MEMORANDUM
RM-5859-PR
DECEMBER 1968

AD 680765

MAXIMUM-LIKELIHOOD PREDICTION AND ESTIMATION FOR NONLINEAR DYNAMIC SYSTEMS

L. D. Attaway

DDC
JAN 8 1969
B

PREPARED FOR:

UNITED STATES AIR FORCE PROJECT RAND

The RAND Corporation
SANTA MONICA • CALIFORNIA

99

MEMORANDUM
RM-5859-PR
DECEMBER 1968

MAXIMUM-LIKELIHOOD PREDICTION AND ESTIMATION FOR NONLINEAR DYNAMIC SYSTEMS

L. D. Attaway

This research is supported by the United States Air Force under Project RAND—Contract No. F44620-67-C-0045—monitored by the Directorate of Operational Requirements and Development Plans, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this study should not be interpreted as representing the official opinion or policy of the United States Air Force.

DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited.

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA • 90406

This study is presented as a competent treatment of the subject, worthy of publication. The Rand Corporation vouches for the quality of the research, without necessarily endorsing the opinions and conclusions of the authors.

Published by The RAND Corporation

PREFACE

A great deal of progress has been made in the development and application of the theory of prediction and estimation in noise for the case of linear problems. A major research frontier today concerns the problems of estimation using noise-corrupted observations in nonlinear systems. In this Memorandum a theoretical approach to maximum-likelihood prediction and estimation is developed for certain such nonlinear situations and is applied to the special case of numerically estimating the initial conditions of a radar-observed reentry body.

The work reported here is part of Rand's continuing basic research in engineering and systems science.

SUMMARY

It is often desirable to estimate certain parameters, such as initial conditions, of a dynamic physical process, using noise-corrupted observations. This Memorandum describes a maximum-likelihood solution when the process is time-invariant nonlinear m-vector and the observation is a linear combination of true process coordinates and noise. It is applied to a white-gaussian-noise case in which the initial conditions have a known a priori joint gaussian distribution and all other process parameters are known. It is desired to estimate the state at $t \geq 0$, given continuous observation over $(0,T)$ and a priori statistics.

Given a maximum-likelihood estimate of initial conditions, a natural estimate for the system state at $t > 0$ simply updates the process differential equation from the estimated initial conditions. The algorithm therefore estimates initial conditions. It can be shown that the maximum-likelihood estimate is th initial condition which minimizes a certain functional on that initial condition, the observation, and the a priori statistics. This functional describes an $(m + 1)$ surface at time T, and the desired estimate corresponds to its minimum; a differential equation is developed which governs the evolution of this estimate with time T.

Using this differential equation, the algorithm calculates as a function of T that maximum-likelihood solution which evolves from the unique solution at time $T = 0$ (given by the a priori mean vector). Using this differential-equation algorithm to stay near the desired solution when updating T, Newtonian techniques may be used to improve the solution for constant T.

Earlier computational results suggest that maximum likelihood is preferred to various approximations. This Memorandum explores computing feasibility for difficult vector cases, rather than extending these numerical comparisons for simpler scalar situations. Both pure differential equation and differential-equation/Newtonian solutions were used to estimate reentry-vehicle initial conditions. A priori statistics presumably result from prior less accurate tracking; the observation from a more accurate tracking system. One coordinate time rate is not measured, and its initial condition serves as an unknown parameter with a priori statistics.

A fixed reentry-path observation was generated using Monte Carlo noise, and initial conditions were estimated. Convergence depended upon the size of integration constants used. The estimate converged to within the order of the Cramer-Rao conditional bound. Accuracy was exchanged between coordinates to improve the overall estimate. The initial angle rate, which could be viewed as an unknown parameter, was handled with the same or better rapidity and accuracy of convergence than the other initial conditions. Readily available programs and the use of a general-purpose computer resulted in processing times much larger than real-time observation. However, efficient programs and special computers could reduce this to on-line realization.

ACKNOWLEDGMENTS

This research was performed in partial fulfillment of the requirements for the degree Doctor of Philosophy in Engineering at the University of California at Los Angeles. It is a pleasure to acknowledge the guidance given me by Professor A. V. Balakrishnan, my advisor.

The computations were performed at Rand, where Jo Ann Lockett gave invaluable aid in the programming and computer runs.

Marjorie Dobson was infinitely patient and accurate in typing the draft and preparing the final manuscripts.

My deepest thanks to them all.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS vii

SYMBOLS xi

Section

I. PROBLEM DEFINITION 1
- General Problem Statement 1
- Special Nonlinear Case Treated 2
- History of the Problem 4

II. NUMERICAL TECHNIQUE 9
- Conditional Expectation 9
- Maximum-Likelihood Estimate 9

III. ERROR STATISTICS 21

IV. AN APPLICATION: REENTRY PREDICTION 27
- General 27
- Equations of Motion 28
- Radar Observables 35
- Computational Approach 36
- Computational Results 46

V. CONCLUSIONS 67
- Theoretical Results 67
- Numerical Results 68
- Future Developments 69

Appendix

A. Derivation of Conditional Probability $p(C|Z(s), 0 \leq s \leq T)$ 71

B. Program Flow Diagram 83

REFERENCES 87

SYMBOLS

A_C = reentry-vehicle frontal (drag) area

$A(\hat{X}), G(\hat{X}), J(\hat{X})$ = functionals on $H(s)$, $X(s;\hat{X})$, R, and $Z(s)$, $0 \le s \le T$

$\bar{a}_V$ = unit vector along the reentry-vehicle line of flight

$B(X_0,T)$ = matrix function of X_0 and T

C_D = reentry-vehicle drag coefficient

$\bar{D}$ = drag-force vector acting on reentry vehicle

$\mathrm{Det}[\cdot]$ = determinant

$E(\cdot)$ = expectation

$E(\cdot|Y)$ = conditional expectation, given Y

$\bar{F}$ = total-force vector acting on reentry vehicle

$$F[X;t;\alpha] = F[X,\alpha] = F[X] = \frac{dX(t)}{dt} = \begin{pmatrix} f_1(X;t;\alpha) \\ \vdots \\ f_m(X;t;\alpha) \end{pmatrix}$$

$f(\hat{X},T)$ = functional on u, Λ, and $Z(s)$, $0 \le s \le T$

G = factor in error statistic

$\bar{G}$ = gravity-force vector acting on reentry vehicle

g = gravitational acceleration

$g^k(s,\hat{X})$ = function related to $r_k(s,u)$, $H_R(u)$, and $X(u;\hat{X})$

$H_i(s)$ = i^{th} row of $H(s)$

$\begin{pmatrix} H_1(s) \\ \vdots \\ H_p(s) \end{pmatrix}$ = matrix which defines the effect of each coordinate on the observation

I = identity matrix

K = normalizing constant

$L(s)$ = vector Wiener process

$M(i;\hat{X})$ = conditional expectation of $Z(i)$, given $\hat{X}$

m = reentry-vehicle mass

$m_i^k(\hat{X})$ = expectation of random variable related to $H_k(s)$, $X(s;\hat{X})$

$N = T/\Delta T$ = number of updatings in T

$N(s)$ = noise in observation at s, $0 \le s \le T$

n = number of Newtonian approximations

$p(\cdot)$ = probability

$R(i)$ = covariance matrix for $Z(i)$

$$R(\xi,s) = \begin{pmatrix} r_1(\xi,s) & & 0 \\ & \ddots & \\ 0 & & r_p(\xi,s) \end{pmatrix} = \text{covariance of the noise}$$

$$R(\xi,s) = \begin{pmatrix} \phi_1 & & 0 \\ & \ddots & \\ 0 & & \phi_p \end{pmatrix} \delta(\xi - s) = \text{covariance of white noise}$$

r = radial distance to reentry vehicle

$\bar{r}_1$ = unit radial vector

T = upper index limit of observation interval

t,s,u = process index, $0 \le s, u \le T$, $0 \le t$

$\bar{V}$ = reentry-vehicle velocity

W = weight of reentry vehicle

X^* = transpose of matrix X

$X(0) = X_0$ = true initial condition of $X(t)$ at $t = 0$

$X(s;\hat{X})$ = $X(s)$, given that $X(0) = \hat{X}$

$\hat{X}(T) = \hat{X}(0;T) = \hat{X}_0$

$$X(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_m(t) \end{pmatrix}, \text{ physical process}$$

$$\hat{X}(t;T) = \begin{pmatrix} \hat{x}_1(t;T) \\ \vdots \\ \hat{x}_m(t;T) \end{pmatrix}$$ = maximum-likelihood estimate of true $X(t)$, $t > 0$, given $Z(s)$, $0 \le s \le T$

$$Z(i) = \begin{pmatrix} z_i^1 \\ \vdots \\ z_i^p \end{pmatrix}$$ = random vector

$$Z(s) = \begin{pmatrix} z_1(s) \\ \vdots \\ z_p(s) \end{pmatrix}$$ = noise-corrupted observation at s, $0 \le s \le T$

$\mathbf{z}_1$ = unit vector along z-coordinate

z_i^k = random variable related to $z_k(s)$, $0 < s \le T$

α = vector of arbitrary parameters

β = inverse of reentry-vehicle ballistic coefficient

γ = weighting factor in maximum-likelihood estimate

$\gamma^k(\hat{X}), \mu^k(\hat{X})$ = infinite sums related to z_i^k, $m_i^k(\hat{X})$, λ_i^k

$\gamma_\ell^k(\hat{X}), \mu_\ell^k(\hat{X})$ = finite sums related to z_i^k, $m_i^k(\hat{X})$, λ_i^k

Δs = index interval used in numerical integration

ΔT = step size at which update $\hat{X}(0;T)$ in T

δ = computer execution time per Δs calculation

$\delta(s - u)$ = delta function

$$\varepsilon = \begin{pmatrix} \varepsilon_r \\ \varepsilon_{\dot{r}} \\ \varepsilon_\theta \\ \varepsilon_{\dot{\theta}} \end{pmatrix} = \begin{pmatrix} |r_0 - \hat{r}_0| \\ |\dot{r}_0 - \hat{\dot{r}}_0| \\ |\theta_0 - \hat{\theta}_0| \\ |\dot{\theta}_0 - \hat{\dot{\theta}}_0| \end{pmatrix}$$

$\zeta_k(s)$ = function related to $r_k(s,u)$ and $z_k(u)$

η = average number of Newtonian iterations per T_i

η_i = number of Newtonian iterations at T_i

θ = angle between zenith (z-axis) and radius vector to reentry vehicle

$\bar{\theta}_1$ = unit angular vector in plane of motion

Λ = covariance of the joint gaussian distribution of $X(0)$ and α

μ = mean of the joint gaussian distribution of $X(0)$ and α

ν = $\Delta T/\Delta s$

ξ_1, ξ_2 = stopping thresholds

ρ = air density at altitude

τ = total computer execution time for processing signal

τ_i = computer execution time at T_i

$\emptyset$ = zero vector

$\{\psi_i^k(s), \lambda_i^k\}$ = orthonormal eigenvectors and related eigenvalues of $r_k(s,u)$

$\omega(X_0,T)$ = bias of the MLE $\hat{X}(0,T)$ of X_0

$\nabla_C w(X), \nabla_{\hat{X}} w(\hat{X})$ = gradient of $w(\cdot)$ with respect to m-vector argument

$\nabla_{CC} w(X), \nabla_{\hat{X}\hat{X}} w(\hat{X})$ = second gradient of $w(\cdot)$ with respect to m-vector argument

$||X||$ = norm of X, over interval (0,T)

I. PROBLEM DEFINITION

GENERAL PROBLEM STATEMENT

A general problem of much interest and importance in prediction, estimation, and control areas is the following: Given an m-vector physical process

$$X(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_m(t) \end{pmatrix}, \tag{1}$$

where $X(t)$ satisfies a differential equation

$$\frac{dX(t)}{dt} = F[X;t;\alpha], \tag{2}$$

and where α is a vector of arbitrary parameters (some known, some unknown), given some a priori statistics on the value assumed by the initial condition $X(0)$, and given a noise-corrupted observation over some (variable) interval $[0,T]$ of a p-vector process related to $X(t)$:

$$Z(s), \qquad 0 \le s \le T, \tag{3}$$

then it is desired to make a best estimate of $X(t)$, $0 \le t$, in terms of the observation $\{Z(s), 0 \le s \le T\}$ and the a priori $X(0)$ statistics. If some components of the parameter vector α are unknown, then they may also have to be estimated in the process.

This problem has been essentially solved for the case where $F(\cdot)$ is linear and the noise-corrupting $Z(t)$ is additive.[1] This Memorandum will construct and apply a numerical algorithm for solving this problem when $F(\cdot)$ is nonlinear, but the corrupting noise is still

additive. The generality of the above statement will be restricted somewhat for present purposes, but the results can then be extended in obvious ways to cover the more general case.

SPECIAL NONLINEAR CASE TREATED

The vector $X(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_m(t) \end{pmatrix}$ will be a real m-vector physical process satisfying the nonlinear, time-invariant differential equation

$$\frac{dX(t)}{dt} = F(X;\alpha) = \begin{pmatrix} f_1(X;\alpha) \\ \vdots \\ f_m(X;\alpha) \end{pmatrix}, \tag{4}$$

where $F(X;\alpha)$ depends upon an unknown but fixed (throughout the observation interval) vector parameter α: known parameters are subsumed in $F(\cdot;\alpha)$. Observations of the p-vector $Z(s)$ over a varying interval $[0,T]$ are available, where it is known that

$$Z(s) = H(s)X(s) + N(s), \qquad 0 < s \leq T, \tag{5a}$$

where

$$H(s) = \begin{pmatrix} h_{11}(s) & \dots & h_{1m}(s) \\ \vdots & & \vdots \\ h_{p1}(s) & \dots & h_{pm}(s) \end{pmatrix}, \tag{5b}$$

$$N(s) = \begin{pmatrix} n_1(s) \\ \vdots \\ n_p(s) \end{pmatrix}, \tag{5c}$$

and

$$Z(s) = \begin{pmatrix} z_1(s) \\ \vdots \\ z_p(s) \end{pmatrix} . \tag{5d}$$

It is assumed that $F(X;\alpha)$ and $H(s)$ are smooth functions, e.g., analytic. Finally, it is assumed that the initial condition on $X(s)$ at time zero, $X(0)$, and the values for α have a joint gaussian distribution for which the distribution mean and covariance are known:

$$E\begin{pmatrix} X(0) \\ \alpha \end{pmatrix} = \mu ,$$

$$E\left\{\left[\begin{pmatrix} X(0) \\ \alpha \end{pmatrix} - \mu\right]\left[\begin{pmatrix} X(0) \\ \alpha \end{pmatrix} - \mu\right]^*\right\} = \Lambda \qquad \text{positive definite,} \tag{6}$$

and that the noise process $N(s)$ has

$$EN(s) = 0 \tag{7}$$

and covariance function

$$E[N(\zeta)N^*(s)] = R(\zeta,s) \qquad \text{positive definite} \tag{8}$$

with n_i independent of n_j, $i \neq j$; thus $R(\zeta,s)$ is a diagonal matrix

$$E[N(\zeta)N^*(s)] = R(\zeta,s) = \begin{pmatrix} r_1(\zeta,s) & & \bigcirc \\ & \ddots & \\ \bigcirc & & r_p(\zeta,s) \end{pmatrix} . \tag{9}$$

The problem is then to construct a best estimate of α and $X(t)$, $0 \le t$, given $\{Z(s),\ 0 \le s \le T\}$.

