec}^2
\end{aligned}
\tag{69}
$$

Notice that m is relatively insensitive to variations in k about the optimal. This is true for the nonstationary case also. The correlogram for each k is graphed below.



When k = 1/3 (less than optimal), the filter is underdriven and successive measurement residuals are positively correlated. When k = 1/2 (optimal), the residuals are uncorrelated [Lemma 3.2]. When k = 3/4 (greater than optimal), the filter is overdriven and successive measurement

residuals are negatively correlated. In both suboptimal filters the correlation is essentially zero between v(i + j) and v(i) when j is greater than 5. Intuitively this means that there is very little information about r and q in c(i + j, i) when j is greater than 5. Hence if c(i) is near 1, on the average, one expects the band matrix approximation to be very good.

The speed and accuracy of the band matrix approximation for this example described in Section IV were examined. The irregular star sighting times were simulated by

$$\Gamma^2(i) = 1 - \frac{1}{2} \sin\left(\frac{\pi}{3} i\right) \tag{70}$$

For simplicity, the following initial conditions were used for all cases:

$$\hat{x}'(1) = x(1) = 0 \qquad m(1) = 90 \text{ arc sec}^2 \tag{71}$$

The random numbers u(i) and w(i) were precomputed and adjusted to have zero sample means and sample variances equal to q and r respectively.

For the case where k = 1/3, the following relative times were required to compute the estimate for various band sizes:

$$\begin{aligned} t &= 0.24 \qquad b = 1 \\ t &= 0.44 \qquad b = 3 \\ t &= 1.00 \qquad b = 24 \end{aligned} \tag{72}$$

The relative difference between the time required for inverting a full matrix and its banded approximation increases as N increases.

The following tabulation is a list of typical estimates of r (= 90 arc sec^2) and q (= 45 arc sec^2) for a single run for various values of b and k.

Parameters	Estimates for k = 1/3			Estimates for k = 3/4		
	b = 1	b = 3	b = 24	b = 1	b = 3	b = 24
q	10.91	12.50	12.50	See Text	14.19	15.76
r	149.08	142.36	142.84	158.09	141.96	134.59

The same sequence of random numbers was used to generate the filter residuals, so the differences in the estimates are due to the different band approximations and the filter gains k. For b = 24 (the last columns), the estimate is the exact maximum likelihood estimate. Notice that the approximate estimate for b = 3 differs only slightly from the exact estimate. The q estimate computed for b = 1 and k = 3/4 was negative. Clearly a better estimate in this case is zero since q is non-negative.

The Cramer-Rao bound [Eq. (24)] was computed for k = 1/3.

$$\operatorname{cov}[\hat{\theta}] \geq \begin{pmatrix} 1608 & -668 \\ -668 & 1251 \end{pmatrix}$$

Since the elements of θ for this example are always positive, a non-negative estimate (which is biased) would have a mean-square error which is less than that of an efficient unbiased estimate. This, in fact, appears to be the case for this example. Ten runs were made and the sample bias and mean-square error were computed for k = 1/3.

$$\frac{1}{10}\sum_{i=1}^{10}(\hat{\theta}_i - \theta) = \begin{pmatrix} 16.1 \\ -13.4 \end{pmatrix}$$

$$\frac{1}{10}\sum_{i=1}^{10}(\hat{\theta}_i - \theta)(\hat{\theta}_i - \theta)^T = \begin{pmatrix} 1216 & -662 \\ -662 & 811 \end{pmatrix}$$

The sample mean-square error is lower than the Cramer-Rao bound. Again, this can occur because the maximum likelihood estimates are biased.

SECTION VI

CONCLUSION AND APPLICATIONS

The maximum likelihood estimates of the covariance parameters of a linear time-discrete system have been derived. Approximations are required, in general, to make their computation tractable.

One application of this work is "adaptive" filtering where the filter gain is made a function of the estimated covariance parameters. References 6 and 8 contain examples of adaptive filters. Another application is off-line filtering, such as post-flight data reduction. The increased computation required to estimate the unknown covariance parameters is not prohibitive in off-line applications.

REFERENCES

1. R. S. Bucy and J. Follin, "Finite Time Adaptive Filtering," IRE Trans. on Automatic Control, 7, 10-19 (July 1962).

2. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME J. Basic Eng., 82, 34-45 (March 1960).

3. A. H. Jazwinski and A. E. Bailie, Adaptive Filtering, Analytical Mechanics Associates, Inc., Report No. 67-6 (March 1967).

4. Paul D. Abramson, Jr., Simultaneous Estimation of the State and Noise Statistics in Linear Dynamic Systems, NASA Technical Report, NASA TR R-332 (March 1970).

5. R. K. Mehra, "Identification of Linear Dynamic Systems," paper presented at the IEEE Symposium on Adaptive Processes (8th), (17-19 November 1969).

6. P. L. Smith, The Estimation of Covariance Parameters of Time-Discrete Linear Systems with Applications to Adaptive Filters, The Aerospace Corporation, TOR-0059(6311)-23 (31 May 1971).

7. M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, Hafner, New York, Vol. 2 (1961).

8. P. L. Smith, Adaptive Filtering Applied to an Attitude Estimation Problem, The Aerospace Corporation, TOR-0059(6311)-14 (12 October 1970).