HISTORY OF THE PROBLEM

This problem intersects related research areas which have been under investigation for several decades: optimum filtering, control theory, and prediction and estimation. An early form was the Wiener filtering problem, now largely solved for linear $F(\cdot)$ and additive gaussian noise. In the linear case the optimality criteria of minimum RMS error, as well as of maximum likelihood, were successfully pursued. It is well known that in all cases (linear or not) the minimum RMS estimator is given by the conditional expectation of $X(0)$, given $Z(s)$, $0 \le s \le T$. In the linear case, it is also well known that the conditional expectation and the maximum-likelihood estimate (given $Z(s)$, $0 \le s \le T$) coincide.

Only in the simplest linear cases is it possible to derive a simple closed expression for the conditional expectation. Rather, it is often necessary to derive the differential equation which the estimator satisfies as a function of t and T, and then to solve this differential equation as T increases. This is done for the linear case in Ref. 1 by Kalman and Bucy.

In the nonlinear case, it is natural to try to extend the results for the conditional expectation, since it satisfies the intuitively appealing minimum RMS criterion. The increased complexity of the nonlinear case forces use of the differential-equation approach, via which the estimator is calculated numerically as T increases. However, the maximum-likelihood estimator no longer coincides in general with the conditional expectation. Since it has certain computational advantages over the conditional expectation, the maximum-likelihood estimator has also been explored via the differential-equation approach. Both

approaches (conditional expectation and maximum-likelihood estimator) involve the conditional-probability function of $X(t)$, given $Z(s)$, $0 \le s \le T$. It, too, can be approached via the differential equation, which it must satisfy as a function of t and T. This is related to the differential equation which the unconditional-probability function of $X(t)$ must satisfy. See Ref. 2 for a discussion of these so-called diffusion equations.

An early attempt to derive such differential equations for the conditional-probability function is given in Ref. 3 by Stratonovich. The continuous-time results were incorrect but have been subsequently derived correctly, as noted below. An interesting and useful early effort is described in Ref. 4, by W. M. Wonham; it addresses the discrete case primarily and suggests a heuristic extension to continuous observations.

A natural way to extend the earlier linear results is to generalize the filters from linear weighting coefficients of the observation to more complicated functionals of the observation. In Ref. 5, Balakrishnan develops the theory of functionals upon the space of random observations, shows that functionals useful for our problem can be approximated by certain kinds of polynomials, and then shows how the method of steepest descent can be used to construct a polynomial approximation to a minimum RMS estimator.

In Refs. 6 and 7, Kushner derives correct differential equations which must be satisfied as functions of t by the conditional-probability function. Kushner there suggests approximating the optimal filter, which is infinite dimensional in the sense of requiring specification

of all moments of the conditional-probability function, by a finite filter--i.e., by ignoring all but a finite number of moments. Bucy in Ref. 8 develops these differential-equation results of Kushner's also.

Another natural way to extend linear iterative results to nonlinear problems is to try to extend the weighting techniques straightforwardly to approximate optimum nonlinear filters. This is done in Ref. 9 by Mowery for the discrete-observation case; he displays some interesting error calculations for the approximate estimators developed. Another approach is that of invariant imbedding; in Ref. 10, Bellman et al. derive a numerical (computer) technique for finding X(t) which minimizes the usual quadratic norm over (0,T) of the difference between X(s), $0 \leq s \leq T$ and the observation. Several examples are calculated out to show convergence performance of the estimation.

In Ref. 11, Friedland and Bernstein derive differential equations for a first-order approximation to the maximum-likelihood estimator for the discrete sample case and then derive an analogous first-order approximation for the continuous-time case. In Ref. 12, Bass et al. derive an approximation for the conditional expectation. Specifically, by neglecting higher-order terms in the noise and in error differences, differential equations governing the evolution with T of the conditional expectation are derived; the approximation for the RMS error (relative to conditional expectation) matrix is similarly derived. These two differential equations theoretically could be jointly solved numerically to calculate the approximated conditional expectation.

Various calculations have been made in order to compare the efficiency of the various possible estimators in very special cases. One

such calculation by Carney and Goldwyn is described in Ref. 13. There, a (scalar) nonlinear estimation problem relative to a linear dynamic system is treated by estimators based on least-squares and maximum-likelihood criteria; in a linearized form of the problem, two versions of the Kalman-Bucy estimator are used. Monte Carlo techniques are used to generate comparative error statistics. In all cases, and over wide ranges of governing parameters, the maximum-likelihood estimate is preferred. Kushner proposes in an exploratory way in Ref. 14 possibly useful finite filters, derived either by assuming specific forms for the conditional-probability function or by similar assumptions on a finite number of the conditional moments. Several special results are then calculated.

In Ref. 15 Kushner derives the exact differential equation which the conditional mode at time T must satisfy. However, solution of this equation is not possible, since it involves higher derivatives of all orders of the conditional probability distribution. Recourse would therefore have to be made to approximations. However, the results of Carney and Goldwyn (Ref. 13) suggest that the maximum-likelihood estimate for the initial condition is preferred to the various approximations tried there and therefore that an exact maximum-likelihood estimate of the initial condition would be preferable to approximations to either the conditional expectation or the maximum-likelihood estimator.

The numerical comparisons made in Ref. 13 and elsewhere are for the scalar case, in order to render the required work load acceptable. This Memorandum is addressed to the vector situation, to examine the feasibility of numerically solving the vector differential equations

satisfied by the maximum-likelihood estimator. The calculations required are sufficiently large that it is not possible to generate error statistics allowing comparison with other techniques. Rather, attention is focused upon a specific observation sample, and the questions of computational feasibility, convergence, and solution usefulness are examined.

II. NUMERICAL TECHNIQUE

In general, the components of $X(0)$ can be expanded to include α, and those of $F(X;\gamma)$ to include $\gamma(t) \equiv \gamma(0) = \gamma$ (i.e., $\frac{d\gamma}{dt} = \emptyset$); and γ therefore can be estimated as part of the generalized $X(0)$. This is assumed done in the following, in which we have set $F(X;\gamma) = F(X)$.

CONDITIONAL EXPECTATION

As discussed in Section I, a desirable criterion for a "best" estimate of $X(t)$, $0 < t$, in terms of $\{Z(s), 0 \le s < T\}$, is that function $\hat{X}(t)$ which is unbiased and minimizes

$$E[\hat{X}(t) - X(t)]^{*}[\hat{X}(t) - X(t)]. \tag{10}$$

It is well known that the conditional expectation of $X(t)$, given $\{Z(s), 0 \le s < T\}$, minimizes (10):

$$\hat{X}(t) = E[X(t) | Z(s), 0 \le s < T]$$

$$= E[X(t) \quad \text{given} \quad Z(s), 0 \le s < T]. \tag{11}$$

In the vector-linear situation, $\hat{X}(t)$ as defined by (11) can be calculated (see Ref. 16, pp. 3-60 - 3-62; also Ref. 1). However, for the vector nonlinear situation under consideration here, such is not the case.

MAXIMUM-LIKELIHOOD ESTIMATE

An estimate alternate to that which minimizes (10) is the so-called maximum-likelihood estimate (MLE) of $X(t)$: the value $\hat{X}(t)$ which, given $Z(s)$, $0 \le s < T$, has the maximum conditional probability of occurrence $p(\hat{X}(t) | Z(s), 0 \le s < T)$. This may be developed as shown

in Appendix A (where necessary and sufficient conditions are also given) and leads to the following results.

Defining

$$X(s;C) = X(s), \quad \text{given that } X(0) = C,$$

and

$$X(0) = X_0,$$

then

$$p(C|Z(s),\ 0 \le s \le T) = \frac{G(C)}{\int_C G(C)\,dC}, \tag{12a}$$

where

$$G(C) = \exp\left[-\frac{1}{2}\left\{(C-\mu)^* \Lambda^{-1}(C-\mu) + \sum_{k=1}^{p}\int_0^T z_k(s)\zeta_k(s)\,ds \right.\right.$$
$$\left.\left. - 2\sum_{k=1}^{p}\int_0^T z_k(s)g^k(s;C)\,ds + \sum_{k=1}^{p}\int_0^T H_k(s)X(s;C)g^k(s;C)\,ds\right\}\right], \tag{12b}$$

and where $\zeta_k(s)$ and $g^k(s;C)$ satisfy

$$\int_0^T \zeta_k(s)r_k(s,u)\,ds = z_k(u), \quad \int_0^T g^k(s;C)r_k(s;u)\,ds = H_k(u)X(u;C). \tag{12c}$$

Necessary and sufficient conditions that these results hold are developed in Appendix A and are

1. $r_k(s,u)$ positive definite, $k = 1, \ldots, p$,

2. $\displaystyle\sum_{i=1}^{\infty} \frac{[m_i^k(C)]^2}{(\lambda_i^k)^2} < \infty$, $k = 1, \ldots, p$,

$$3. \quad \sum_{i=1}^{\infty} \frac{[z_i^k]^2}{(\lambda_i^k)^2} < \infty, \; k = 1, \ldots, p.$$

where z_i^k and $m_i^k(C)$ are defined by

$$z_i^k = \int_0^T z_k(s)\phi_i^k(s)ds,$$

$$m_i^k(C) = E[z_i^k|C] = \int_0^T H_k(s)X(s;C)\phi_i^k(s)ds,$$

and $\{\phi_i^k(s), \lambda_i^k\}$ are the orthonormal eigenvectors and related eigenvalues of $r_k(s,u)$; that is, $\{\phi_i^k(s), \lambda_i^k\}$ satisfy

$$\int_0^T r_k(s;u)\phi_i^k(u)du = \lambda_i^k\phi_i^k(s).$$

Equation (12) simplifies greatly when the noise N(s) is such as to permit its development as in Appendix A <u>and</u> the $r_k(s,u)$ can be treated as delta functions:

$$r_k(s,u) = \Phi_k\delta(s - u), \; k = 1, \ldots, p.$$

Then (12) becomes

$$p(C|Z(s), 0 \le s \le T) = \frac{J(C)}{\int_C J(C)dC}, \qquad (13a)$$

where

$$J(C) = \exp\left[-\frac{1}{2}\left\{(C-\mu)^{*}\Lambda^{-1}(C-\mu)\right.\right.$$

$$\left.\left.+\int_0^T [Z(s) - H(s)X(s;C)]^{*}R^{-1}[Z(s) - H(s)X(s;C)]ds\right\}\right], \qquad (13b)$$

$$R = \begin{pmatrix} \Phi_1 & & 0 \\ & \ddots & \\ 0 & & \Phi_p \end{pmatrix}. \qquad (13c)$$

Equation (12) can now be used to write a general expression for the MLE. However, in order to reduce the computing complexities, the MLE will be derived and applied for Eq. (13), the specialization of (12) to the case of white noise. This is an important case in its own right and can be generalized straightforwardly.

Since the denominator of (13), for given $Z(s)$, $0 \le s \le T$, is a constant, the MLE is that C which maximizes the numerator, which is clearly that C which <u>minimizes</u>

$$f(C,T) = (C-\mu)^{*}\Lambda^{-1}(C-\mu)$$

$$+\int_0^T [Z(s) - H(s)X(s;C)]^{*}R^{-1}[Z(s) - H(s)X(s;C)]ds. \qquad (14)$$

Direct minimization of (14) for fixed T is in general impossible because the available numerical techniques require an initial-estimate C close enough to the sought $\hat{X}(T)$ to guarantee convergence.* For arbitrary T this is not available. Rather, an iterative numerical

*$\hat{X}(T)$ may or may not equal X_0, the true initial condition.

procedure which builds upon preceding estimates for increasing values of T must be devised. This can be done by developing the differential equation which the best estimate must satisfy as a function of time T, the upper limit of the interval of observation. As indicated below, the MLE of $X(t)$, $0 < t$, follows naturally from the MLE of $X(0)$, so the latter will be estimated.

Let us assume that for each T over a range of T beginning at $T = 0$ there does exist a unique value of the vector C, $\hat{X}(T)$, which minimizes (14). Then $\hat{X}(T)$ must be the solution, for each T, of

$$\nabla_C f(C,T) = 2\Lambda^{-1}(C - \mu) - 2\int_0^T \nabla_C X^*(s;C)H^*(s)R^{-1}[Z(s) - H(s)X(s;C)]ds = \emptyset. \tag{15}$$

That is, $\hat{X}(T)$ must satisfy

$$\hat{X}(T) = \mu + \Lambda \int_0^T \nabla_C X^*(s;\hat{X})H^*(s)R^{-1}[Z(s) - H(s)X(s;\hat{X})]ds. \tag{16}$$

But if $\hat{X}(T)$ satisfies (16) over the entire assumed T range, then it satisfies there

$$\frac{d\hat{X}}{dT} = \frac{\partial\hat{X}}{\partial T}\frac{dT}{dT} + \nabla_C\hat{X}(T)\frac{dX}{dT}. \tag{17}$$

It is useful to write (17) as shown, because it explicitly identifies the variation in $\hat{X}(T)$ due to variation in T and that due to variation in $\hat{X}$. However, because of the white-noise component in $Z(s)$,

(17) can only be viewed as a symbolic representation of a corresponding differential expression:

$$d\hat{X} = \nabla_C X^*(T,\hat{X})H^*(T)R^{-1}H(T)[X(T;X_0) - X(T;\hat{X})]dT$$
$$+ \nabla_C X^*(T;\hat{X})H^*(T)R^{-1}dL(T) + \nabla_C\hat{X}(T)d\hat{X} ,$$

where

$$L(s) = \begin{pmatrix} \ell_1(s) \\ \vdots \\ \ell_p(s) \end{pmatrix} ,$$

and the $\ell_i(s)$ are appropriate independent scalar Wiener processes.[2,4] In the following the convenient symbolism of (17) will be used. The results are correct as long as only differential expressions are used in the calculations; this is the case. Of course, a true differential equation analogous to (17) does apply in the colored-noise case. However, computations then are more difficult because of the need to calculate $g^k(s;C)$ of Eq. (12) as T and C change.

Equation (17) is equivalent to

$$[I - \nabla_C\hat{X}(T)]\frac{d\hat{X}}{dT} = \frac{\partial\hat{X}}{\partial T} = [\nabla_C X^*(T;\hat{X})H^*(T)R^{-1}[Z(T) - H(T)X(T,\hat{X})]. \quad (18)$$

From (16),

$$\nabla_C\hat{X}(T) = \int_0^T \nabla_{CC}X^*(s;\hat{X})H^*(s)R^{-1}[Z(s) - H(s)X(s;\hat{X})]ds$$
$$- \int_0^T \nabla_C X^*(s;\hat{X})H^*(s)R^{-1}H(s)\nabla_C X(s;\hat{X})ds,$$

or

$$I - \nabla_C \hat{X}(T) = \left[P^{-1} + \int_0^T \nabla_C X^*(s;\hat{X}) H^*(s) R^{-1} H(s) \nabla_C X(s;\hat{X}) ds \right.$$

$$\left. - \int_0^T \nabla_{CC} X^*(s;\hat{X}) H^*(s) R^{-1} [Z(s) - H(s) X(s;\hat{X})] ds \right].$$

But from (15),

$$\nabla_{CC} f(\hat{X},T) = 2P^{-1} + 2 \int_0^T \nabla_C X^*(s;\hat{X}) H^*(s) R^{-1} H(s) \nabla_C X(s;\hat{X}) ds$$

$$- 2 \int_0^T \nabla_{CC} X^*(s;\hat{X}) H^*(s) R^{-1} [Z(s) - H(s) X(s;\hat{X})] ds, \qquad (19)$$

and thus

$$I - \nabla_C \hat{X}(T) = \frac{1}{2} \nabla_{CC} f(\hat{X},T),$$

so that (18) becomes

$$\nabla_{CC} f(\hat{X},T) \frac{d\hat{X}}{dT} = 2 \nabla_C X^*(T;\hat{X}) H^*(T) R^{-1} [Z(T) - H(T) X(T;\hat{X})]. \qquad (20)$$

Expression (20) is the fundamental differential equation of interest; and if $\nabla_{CC} f(\hat{X},T)$ is nonsingular, it can be written as

$$\frac{d\hat{X}}{dT} = 2[\nabla_{CC} f(\hat{X},T)]^{-1} \nabla_C X^*(T;\hat{X}) H^*(T) R^{-1} [Z(T) - H(T) X(T;\hat{X})]. \qquad (21)$$

It was assumed above that there did exist a unique solution to (15) which minimized (14) over some T interval beginning at T = 0. A necessary and sufficient condition for this is that $\nabla_{CC} f(C,T)$ be positive

definite at $C = \hat{X}$ (for constant T), which requires that $\nabla_{CC} f(C,T)$ be nonsingular in that interval. Therefore, under this assumption, (20) can be solved for $\frac{d\hat{X}}{dT}$ in the form of (21).

Further, at T = 0, (16) and (19) imply

$$\hat{X}(0) = \mu \qquad \text{and} \qquad \nabla_{CC} f(\hat{X},0) = 2\Lambda^{-1}. \tag{22}$$

Thus at T = 0 we are guaranteed the nonsingularity of $\nabla_{CC} f(\hat{X},T)$; the legitimacy of the form (21); and the initial conditions with which to start a numerical solution:

$$\hat{X}(0) = \mu$$

$$\left.\frac{d\hat{X}}{dT}\right|_{T=0} = \Lambda H^{*}(0)R^{-1}[Z(0) - H(0)\mu], \tag{23}$$

since $\nabla_C X(0;\hat{X}) = I$.

Furthermore, (22) implies that $\nabla_{CC} f(\hat{X},T)$ is nonsingular in some finite T neighborhood to the right of T = 0; since

$$\int_0^T \nabla_C X^{*}(s;\hat{X})H^{*}(s)R^{-1}H(s)\nabla_C X(s;\hat{X})ds$$

$$- \int_0^T \nabla_{CC} X^{*}(s;\hat{X})H^{*}(s)R^{-1}[Z(s) - H(s)X(s;\hat{X})]ds \tag{24}$$

is a continuous function of T, it follows from (19) that $\nabla_{CC} f(\hat{X},T)$ must be nonsingular over some T neighborhood of T = 0. Finally, $\nabla_{CC} f(\hat{X},0) = 2\Lambda^{-1}$, so that Λ being positive definite makes f(C,T) convex at (C,T) = (μ,0). Again, the continuity in T of (24) guarantees that $\nabla_{CC} f(\hat{X},T)$ will be convex in some T neighborhood of T = 0.

The existence of a minimizing solution to $\nabla_C f(C,T) = \emptyset$ has been shown in Ref. 17 to exist as per the following:

If

$$\int_0^T X^*(s)F(X(s),s)ds \le K(T)(1 + ||X||^2)$$

for each T, where $\dot{X}(s) = F(X(s),s)$ and $||X||$ is the norm of $X(s)$ over $(0,T)$, then for each T there exists a minimizing solution $\hat{X}(T)$, and at each such T the matrix $\nabla_{CC} f(\hat{X},T)$ is positive definite. This condition is widely met. Summarizing the foregoing, we have:

<u>THEOREM</u>. Given Λ and R positive definite and that for all T of interest

$$\int_0^T X^*(s)F(X(s),s)ds \le K(T)(1 + ||X||^2),$$

then there exists a solution $\hat{X}(T)$ to

$$\nabla_C f(\hat{X},T) = 2\Lambda^{-1}(\hat{X} - \mu)$$

$$- 2\int_0^T \nabla_C X^*(s;\hat{X})H^*(s)R^{-1}[Z(s) - H(s)X(s;\hat{X})]ds = \emptyset \qquad (25)$$

which minimizes

$$f(C,T) = (C - \mu)^*\Lambda^{-1}(C - \mu)$$

$$+ \int_0^T [Z(s) - H(s)X(s;C)]^*R^{-1}[Z(s) - H(s)X(s;C)]ds. \qquad (26)$$

Further, $\nabla_{CC} f(\hat{X},T)$ is positive definite, and $\hat{X}(T)$ satisfies the differential equation

$$\frac{d\hat{X}}{dT} = 2\left[\nabla_{CC} f(\hat{X},T)\right]^{-1} \nabla_C X^*(T,\hat{X}) H^*(T) R^{-1}\left[Z(T) - H(T)X(T;\hat{X})\right] \tag{27}$$

with initial conditions at $T = 0$:

$$\hat{X}(0) = \mu \tag{28a}$$

$$\left.\frac{d\hat{X}}{dT}\right|_{T=0} = \left[H^*(0)R^{-1}\left[Z(0) - H(0)\mu\right]\right] \tag{28b}$$

The usefulness of these results in calculating the MLE for $X(0)$ will depend upon

1. The ease with which a numerical solution of (27) can be implemented,
2. The integration interval which must be used in order to guarantee continuation in the neighborhood of the desired solution,
3. The nature of the solution to which $\hat{X}(T)$ converges with increasing T,
4. The rate at which convergence of $\hat{X}(T)$ occurs.

These depend, naturally, upon R and μ in a critical way.

Since the solution of

$$\frac{dX}{dt} = F(X)$$

corresponding to a given $X(0)$ is unique, then one may immediately extend a MLE for $X(0)$ to a MLE for $X(t)$, $t < 0$. For in that case

it is natural to take as the MLE of $X(t)$, given $\hat{X}(0)$,

$$\hat{X}(t;T) = X(t;\hat{X}(0);T), \tag{29}$$

where to show the dependence upon both t and T we have written $\hat{X}(t;T)$ = MLE of $X(t)$.

It therefore suffices to estimate $X(0)$ and then numerically integrate

$$\frac{dX(t)}{dt} = F(X) \tag{30}$$

from this initial condition to find the MLE of $X(t)$. This is the approach assumed in the application. However, it is of interest to develop the differential equation which $\hat{X}(t;T)$ satisfies.

By (29)

$$\hat{X}(t;T) = X(t;\hat{X}(0);T),$$

where $\hat{X}(0;T)$ = MLE of $X(0)$ given $Z(s)$, $0 \le s \le T$.

Then

$$\frac{d\hat{X}(t;T)}{dT} = \frac{\partial X(t;\hat{X}(0;T))}{\partial T} + \nabla_{\hat{X}} X(t;\hat{X}(0;T)) \frac{d\hat{X}(0;T)}{dT};$$

but

$$\frac{\partial X(t;\hat{X}(0;T))}{\partial T} = \begin{cases} 0, & t \le T \\ F(X(T)) = F(X), & t = T \end{cases}$$

and (27) then imply

$$\frac{d\hat{X}(t;T)}{dT} = \begin{cases} 2\nabla_C X(t,\hat{X}[0;T],T)[\nabla_{CC} f(\hat{X}[0;T],T)]^{-1}\nabla_C X^*(T;\hat{X}[0;T])H^*(T) \\ \qquad R^{-1}[Z(T) - H(T)X(T;\hat{X}[0;T])] \\ \qquad\qquad \text{for } t \neq T \\ \qquad\qquad\qquad \text{and} \\ F[X(T;\hat{X}[0;T])] + 2\nabla_C X(T;\hat{X}[0;T]) \\ \qquad [\nabla_{CC} f(\hat{X}[0;T],T)]^{-1}\nabla_C X^*(T;\hat{X}[0;T]) \\ \qquad H^*(T)R^{-1}[Z(T) - H(T)X(T;\hat{X}[0;T])] \\ \qquad\qquad \text{for } t = T. \end{cases} \tag{31}$$

Equation (31) may be rewritten as

$$\frac{d\hat{X}(t;T)}{dT} = \begin{cases} 2\nabla_C \hat{X}(t;T)[\nabla_{CC} f(\hat{X}[0;T],T)]^{-1}\nabla_C \hat{X}^*(T;T)H^*(T)R^{-1} \\ \qquad [Z(T) - H(T)\hat{X}(T;T)] \\ \qquad\qquad \text{for } t \neq T \\ \qquad\qquad\qquad \text{and} \\ F[\hat{X}(T;T)] + 2\nabla_C \hat{X}(T;T)[\nabla_{CC} f(\hat{X}[0;T],T)]^{-1}\nabla_C \hat{X}^*(T;T)H^*(T)R^{-1} \\ \qquad [Z(T) - H(T)\hat{X}(T;T)] \\ \qquad\qquad \text{for } t = T. \end{cases} \tag{32}$$

In the above, $\hat{X}(t;T)$ has been used to indicate dependence upon t and T; note that $\hat{X}(T) = \hat{X}(0,T)$, from earlier nomenclature. We continue to use $\hat{X}(T)$ in the following.

III. ERROR STATISTICS

A lower bound for the conditional-error statistics of the MLE developed in Section II will be derived--specifically, the unbiased Cramer-Rao bound. It has not been possible to derive an exact error estimate, or an upper bound which is useful. When T is large, the estimate can be expected to become more and more independent of the a priori matrix Λ. The bound developed here does not involve Λ and could only apply for large T. As such, it is a useful bound with which to compare convergence of the estimator for large T.

Defining the bias of the MLE $\hat{X}(T)$ of X_0 as

$$a(X_0,T) = E(\hat{X}(T) - X_0 | X_0), \tag{33}$$

then a direct generalization of the results of Ref. 16 (pp. 3-2 - 3-5) to the continuous sample time case gives

$$\begin{aligned} \varepsilon^2(X_0,T) &= \text{Trace } E\{(\hat{X}(T) - X_0)(\hat{X}(T) - X_0)^* | X_0\} \\ &\geq \text{Trace } \{(I + \nabla_C a(X_0,T))^* (I + \nabla_C a(X_0,T)) G^{-1}\}, \end{aligned} \tag{34a}$$

where the information matrix

$$G = E\{[\nabla_C \log p(Z(s), 0 \leq s \leq T | X_0)][\nabla_C \log p(Z(s), 0 \leq s \leq T | X_0)]^* | X_0\}. \tag{34b}$$

This is the Cramer-Rao bound; and that (34) holds for the continuous sample time case follows from a limiting development analogous to that in Appendix A. Directly from Appendix A follows the expression

$$p(Z(s),\ 0 < s < T|X_0) = K \exp\left\{-\frac{1}{2}\int_0^T [Z(s) - H(s)X(s;X_0)]^* R^{-1} [Z(s) - H(s)X(s;X_0)]ds\right\}, \tag{35}$$

where K is an appropriate normalizing constant.

We then have

$$4G = E\left\{\int_0^T\int_0^T \nabla_C X^*(s;X_0)H^*(s)R^{-1}[H(s)X(s;X_0) + N(s) - H(s)X(s;X_0)]^* [H(u)X(u;X_0) + N(u) - H(u)X(u;X_0)]^* R^{-1}H(u)\nabla_C X(u;X_0)ds\,du|X_0\right\}$$

$$= \int_0^T\int_0^T \nabla_C X^*(s;X_0)H^*(s)R^{-1}R\delta(s-u)R^{-1}H(u)\nabla_C X(u;X_0)ds\,du, \tag{36}$$

or

$$G = \frac{1}{4}\int_0^T \nabla_C X^*(s;X_0)H^*(s)R^{-1}H(s)\nabla_C X(s;X_0)ds. \tag{37}$$

Substituting (37) into (34) gives

$$\epsilon^2(X_0,T) > 4\ \mathrm{Trace}\left\{[I + \nabla_C a(X_0,T)]^*[I + \nabla_C a(X_0,T)]\left[\int_0^T \nabla_C X^*(s;X_0)H^*(s)R^{-1}H(s)\nabla_C X(s;X_0)ds\right]^{-1}\right\}. \tag{38}$$

A singularity does not occur in (38) at T = 0, since at T = 0 the coefficient $[I + \nabla_C a(X_0,T)]$ is zero. For, from (16) the MLE $\hat{X}(T)$ satisfies (suppressing t' = T in X[T] in the right-hand side)

$$\hat{X}(T) = \mu + \int_0^T \nabla_C X^*(s;\hat{X})H^*(s)R^{-1}[Z(s) - H(s)X(s;\hat{X})]ds. \tag{39}$$

Subtracting X_0 from (39) and taking the expectation gives

$$\omega(X_0,T) = \mu - X_0 + \Lambda E\left\{\int_0^T \nabla_C X^*(s;\hat{X})H^*(s)R^{-1}[Z(s) - H(s)X(s;\hat{X})]ds\,|X_0\right\}, \tag{40}$$

from which

$$\nabla_C\omega(X_0,T) = -I + \Lambda E\left\{\int_0^T \nabla_{CC}X^*(s;\hat{X})H^*(s)R^{-1}[Z(s) - H(s)X(s;\hat{X})]ds\,|X_0\right\}$$

$$- \Lambda E\left\{\int_0^T \nabla_C X^*(s;\hat{X})H^*(s)R^{-1}H(s)\nabla_C X(s;\hat{X})ds\,|X_0\right\}, \tag{41}$$

so that

$$I + \nabla_C\omega(X_0,T) = \Lambda E\left\{\int_0^T \nabla_{CC}X^*(s;\hat{X})H^*(s)R^{-1}[Z(s) - H(s)X(s;\hat{X})]ds\,|X_0\right\}$$

$$- \Lambda E\left\{\int_0^T \nabla_C X^*(s;\hat{X})H^*(s)R^{-1}H(s)\nabla_C X(s;\hat{X})ds\,|X_0\right\}, \tag{42}$$

which is zero for $T = 0$.

The expression (42) for $I + \nabla_C\omega(X_0,T)$ has not been evaluated exactly. Using gross eliminations, we heuristically show that $\nabla_C\omega(X_0,T)$ decreases monotonically as T increases. This amounts to linearizing the error expressions, so that the resulting expressions

are correct for the linear case but are only approximate in the non-linear situation. Assuming that $X(s;\hat{X})$ and $\nabla_C X(s;\hat{X})$ can be expanded about X_0 in power series, we have

$$X(s;\hat{X}) = X(s;X_0) + \nabla_C X(s;X_0)(\hat{X} - X_0)$$

$$+ \frac{1}{2!}(\hat{X} - X_0)^* \nabla_{CC} X(s;X_0)(\hat{X} - X_0) + \ldots , \tag{43}$$

and

$$\nabla_C X(s;\hat{X}) = \nabla_C X(s;X_0) + \nabla_{CC} X(s;X_0)(\hat{X} - X_0)$$

$$+ \frac{1}{2!}(\hat{X} - X_0)^* \nabla_{CCC} X(s;X_0)(\hat{X} - X_0) + \ldots . \tag{44}$$

Substituting (43) and (44) into (39) and linearizing the error expressions by dropping all terms in $(\hat{X} - X_0)$ higher than the first and in $(\hat{X} - X_0)^n$ and $N(s)$ where $n > 1$, there results

$$\hat{X} - X_0 = -(X_0 - \mu) - \Lambda \int_0^T \nabla_C X^*(s;X_0) H^*(s) R^{-1} H(s) \nabla_C X(s;X_0) ds (\hat{X} - X_0)$$

$$+ \Lambda \int_0^T \nabla_C X^*(s;X_0) H^*(s) R^{-1} N(s) ds .$$

Factoring out and solving for $\hat{X} - X_0$ gives (45) below, which is exact for the linear case: this fact can be shown simply by substituting $Z(s;X_0)$ for the linear case into (39), subtracting X_0 from both sides, and solving for $\hat{X} - X_0$.

$$\hat{X} - X_0 = \left[\Lambda^{-1} + \int_0^T \nabla_C X^*(s;X_0)H^*(s)R^{-1}H(s)\nabla_C X(s;X_0)ds\right]^{-1}$$

$$\left[-\Lambda^{-1}(X_0 - \mu) + \int_0^T \nabla_C X^*(s;X_0)H^*(s)R^{-1}N(s)ds\right]. \qquad (45)$$

Thus for large T

$$\omega(X_0,T) = E[\hat{X}(T) - X_0|X_0] = B^{-1}(X_0,T)$$

$$\left[-\Lambda^{-1}(X_0 - \mu) + \int_0^T \nabla_C X^*(s;X_0)H^*(s)R^{-1}E[N(s)]ds\right],$$

or

$$\omega(X_0,T) = -B^{-1}(X_0,T)\Lambda^{-1}(X_0 - \mu), \qquad (46a)$$

where

$$B(X_0,T) = \Lambda^{-1} + \int_0^T \nabla_C X^*(s;X_0)H^*(s)R^{-1}H(s)\nabla_C X(s;X_0)ds. \qquad (46b)$$

Equation (46), arrived at via the preceding linearization, can be used with (38) to estimate the biased Cramer-Rao bound. However, $B(X_0,T)$ can be written as

$$B(X_0,T) = \Lambda^{-1} + \int_0^T [R^{-\frac{1}{2}}H(s)\nabla_C X(s;X_0)]^*[R^{-\frac{1}{2}}H(s)\nabla_C X(s;X_0)]ds, \qquad (47)$$

which is the sum of two squares and therefore is monotonically increasing. If its limit as T increases is sufficiently large, then

$$B^{-1}(X_0,T) \underset{T\to\infty}{\longrightarrow} \mathfrak{c}(X_0) \doteq 0. \tag{48}$$

For such X_0 and for large T the inequality (38) becomes

$$\epsilon^2(X_0,T) > 4 \text{ Trace} \left\{ \left[\int_0^T \nabla_C X^*(s;X_0) H^*(s) R^{-1} H(s) \nabla_C X(s;X_0)\, ds \right]^{-1} \right\}, \tag{49}$$

which is the unbiased Cramer-Rao bound. The lower bound given in the right side of (49) is probably too small, in view of the assumptions employed in its heuristic development, and because (48) will not always be satisfied. In fact, comparing (46b) and (37), it is evident that (except for very large Λ^{-1}) $B^{-1}(X_0,T)$ very small implies that G is also very small; i.e., the unbiased Cramer-Rao bound is then also essentially zero. When (48) is not satisfied it is possible to approximate the biased Cramer-Rao bound using the gradient of $\omega(X_0,T)$ as found from (40a). However, the bound given by (49) will be displayed later for the application being made. As mentioned earlier, this result could only apply for large T, when it is expected that the estimate becomes independent of Λ, which does not appear in (49). In any case, the unbiased Cramer-Rao bound is an interesting standard with which to compare the estimate.

IV. AN APPLICATION: REENTRY PREDICTION

GENERAL

As an example application of the foregoing theoretical results, they will be used to estimate the initial conditions of a reentry vehicle (RV) which has been detected and is being tracked. It is assumed that a priori information is available relative to the space point at which the RV will be detected and to its velocity at that point. In the particular case at hand, the acquisition and tracking take place during the last 5 seconds before impact, so that it is natural to think of these a priori data as having been supplied by a long-range surveillance radar for purposes of acquisition and track initiation. In such a situation, it is natural to ask how to use these a priori data following acquisition, in order to make the best estimates and predictions during the tracking phase.

Depending upon the circumstances, it may be necessary to estimate certain RV parameters, either initially or continuously, in order to predict its reentry path. In particular, the RV lift and drag coefficients are often poorly known constants or functions of time. These, and other unknown parameters, could be estimated within the framework of the preceding theory, simply by including them as extra components of the vector $X(t)$ and including components describing their variation with time in $F(X)$. Similar remarks apply to certain variable environmental factors, such as air density at the reference altitude, which vary continuously. However, inclusion of these extra parameters in the estimation example greatly increases the complexity and difficulty of the computing task without adding very much to the usefulness

of the example. Therefore, the estimation procedure will assume that all RV and environmental parameters are known accurately.

Similar reductions in computing complexity can be made by restricting the RV motion to a well-known vertical plane containing the tracking radar. Estimation and prediction within just that plane is still a sufficiently rich problem to display the technique. Also RV characteristics are assumed to cause only drag accelerations; no lift or deflection forces operate (the latter is ruled out by the planar-motion assumption). Finally, for simplicity a flat earth will be assumed, as well as a force of gravity that is constant throughout the altitudes concerned and directed along the radar vertical. In the case at hand these latter are both acceptable assumptions because the ground range and altitude variations are so slight. However, even if they were not slight, an estimation example which treated them as constant would be as useful as if they were otherwise treated, as long as the equations of motion actually used are correctly reflected in the trajectory estimation procedure.

EQUATIONS OF MOTION

Given these restrictions, the situation can be displayed in spherical coordinates as in Fig. 1, where the vector meanings are noted. In spherical coordinates the equation of motion in one plane is[18]

$$\dot{\bar{V}} = \bar{r}_1[\ddot{r} - r\dot{\theta}^2] + \bar{\theta}_1[2\dot{r}\dot{\theta} + r\ddot{\theta}] = \bar{F}/m = [\bar{D} + \bar{G}]/m, \qquad (50)$$

where the dot represents the time derivative, the bar represents a vector, and

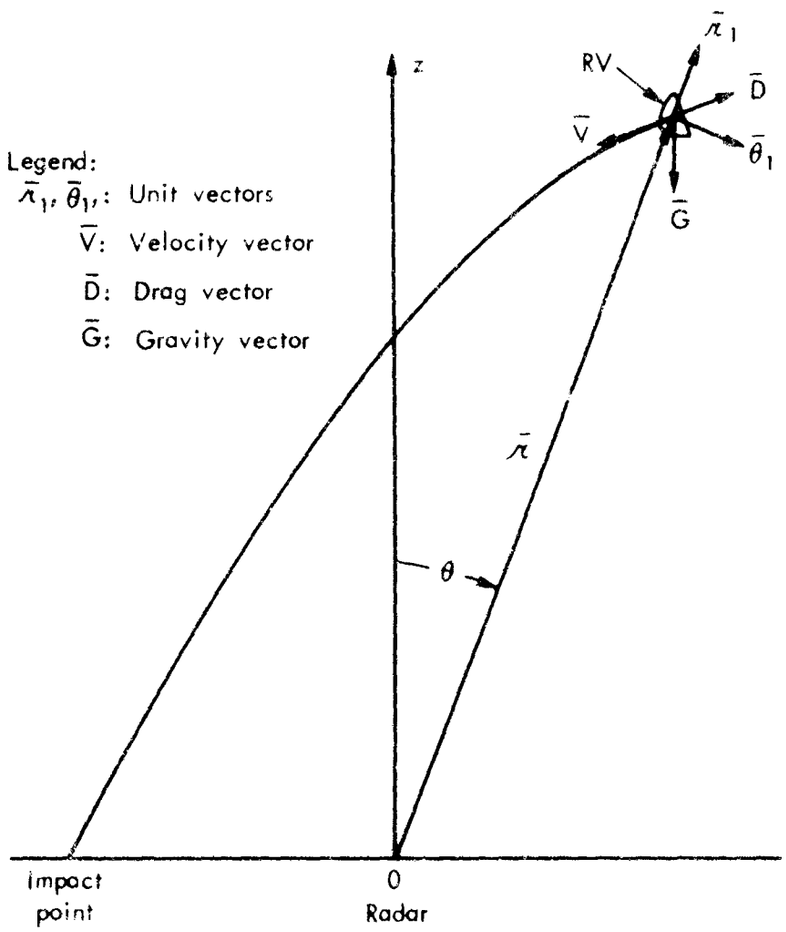

Fig. 1 — Space and vector relationships

$\bar{V}$ = RV velocity

$\bar{r}_1$ = unit radial vector

$\bar{a}_1$ = unit angular vector in plane of motion

r = radial distance to RV

θ = angle between zenith (z-axis) and radius vector to RV

$\bar{F}$ = total-force vector acting on RV

$\bar{D}$ = drag-force vector acting on RV

$\bar{G}$ = gravity-force vector acting on RV

m = RV mass

Force $\bar{D}$ may be written as

$$\bar{D} = -\frac{1}{2} C_D A_C \rho V^2 \bar{a}_V ,$$

where

C_D = RV drag coefficient

A_C = RV frontal (drag) area

ρ = air density at altitude

$\bar{a}_V$ = unit vector along the RV line of flight, or

$$\bar{a}_V = \frac{\bar{V}}{|V|} = \frac{\dot{r}\bar{r}_1 + r\dot{\theta}\bar{\theta}_1}{(\dot{r}^2 + r^2\dot{\theta}^2)^{1/2}}$$

Force $\bar{G}$, because of the simplifying assumption about its direction, may be written as

$$\bar{G} = -mg\bar{z}_1 = -mg \cos\theta\bar{r}_1 + mg \sin\theta\bar{\theta}_1 ,$$

where

g = gravitational acceleration

$\bar{z}_1$ = unit vector along z-coordinate

Thus

$$\bar{F} = \left[\left(-\frac{1}{2} C_D A_C \rho V^2\right) \frac{\dot{r}}{|V|} - mg \cos\theta\right] \bar{r}_1$$
$$+ \left[\left(-\frac{1}{2} C_D A_C \rho V^2\right) \frac{r\dot{\theta}}{|V|} + mg \sin\theta\right] \bar{\theta}_1 ,$$

and combining this with (50) gives

$$[\ddot{r} - r\dot{\theta}^2]\bar{r}_1 + [2\dot{r}\dot{\theta} + r\ddot{\theta}]\bar{\theta}_1 = -\left[\frac{1}{2} C_D A_C \rho V^2 \frac{\dot{r}}{|V|m} + g \cos\theta\right]\bar{r}_1$$
$$- \left[\frac{1}{2} C_D A_C \rho V^2 \frac{r\dot{\theta}}{|V|m} - g \sin\theta\right]\bar{\theta}_1 . \tag{51}$$

Equating coefficients of $\bar{r}_1$ and $\bar{\theta}_1$ in (51),

$$\ddot{r} - r\dot{\theta}^2 = -\frac{1}{2} C_D A_C \rho |V| \frac{\dot{r}}{m} - g \cos\theta , \tag{52a}$$

$$2\dot{r}\dot{\theta} + r\ddot{\theta} = -\frac{1}{2} C_D A_C \rho |V| \frac{r\dot{\theta}}{m} + g \sin\theta . \tag{52b}$$

Substituting $|V| = (\dot{r}^2 + r^2\dot{\theta}^2)^{1/2}$ into (52) gives

$$\ddot{r} - r\dot{\theta}^2 = -\frac{1}{2} C_D A_C \rho (\dot{r}^2 + r^2\dot{\theta}^2)^{1/2} \frac{\dot{r}}{m} - g \cos\theta , \tag{53a}$$

$$2\dot{r}\dot{\theta} + r\ddot{\theta} = -\frac{1}{2} C_D A_C \rho (\dot{r}^2 + r^2\dot{\theta}^2)^{1/2} \frac{r\dot{\theta}}{m} + g \sin\theta . \tag{53b}$$

If we now define

$$\beta = \frac{1}{W/C_D A_C} = \frac{1}{mg/C_D A_C} ,$$

where

$$W = mg$$

is the weight of the RV, there results the final form for the

equations of motion:

$$\ddot{r} - r\dot{\theta}^2 = -\frac{1}{2}\beta g\rho\dot{r}(\dot{r}^2 + r^2\dot{\theta}^2)^{1/2} - g\cos\theta, \tag{54a}$$

$$2\dot{r}\dot{\theta} + r\ddot{\theta} = -\frac{1}{2}\beta g\rho r\dot{\theta}(\dot{r}^2 + r^2\dot{\theta}^2)^{1/2} + g\sin\theta. \tag{54b}$$

We will take as the basic vector to be estimated or predicted

$$X(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{pmatrix} = \begin{pmatrix} r(t) \\ \dot{r}(t) \\ \theta(t) \\ \dot{\theta}(t) \end{pmatrix}. \tag{55}$$

From (55),

$$\frac{dx_1}{dt} = x_2(t) \qquad \text{and} \qquad \frac{dx_3}{dt} = x_4(t).$$

From (54),

$$\frac{dx_2}{dt} = \ddot{r} = r\dot{\theta}^2 - \frac{1}{2}\beta g\rho\dot{r}(\dot{r}^2 + r^2\dot{\theta}^2)^{1/2} - g\cos\theta$$

$$= x_1x_4^2 - \frac{1}{2}\beta g\rho x_2(x_2^2 + x_1^2x_4^2)^{1/2} - g\cos x_3,$$

and

$$\frac{dx_4}{dt} = \ddot{\theta} = \frac{1}{r}\left\{-2\dot{r}\dot{\theta} - \frac{1}{2}\beta g\rho r\dot{\theta}(\dot{r}^2 + r^2\dot{\theta}^2)^{1/2} + g\sin\theta\right\}$$

$$= \frac{1}{x_1}\left\{-2x_2x_4 - \frac{1}{2}\beta g\rho x_1x_4(x_2^2 + x_1^2x_4^2)^{1/2} + g\sin x_3\right\}.$$

Thus we can write

$$\frac{dX}{dt} = F(X) = \begin{pmatrix} x_2 \\ x_1 x_4^2 - \frac{1}{2}\beta g \rho x_2 (x_2^2 + x_1^2 x_4^2)^{1/2} - g \cos x_3 \\ x_4 \\ \frac{1}{x_1}\left\{- 2x_2 x_4 - \frac{1}{2}\beta g \rho x_1 x_4 (x_2^2 + x_1^2 x_4^2)^{1/2} + g \sin x_3\right\} \end{pmatrix}. \tag{56}$$

Here ρ varies with altitude and for this application will be assumed to vary as

$$\rho = \rho_0 \exp(-\alpha x) = \rho_0 \exp(-\alpha r \cos\theta). \tag{57}$$

This uses Z as the effective altitude for estimating ρ but is acceptable, given the small ground ranges met in the current application. Equation (56) can then be rewritten as

$$\frac{dX}{dt} = F(X) = \begin{pmatrix} x_2 \\ x_1 x_4^2 - \frac{1}{2}\beta g \rho_0 x_2 (x_2^2 + x_1^2 x_4^2)^{1/2} \exp(-\alpha x_1 \cos x_3) \\ \qquad - g \cos x_3 \\ x_4 \\ \frac{1}{x_1}\left\{- 2x_2 x_4 - \frac{1}{2}\beta g \rho_0 x_1 x_4 (x_2^2 + x_1^2 x_4^2)^{1/2} \exp(-\alpha x_1 \cos x_3)\right. \\ \qquad \left. + g \sin x_3\right\} \end{pmatrix}. \tag{58}$$

Equations (55) and (58) describe the physics of the process. Values for β, and the time interval of observation, have been selected so as to result in highly nonlinear variations in position and velocity coordinates. This is illustrated in Fig. 2, which gives these coordin-

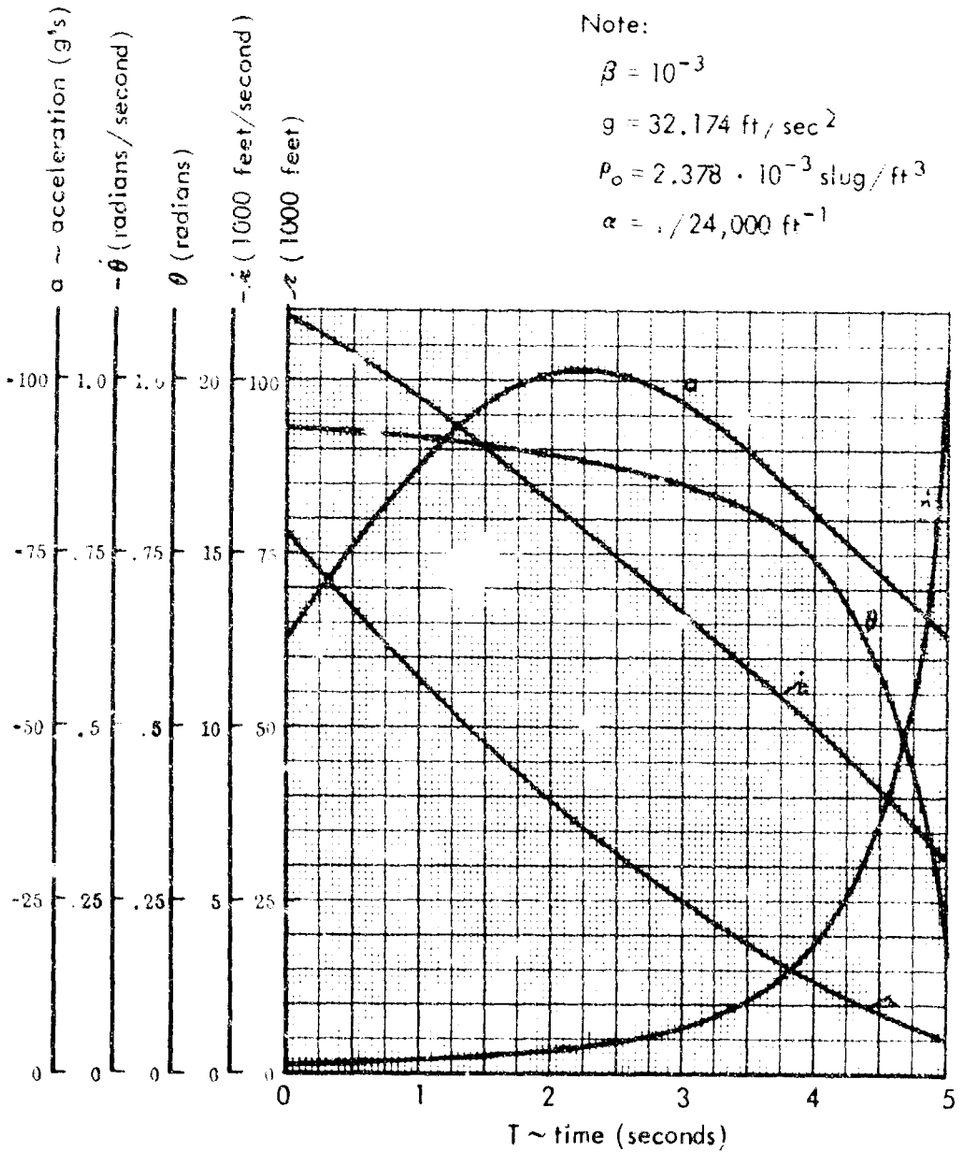

Fig. 2 — Reentry path used

ates as a function of time, as well as the values used for β, g, ρ_0, and γ. Also shown is the axial RV acceleration, to illustrate the highly nonlinear nature of the RV forces and motion. These results were computed using a Rand computer program called ROCKET.[19]

RADAR OBSERVABLES

The radar is assumed able to measure directly the RV coordinates $x_1 = r$, $x_2 = \dot{r}$, $x_3 = \theta$, with measurement being exact except for additive noise. Any multiplicative factors could have been inserted into the measurement of these three coordinates with no added complexity. This is equivalent to taking

$$H(s) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = H,$$

so that

$$Z(s) = H(s)X(s) + N(s) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1(s) \\ x_2(s) \\ x_3(s) \\ x_4(s) \end{pmatrix}$$

$$+ \begin{pmatrix} n_1(s) \\ n_2(s) \\ n_3(s) \end{pmatrix} = \begin{pmatrix} x_1(s) \\ x_2(s) \\ x_3(s) \end{pmatrix} + \begin{pmatrix} n_1(s) \\ n_2(s) \\ n_3(s) \end{pmatrix}.$$

COMPUTATIONAL APPROACH

Mixed Differential-Equation/Newtonian Approach

The objective of a computational scheme is to calculate the solution $\hat{X}(T)$ to (25) which minimizes (26):

$$\nabla_C f(\hat{X},T) = 2\Lambda^{-1}(\hat{X} - \mu) - 2\int_0^T \nabla_C X^*(s;\hat{X})H^*R^{-1}[Z(s) - HX(s;\hat{X})]ds = \emptyset, \tag{25}$$

$$f(C,T) = (C - \mu)^*\Lambda^{-1}(C - \mu) + \int_0^T [Z(s) - HX(s;C)]^*R^{-1}[Z(s) - HX(s;C)]ds. \tag{26}$$

Two applications are possible: real-time, on-line computations, and nonreal-time, off-line computations. Both applications can be approached via the preceding theoretical developments. The approach involves solving (27) for $\hat{X}(T)$ as the observation interval for $Z(s)$, $(0,T)$ grows with time T; the solution is initiated by using the initial conditions given by (28):

$$\frac{d\hat{X}}{dT} = 2[\nabla_{CC} f(\hat{X},T)]^{-1}\nabla_C X^*(T;\hat{X})H^*R^{-1}[Z(T) - HX(T;\hat{X})], \tag{27}$$

$$\hat{X}(0) = \mu; \qquad \left.\frac{d\hat{X}}{dT}\right|_{T=0} = \Lambda H^*R^{-1}[Z(0) - H\mu]. \tag{28}$$

Since the solution of (27) is to be done numerically, it really amounts to taking

$$\hat{X}(T + \Delta T) = \hat{X}(T) + \frac{d\hat{X}(T)}{dT} \Delta T, \tag{59}$$

using (28) as the starting condition; actually a suitable numerical integration technique is used to effect (59), specifically Runge-Kutta. The question of suitable ΔT size immediately arises. Depending upon the magnitude of the factors in (27), very small ΔT values may be required to keep the resulting $\hat{X}(T + \Delta T)$ from diverging from the desired exact solution to (27). When this situation applies, computing times can become so great as to be impossible, even for off-line applications.

Since we are estimating the initial condition $X(0)$, $\hat{X}(T)$ need not change rapidly with T; for example, it certainly need not change as per $F(X)$. It is reasonable, therefore, to consider weighting $\frac{d\hat{X}}{dT}$ in (59) so as to decrease its destabilizing effect for a given size ΔT:

$$\hat{X}(T + \Delta T) = \hat{X}(T) + \gamma \frac{d\hat{X}(T)}{dT} \Delta T, \tag{60}$$

where γ is a constant, $0 \leq \gamma \leq 1$.

Naturally, for finite ΔT and constant γ the value for $\hat{X}(T + \Delta T)$ resulting from (60) will differ from that value which truly minimizes (26). However, if it is close enough to the correct value, then steepest-descent or Newtonian techniques can be used to improve the estimate for constant T. For example, if Newton's method is applied, then a sequence of improving estimates for $\hat{X}(T)$ results from taking

$$\hat{X}^{n+1}(T) = \hat{X}^{n}(T) - [\nabla_{CC} f(\hat{X}^{n}[T],T)]^{-1} \nabla_{C} f(\hat{X}^{n}[T],T). \tag{61}$$

It is immediately clear that several exchanges can be made between these two techniques, in order to gain either computational accuracy

or efficiency, or both. Note that even for $\gamma = 0$, the application of (60) and (61) affords an approach as near as desired to the correct $\hat{X}(T)$, provided the conditions for Newtonian convergence are met. These conditions are (see Ref. 20, p. 63)

If the left member of

$$\nabla_C f(C,T) = \begin{pmatrix} \frac{\partial f(C,T)}{\partial c_1} \\ \vdots \\ \frac{\partial f(C,T)}{\partial c_m} \end{pmatrix} = \emptyset$$

satisfies

1. $\left|\frac{\partial f(\hat{X}^0,T)}{\partial c_j}\right| \le d_1$, $j = 1, \ldots, m$,

2. The matrix $\nabla_{CC} f(\hat{X}^0,T)$ has a nonvanishing determinant D (with absolute value $|D|$) and, for the absolute value $|A_{ij}|$ of its cofactors A_{ij},

$$\max_i \frac{1}{|D|} \sum_{j=1}^{n} |A_{ij}| \le d_2,$$

3. The elements of $\nabla_{CCC} f(\hat{X},T)$, $s_{ijk} = \frac{\partial^3 f(\hat{X},T)}{\partial c_i \partial c_j \partial c_k}$, are all bounded:

$$|s_{ijk}| < d_3$$

in the region

4. $\max |\hat{x}_i - \hat{x}_i^0| \le \frac{1 - \sqrt{1 - 2b_0}}{b_0} d_1 d_2$,

where $\hat{X} = \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_m \end{pmatrix}$ and

b_0 is defined by and satisfies

$$b_0 = d_2^2 d_1 d_3 m^2 \leq 1/2,$$

then the system $\nabla_C f(\hat{X},T) = \emptyset$ has a solution which can be obtained by Newton's method (61).

Joint application of (60) and (61) would then proceed as follows. At $T = 0$, take $\hat{X}(0) = \mu$; compute $\left.\frac{d\hat{X}}{dT}\right|_{T=0}$ from (28) and apply (60) to estimate

$$\hat{X}^0(\Delta T_1) = \mu + \gamma\Lambda H^* R^{-1}[Z(0) - H\mu]\Delta T_1.$$

Then apply (61) until $\hat{X}^n(\Delta T_1)$ satisfies certain conditions, such as all coordinates of $\nabla_C f(\hat{X}^n[\Delta T_1], \Delta T_1)$ being less than some constant, or all coordinates of $[\hat{X}^n(\Delta T_1) - \hat{X}^{n-1}(\Delta T_1)]/||\hat{X}^{n-1}(\Delta T_1)||$ being less than some constant. Taking this last estimate for $\hat{X}(\Delta T_1)$, then apply (60) to compute $\hat{X}^0(\Delta T_1 + \Delta T_2)$ and so forth until $\sum_i \Delta T_i = T$, the upper limit of the observation interval. We will refer to each application of (61) as a Newtonian iteration. Application of either (60) or (61) involves numerical calculation of the various factors occurring therein, and as these equations suggest, involve essentially the same amount of computing time. Assuming this to be case, it is possible to make a rough estimate of the alterations in computing time possible via (60) and (61), as follows.

Computation Time

In general, the integration interval Δs used in the numerical calculation of the factors of (60) and (61) need not equal ΔT; in general it will be preferable not to have $\Delta s = \Delta T$, in order to permit sufficient data density over $(0,T)$ even though ΔT is fairly large. For reasonably large T, the calculation time is dominated by the per Δs integrations.

Let

ΔT = step size at which update $X(T)$ in time T

Δs = time interval used in numerical integration for calculating the factors in (60) and (61)

$\upsilon = \Delta T/\Delta s$

δ = computer execution time per Δs calculation, assumed the same in application of either (60) or (61)

η_i = number of Newtonian iterations at time T_i

η = average number of Newtonian iterations per time T_i

T = total time of observation to date

τ_i = computer execution time at time T_i

τ = total computer execution time for processing signal over $(0,T)$

$N = T/\Delta T$ = number of updatings

At time T_i, the computer must process numerical integrations across $\frac{T_i}{\Delta s}$ intervals in each calculation of the factors of (61), plus once again for the updating via (60). Thus we can write

$$\tau_i = \frac{\delta(\eta_i + 1)T_i}{\Delta s} = \frac{\delta(\eta_i + 1)}{\Delta s} i\Delta T = i\delta(\eta_i + 1)\frac{\Delta T}{\Delta s},$$

or

$$\tau_i = i\delta(\eta_i + 1)\nu. \tag{62}$$

Then

$$\tau = \sum_{i=1}^{N} \tau_i = \nu\delta \sum_{i=1}^{N} i(\eta_i + 1) \doteq \nu\delta(\eta + 1) \sum_{i=1}^{N} i \quad \frac{\nu\delta(\eta + 1)N(N + 1)}{2},$$

so that from $N = \frac{T}{\Delta T}$ and $N + 1 \doteq N$ there follows

$$\tau \doteq \frac{\nu\delta(\eta + 1)}{2} \left(\frac{T}{\Delta T}\right)^2 = \frac{\delta(\eta + 1)}{2} \frac{T^2}{\Delta s \Delta T}. \tag{63}$$

It is the variation of total computer execution time with $\frac{1}{\Delta s \Delta T}$ that forces the use of rather large Δs values, as well as the use of rather large ΔT.

Even with only a four-coordinate vector $X(s)$, the number of equations which must be integrated numerically to apply (60) and (61) gets excessive unless the computer is of special design. The computers available are far from specially designed, naturally, and further, already-programmed numerical procedures have been used whenever possible to reduce programming effort. The result is that the procedure used here for computation is very inefficient, timewise, taking about one second per complete Δs calculation (integration)--i.e., $\delta \doteq 1$ second. The resulting effect upon τ, given by (63), is plotted in Fig. 3 as a function of T for $\frac{\delta(\eta + 1)}{\Delta s \Delta T} = 10^4$. Since τ scales linearly with $\frac{\delta(\eta + 1)}{\Delta s \Delta T}$, Fig. 3 can easily be read for values other than 10^4 for this factor. Except for rather small values of T, Fig. 3 suggests $\hat{X}(T)$ calculation times two or three orders of magnitude greater than the real-time observation being processed. Since it appears that such

factors can be recovered by specially designed computers, and since we are here principally concerned with the potential capabilities of the proposed technique, this is an acceptable limitation.

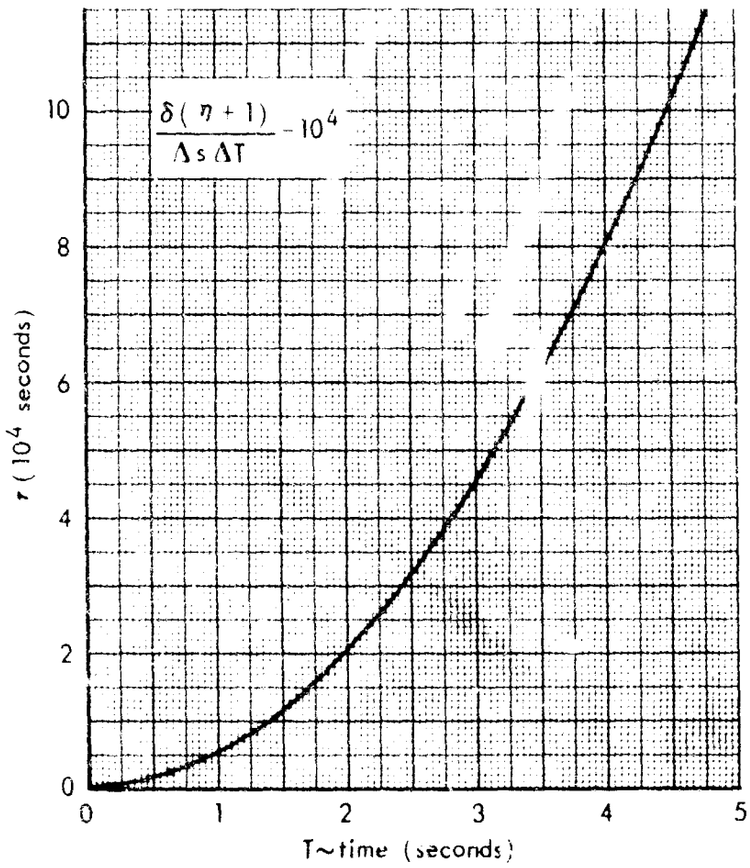

Fig.3—Computing-time variation

Integration Routines

It remains in this section (Computational Approach) to describe the specific equations which were numerically solved in the application of (60) and (61).

Given an $\hat{X}(T)$, to apply (60) to the calculation of $\hat{X}(T + \Delta T)$ requires the calculation of the factors in (27), that is, $\nabla_{CC} f(\hat{X},T)^{-1}$, $\nabla_C X(T;\hat{X})$, $X(T;\hat{X})$, and the performance of manipulations indicated in (60). First consider $X(T;\hat{X})$. A basic given in the problem statement is the differential equation

$$\frac{dX}{dT} = F(X). \tag{64}$$

The Runge-Kutta integration routine in the Rand program library has been directly utilized to compute $X(T;\hat{X})$ via (64). Directly from (64),

$$X(T;\hat{X}) = \int_0^T \frac{dX(s;\hat{X})}{ds}\, ds + \hat{X} = \int_0^T F[X(s;\hat{X})]ds + \hat{X}, \tag{65}$$

so that

$$\frac{\partial X(T;\hat{X})}{\partial c_k} = \int_0^T \frac{\partial F[X(s;\hat{X})]}{\partial c_k}\, ds + I_k = \int_0^T \nabla_{X(s;\hat{X})} F[X(s;\hat{X})] \frac{\partial X(s;\hat{X})}{\partial c_k}\, ds + I_k, \tag{66a}$$

where

$$\hat{X} = \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_m \end{pmatrix} \qquad C = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} \qquad I_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ kth place.} \tag{66b}$$

Equation (66a) may be rewritten as

$$\nabla_C X(T;\hat{X}) = \int_0^T \nabla_{X(s;\hat{X})} F[X(s;\hat{X})] \nabla_C X(s;\hat{X})\, ds + I, \tag{67}$$

where I is the mXm identity matrix.

Equation (67) can then be written as

$$\frac{\partial \nabla_C X(T;\hat{X})}{\partial T} = \nabla_{X(T;\hat{X})} F[X(T;\hat{X})] \nabla_C X(T;\hat{X}), \tag{68}$$

and the same Runge-Kutta integration can be used to calculate $\nabla_C X(T;\hat{X})$ via (68), once $X(T;\hat{X})$ is known via (64), using as the initial conditions

$$\nabla_C X(0,\hat{X}) = I, \qquad F[X(0;\hat{X})] = F(\hat{X}). \tag{69}$$

From (19), $\dfrac{\partial \nabla_{CC} f(\hat{X},T)}{\partial T}$ may be written as

$$\frac{\partial \nabla_{CC} f(\hat{X},T)}{\partial T} = 2\nabla_C X^*(T;\hat{X}) H^* R^{-1} H \nabla_C X(T;\hat{X})$$

$$- 2\nabla_{CC} X^*(T;\hat{X}) H^* R^{-1} [Z(T) - HX(T;\hat{X})], \tag{70a}$$

with

$$\nabla_{CC} f(\hat{X},0) = 2A^{-1}. \tag{70b}$$

The vectors $X(T;\hat{X})$ and $\nabla_C X(T;\hat{X})$ are available from (64) and (68)-(69); similarly, from (67),

$$\frac{\partial \nabla_C X(T;\hat{X})}{\partial c_k} = \int_0^T \frac{\partial [\nabla_{X(s;\hat{X})} F[X(s;\hat{X})] \nabla_C X(s;\hat{X})]}{\partial c_k}\, ds$$

$$= \int_0^T \left\{ \frac{\partial \left[\nabla_{X(s;\hat{X})} F^T X(s;\hat{X}) \right]}{\partial c_k} \nabla_C X(s;\hat{X}) \right.$$

$$\left. + \nabla_{X(s;\hat{X})} F^T X(s;\hat{X}) \frac{\partial \nabla_C X(s;\hat{X})}{\partial c_k} \right\} ds$$

$$= \int_0^T \left\{ \nabla_{X(s;\hat{X})X(s;\hat{X})} F^T X(s;\hat{X}) \frac{\partial X(s;\hat{X})}{\partial c_k} \nabla_C X(s;\hat{X}) \right.$$

$$\left. + \nabla_{X(s;\hat{X})} F^T X(s;\hat{X}) \frac{\partial \nabla_C X(s;\hat{X})}{\partial c_k} \right\} ds,$$

which is equivalent to

$$\nabla_{CC} X(T;\hat{X}) = \int_0^T \nabla_{X(s;\hat{X})X(s;\hat{X})} F^T X(s;\hat{X}) \nabla_C X(s;\hat{X}) \nabla_C X(s;\hat{X}) ds$$

$$+ \int_0^T \nabla_{X(s;\hat{X})} F^T X(s;\hat{X}) \nabla_{CC} X(s;\hat{X}) ds,$$

which can be rewritten as

$$\frac{\partial \nabla_{CC} X(T;\hat{X})}{\partial T} = \nabla_{X(T;\hat{X})X(T;\hat{X})} F^T X(T;\hat{X}) \nabla_C X(T;\hat{X}) \nabla_C X(T;\hat{X})$$

$$+ \nabla_{X(T;\hat{X})} F^T X(T;\hat{X}) \nabla_{CC} X(T;\hat{X}), \tag{71a}$$

where

$$\nabla_{CC} X(0;\hat{X}) = 0. \tag{71b}$$

Again, the same Runge-Kutta techniques can be used to solve (71), given outputs from similar Runge-Kutta applications to (64) and (68)

for $X(T;\hat{X})$ and $\nabla_C X(T;\hat{X})$. All the inputs for solving (70) are then available, which can also be solved via the same Runge-Kutta technique. $\nabla_{CC} f(\hat{X},T)$ is then inverted, and $\frac{d\hat{X}}{dT}$ formed via the multiplications shown in (27). We are, in effect, solving simultaneously via Runge-Kutta techniques the equations (64), (68), (70), and (71). This involves $m + m^2 + m^3 + m^2 = m + 2m^2 + m^3$ simultaneous equations; for $m = 4$ this total is 100. Given this solution for a given T, the same Runge-Kutta technique then used to update $\hat{X}(T)$ to $\hat{X}(T + \Delta T)$ as per (60).

In addition to the preceding, it is necessary to calculate $\nabla_C f(\hat{X},T)$ in order to apply (61). From (25),

$$\frac{\partial \nabla_C f(\hat{X},T)}{\partial T} = -2\nabla_C X^*(T;\hat{X})H^* R^{-1}[Z(T) - HX(T;\hat{X})], \tag{72}$$

which can be solved immediately with the preceding set, leading to a total of 104 simultaneous equations. This explains the large computer execution times per Δs.

The actual implementation of the preceding is sketched in the program flow diagram in Appendix B.

COMPUTATIONAL RESULTS

The objectives of the calculations performed are several. The first is simply to show that the maximum-likelihood estimator derived above can be numerically calculated. Another aim is to show the rate of its convergence and the usefulness of the limit to which the MLE tends. A third goal is to illustrate some of the computational choices which must be made (e.g., magnitudes of ΔT and Δs) and their

effects. Finally, it is desired to display relative to the computed solutions some numerically derived values for the Cramer-Rao bound.

Computations

The nonlinear and complicated nature of F(X) makes it necessary to calculate $X(s;X_0)$ for any given X_0 numerically. Although this can be done quite easily with the ROCKET[19] program, it is relatively expensive in computer time. In part for this reason, the computations to be described were made for only one X_0, that corresponding to Fig. 2:

$$X_0 = \begin{pmatrix} 77929 \text{ ft} \\ -\ 21836 \text{ ft/sec} \\ .83200 \text{ rad} \\ -.011849 \text{ rad/sec} \end{pmatrix}.$$

Variations in X_0 relative to μ, the expected value of X_0, were accomplished by changing μ. Two such variations were examined, one corresponding roughly to $X_0 - \mu$ coordinates of two, twenty, twenty, and two times their standard deviations (about their μ values, respectively); and another corresponding to $X_0 - \mu$ coordinates of three times the above differences. The smaller excursion was used as the principal case and corresponds to an excursion which should not usually be exceeded; the relative magnitudes were chosen to test for convergence by coordinate. The more extreme excursion was used simply to test initiation and convergence for an extreme deviation from the initial value of the estimator at T = 0: μ.

The trajectory corresponding to the above X_0 was calculated at 10^{-3}-second steps over a 5-second interval preceding (and near) RV impact. This served as the basic trajectory. For each noise matrix R used, a random sequence of noise inputs was generated at 10^{-3}-second steps, and a sample observation waveform $Z(s)$, $0 \le s \le 5$ seconds, calculated from

$$Z(s) = HX(s;X_0) + N(s).$$

Given the random (noise) component in $Z(s)$, it is impossible to store $Z(s)$ over all points of a time continuum in closed form; but $\Delta s = 10^{-3}$ second was an adequate density for our purposes.

Only two noise matrices (R) were used, one representing a nominal radar performance, the other an extremely accurate performance to test the impact of R upon estimator initiation and convergence. The former (nominal) R was used as the principal case. Those used were, the first being the principal case.

$$R = \begin{pmatrix} 10^4 \text{ft}^2 & & 0 \\ & 10^2 (\text{ft/sec})^2 & \\ 0 & & 3\text{X}10^{-4} \text{rad}^2 \end{pmatrix},$$

$$R = \begin{pmatrix} 10^2 \text{ft}^2 & & 0 \\ & 1(\text{ft/sec})^2 & \\ 0 & & 3.04\text{X}10^{-6} \text{rad}^2 \end{pmatrix}.$$

It is useful to express the last component of R in square feet, to make

it visibly comparable to the first two components. It becomes at X_0

$$R = \begin{pmatrix} 10^4 ft^2 & & \bigcirc \\ & 10^2 (ft/sec)^2 & \\ \bigcirc & & 1.82X10^6 ft^2 \end{pmatrix},$$

$$R = \begin{pmatrix} 10^2 ft^2 & & \bigcirc \\ & 1(ft/sec)^2 & \\ \bigcirc & & 1.85X10^4 ft^2 \end{pmatrix}.$$

With R in this form it is more clear that the angular accuracy ascribed to the radar is not out of line with the range accuracy.

Given the preceding selections, there remains only Λ unspecified. Its value principally has two effects. First, its magnitude coupled with that of R influences (in some complex way) the feasibility of numerically inverting $\nabla_{CC} f(\hat{X},T)$--and therefore that the basic differential equation (27) will be applicable and soluble for all T of interest. This interaction between Λ and R was explored in a very limited way by varying R as per above. Second, the magnitudes of Λ and R help determine what values of ΔT (the step size used in the numerical updating in time of the estimator) can be used. That this is the case can be seen by examining the expression (28) for $\frac{d\hat{X}}{dT}$ at T = 0:

$$\frac{d\hat{X}}{dT} = \Lambda H^* R^{-1} [Z(0) - H\mu]. \qquad (28')$$

The vector $\frac{d\hat{X}}{dT}$ gives the <u>direction</u> in which $\hat{X}(T)$ should be changed; however, use of too large a value for ΔT in (60),

$$\hat{X}(T + \Delta T) = \hat{X}(T) + \gamma \frac{d\hat{X}}{dT} \Delta T, \tag{60}$$

will cause divergence from the required solution. It is clear from (28') that large Λ tends to emphasize the influence of the radar measurements on $\frac{d\hat{X}}{dT}$ and large R to emphasize the a priori statistics, as should be the case. Altering ΔT does not alter the direction of change given by $\frac{d\hat{X}}{dT}$; it only alters the magnitude of the change in that direction.

As explained earlier, the computing routine used here employs as many existing programs as possible and is therefore expensive in computer time; therefore, relatively large ΔT values are preferred--in the range 10^{-3} to 10^{-1} seconds. These factors constrained the computations to the use of just one Λ and the use of various values of ΔT in order to examine its influence upon initiation and convergence. The Λ used was

$$\Lambda = \begin{pmatrix} 10^6 \text{ft}^2 & & & \bigcirc \\ & 10^2 (\text{ft/sec})^2 & & \\ & & 3\text{X}10^{-4} \text{rad}^2 & \\ \bigcirc & & & 10^{-6} (\text{rad/sec})^2 \end{pmatrix}.$$

It is of interest to express all coordinates of Λ in feet and display $\Lambda^{1/2}$, in order to indicate the comparability of all the error components and to make clearer the a priori error magnitudes involved:

$$\Lambda^{1/2} = \begin{pmatrix} 10^3 & & & \bigcirc \\ & 10\text{ft/sec} & & \\ & & 1.34\text{X}10^3\text{ft} & \\ \bigcirc & & & 78\ \text{ft/sec} \end{pmatrix}.$$

There remains to be assumed the step size, Δs, to be used inside the integration routines for fixed T--e.g., in the calculation of $X(T;X_0)$. In these calculations, this has been taken equal to ΔT and thus has fallen in the range of 10^{-3} to 10^{-1} seconds and was varied with ΔT to examine the effect upon estimator performance. The principal effect that Δs size has is to introduce a granular structure into the surface $f(C,T)$, the minimum of which is being estimated. Clearly, the accuracy of estimation possible will be limited by this granularity and therefore by Δs.

Computational results will be shown in terms of the absolute error between individual components of the estimator and the actual initial condition X_0; also shown will be the same errors transformed to the equivalent absolute errors in distance and in velocity. For

$$X_0 = \begin{pmatrix} x_{10} \\ x_{20} \\ x_{30} \\ x_{40} \end{pmatrix} = \begin{pmatrix} r_0 \\ \dot{r}_0 \\ \theta_0 \\ \dot{\theta}_0 \end{pmatrix},$$

$$\hat{X} = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \\ \hat{x}_4 \end{pmatrix} = \begin{pmatrix} \hat{r}_0 \\ \hat{\dot{r}}_0 \\ \hat{\theta}_0 \\ \hat{\dot{\theta}}_0 \end{pmatrix},$$

we define the errors of interest as

$$\epsilon_r = |r_0 - \hat{r}_0|$$

$$\epsilon_{\dot{r}} = |\dot{r}_0 - \hat{\dot{r}}_0|$$

$$\epsilon_\theta = |\theta_0 - \hat{\theta}_0|$$

$$\epsilon_{\dot{\theta}} = |\dot{\theta}_0 - \hat{\dot{\theta}}_0|$$

$$\epsilon_D = [\epsilon_r^2 + r_0^2\epsilon_\theta^2]^{1/2}$$

$$\epsilon_V = [\epsilon_{\dot{r}}^2 + r_0^2\epsilon_{\dot{\theta}}^2]^{1/2},$$

where ϵ_D and ϵ_V are the errors in distance and velocity, respectively.

Pure Differential-Equation Solution

In Figs. 4 and 5 are shown these errors as a function of time T for several fixed values of $\Delta s = \Delta T$ (10^{-1}, 10^{-2}, and 10^{-3} seconds). In these estimates $\gamma = 1$ and no Newtonian improvements to the pure differential-equation solution were employed. The different ranges of T over which the errors are shown (for different Δs) result because of the greater computing time required for the smaller Δs values. The intent is to show enough T range collectively to indicate the performance of the estimator.

Figure 4a indicates that for $\Delta s = 10^{-1}$ seconds $\hat{r}_0$ is either diverging or tending to a limit some distance from r_0. Three things are probably occurring. First, the granularity introduced into the surface $f(\hat{X},T)$ by a finite sampling interval $\Delta s = 10^{-1}$ seconds is probably such that great accuracy is not possible. This is borne out later by the differential-equation/Newtonian calculation, which

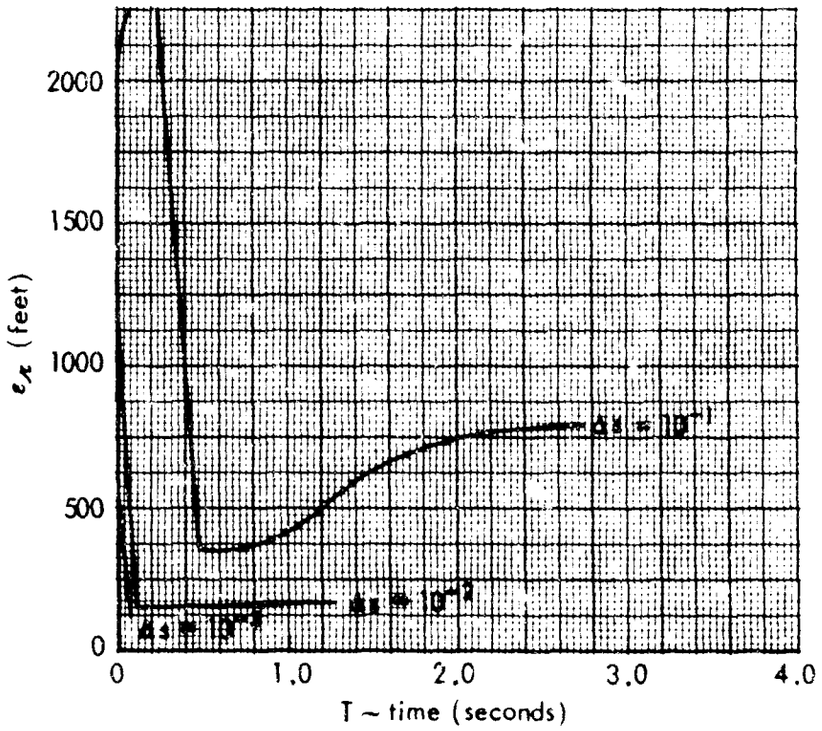

Fig.4a—Range error for pure differential-equation solution, for various Δs and varying T

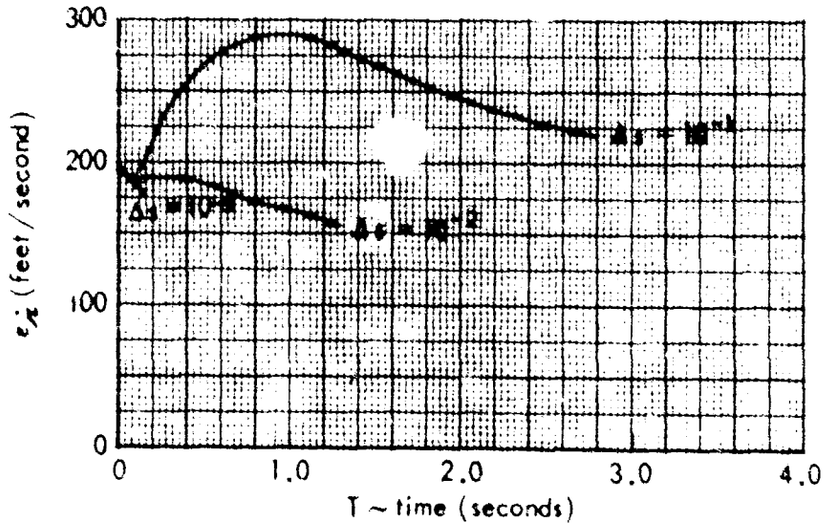

Fig.4b—Range-rate error for pure differential-equation solution for various Δs and varying T

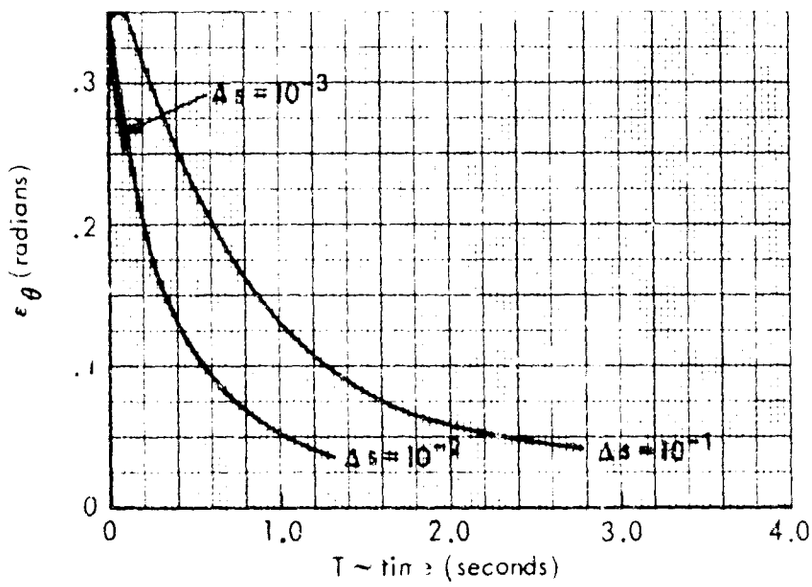

Fig.4c—Angle error for pure differential-equation solution, for various Δs and varying T

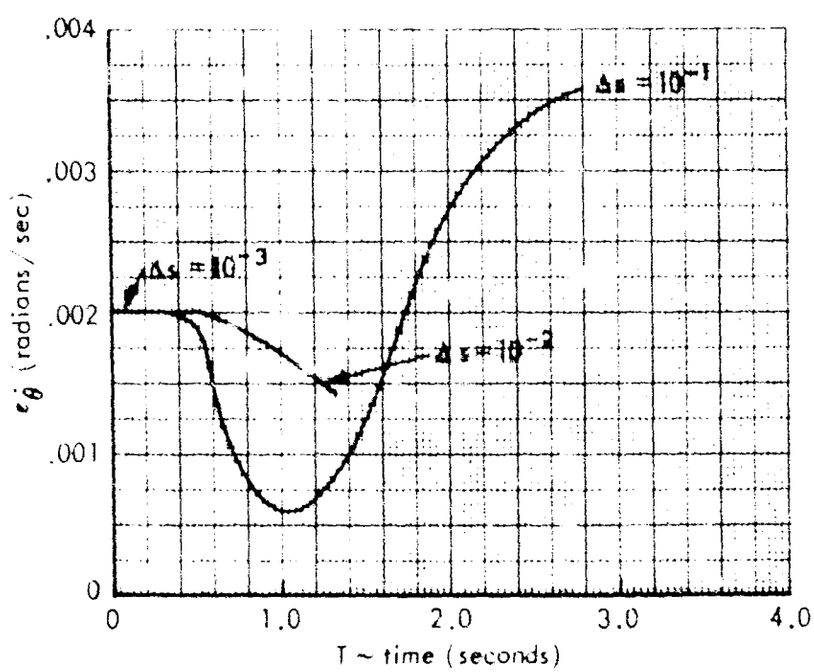

Fig.4d—Angle-rate error for pure differential-equation solution, for various Δs and varying T

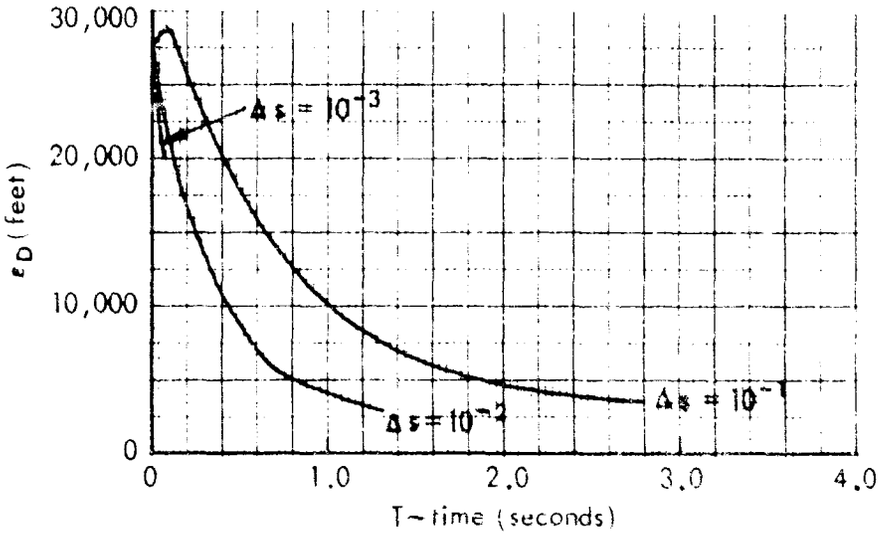

Fig. 5a — Miss-distance error for pure differential-equation solution, for various Δs and varying T

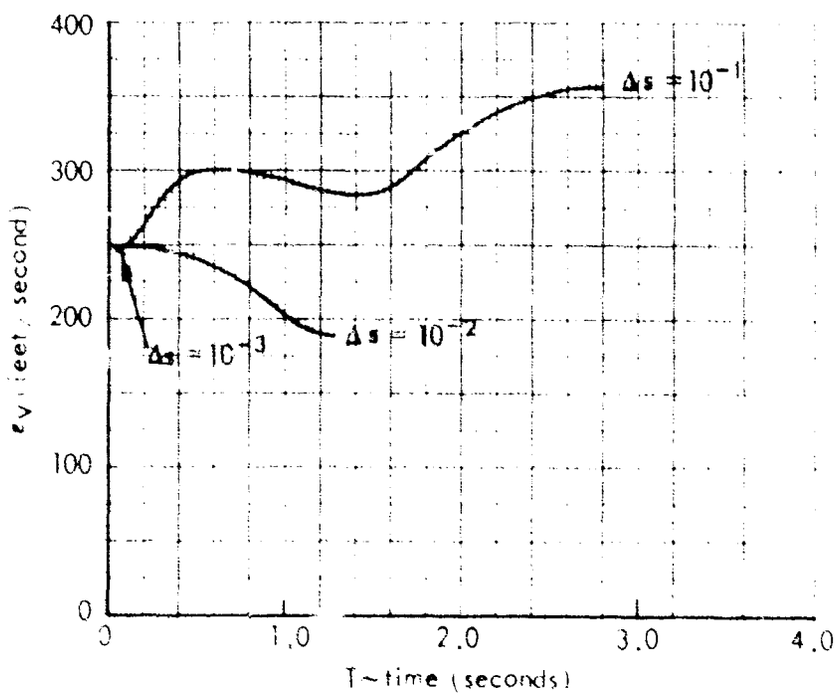

Fig. 5b — Velocity error for pure differential-equation solution, for various Δs and varying T

apparently has an ϵ_r asymptote at about 1500 feet which should depend mainly upon this granularity. Second, $\Delta T = 10^{-1}$ seconds is probably just too large, with resulting divergence from the desired solution. Third, as will be shown below, this estimator will sacrifice performance in one coordinate in order to improve overall performance, and this may be occurring. Similar remarks apply to Fig. 4b for $\Delta s = 10^{-1}$ seconds, which displays $\epsilon_{\dot{r}}$. In Fig. 4c, it appears that $\hat{\theta}_0$ is converging usefully. Figure 4d, however, indicates that for $\Delta s = 10^{-1}$ seconds $\hat{\dot{\theta}}$ is diverging rapidly. In Fig. 5 are shown the errors ϵ_D and ϵ_V. It appears that ϵ_D is converging to zero; since ϵ_D is a combination of ϵ_r and ϵ_θ, it would appear that the behavior of $\hat{r}_0$ depicted for $\Delta s = 10^{-1}$ seconds in Fig. 4a was in part an attempt by the estimator to produce a better overall miss-distance result. Similar remarks apply to ϵ_V for $\Delta s = 10^{-1}$ seconds, presented in Fig. 5b. However, it is clear from Fig. 5b that $\Delta T = \Delta s = 10^{-1}$ seconds is too large for convergence.

Figures 4 and 5 indicate that Δs values of 10^{-2} and 10^{-3} seconds provide convergence. The slight divergence of ϵ_r for $\Delta s = 10^{-2}$ seconds (Fig. 4a) is shown in Fig. 5 to be the result of an overall improvement in ϵ_D and ϵ_V, as discussed above. However, two results shown in Figs. 4 and 5 deserve special discussion.

First, although ϵ_{r_0}, ϵ_{θ_0}, and $\epsilon_{\dot{\theta}_0}$ all appear to converge rapidly to limits near zero $\epsilon_{\dot{r}_0}$ does not (Fig. 4b); similarly for the rapid convergence of ϵ_D toward zero, but not ϵ_V. Near T = 0 this is due in part to the relative emphasis placed upon the radar measurements by the elements of Λ and R and can best be visualized by examining $\frac{d\hat{X}}{dT}$ at

$T = 0$ given by (28). Given the diagonal nature of Λ, R and H, $\left.\frac{d\hat{X}}{dT}\right|_{T=0}$ takes the form, substituting in the values of $\{a_{ii}\}$ and $\{\phi_i\}$ from Λ and (nominal) R,

$$\left.\frac{d\hat{X}}{dT}\right|_{T=0} = \begin{pmatrix} 100[z_1(0) - \mu_1] \\ z_2(0) - \mu_2 \\ z_3(0) - \mu_3 \\ 0 \end{pmatrix}.$$

This value of 0 in the fourth coordinate will change, of course, as T increases. It is apparent from this expression that the r_0 estimate will emphasize the radar measurements much more than will the estimate of $\dot{r}$ in the vicinity of $T = 0$. The relative performance of the two estimates shown in Figs. 4a and 4b near $T = 0$ is therefore to be expected, and similarly for Figs. 4c and 4d, inasmuch as the radar errors are rather small. The degree of radar-measurement emphasis away from $T = 0$ will then depend upon the growth of the product of $[\nabla_{CC} f(\hat{X},T)]^{-1}$ and $\nabla_C X^*(T;\hat{X})$ (see (27)). It would appear that such growth is not as favorable to convergence of $\hat{f}_0$ as for the other estimates. Naturally, the poor convergence of ϵ_V in Fig. 5b results because ϵ_V contains ϵ_f. Is the poor convergence shown by ϵ_f and ϵ_V worth the effort? The answer lies in two quarters. First, if the initial error (between μ_2 and f_0) had been larger, then convergence to the shown level would have been more impressive; this is discussed later relative to such large excursions (see Fig. 8). The second aspect deals with the other result of Figs. 4 and 5 deserving greater discussion: estimation of $\dot{\theta}_0$ shown in Fig. 4d.

In the present application the variable $\dot{\theta}$ cannot be measured directly by the radar and so must be estimated entirely in terms of the other observables. Therefore, $\dot{\theta}_0$ can be considered an unknown parameter (which has an a priori statistical distribution). It is seen that $\hat{\dot{\theta}}_0$ converges for appropriate Δs values; ϵ_V, which includes $\epsilon_{\dot{\theta}}$, also behaves appropriately with increasing T. This is discussed in greater detail below, under the differential-equation/Newtonian solution.

Differential-Equation/Newtonian Solution

If Δs were infinitesimal, then use of the Newtonian technique to improve the pure differential-equation solution for finite ΔT should result in an exact solution corresponding to infinitesimal ΔT (given convergence of the Newtonian iterations and within the limits of computational accuracy). Since Δs is finite in these calculations, only smaller improvement relative to this theoretical maximum can be expected. In the results that follow, once convergence for the differential-equation/Newtonian estimator was indicated, values for small Δs (10^{-3} seconds) and large T were calculated by setting $\frac{d\hat{X}}{dT} = \emptyset$ and using just the Newtonian iterations. That is, for such computer-expensive cases $\gamma = 0$ was used in (60). Alternatively, large ΔT was used to get an approximate starting solution for large T, at which T smaller Δs was then used in the Newtonian iterations to estimate the solution for $\Delta T = \Delta s$.

Results for the joint routine are displayed in Figs. 6 and 7. It is seen that again for $\Delta s = 10^{-1}$ seconds (as discussed earlier) ϵ_r approaches a limit considerably greater than zero. All the other

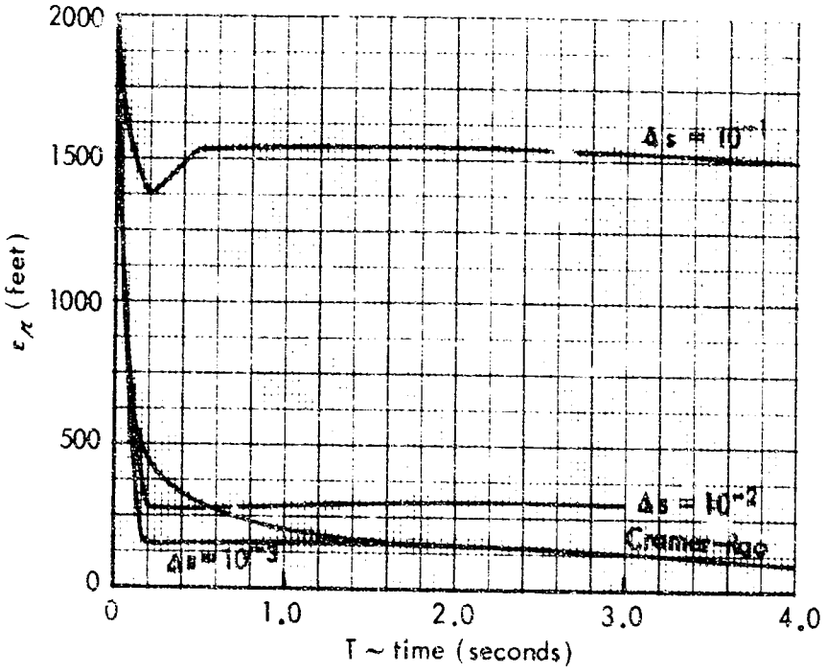

Fig. 6a — Range error for differential-equation/Newtonian solution, and Cramer-Rao term, for various Δs and varying T

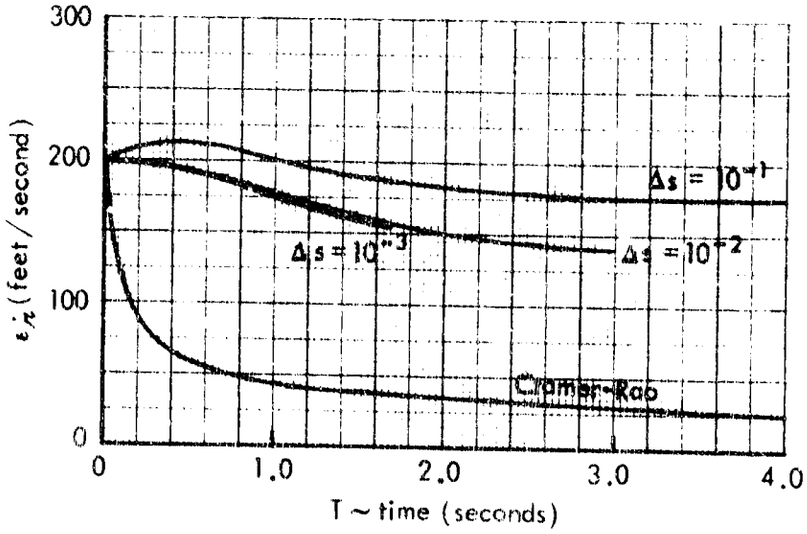

Fig. 6b — Range-rate error for differential-equation/Newtonian solution, and Cramer-Rao term, for various Δs and varying T

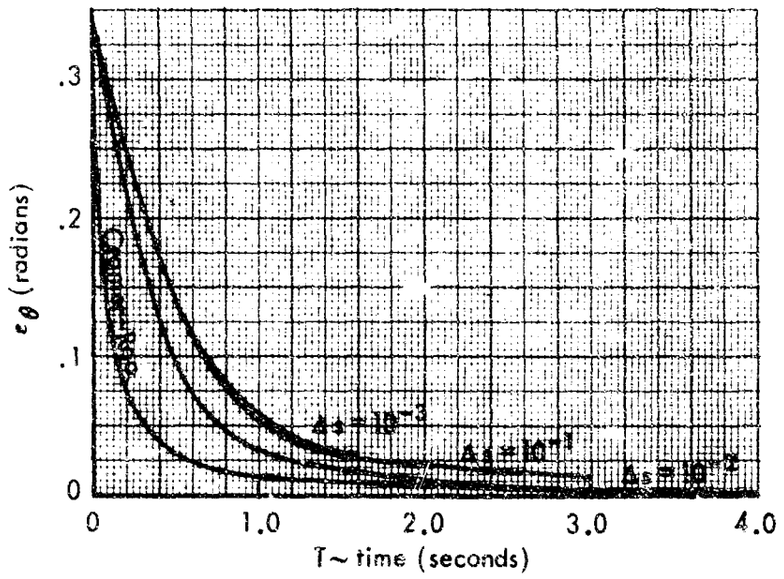

Fig.6c —Angle error for differential-equation/Newtonian solution, and Cramer-Rao term, for various Δs and varying T

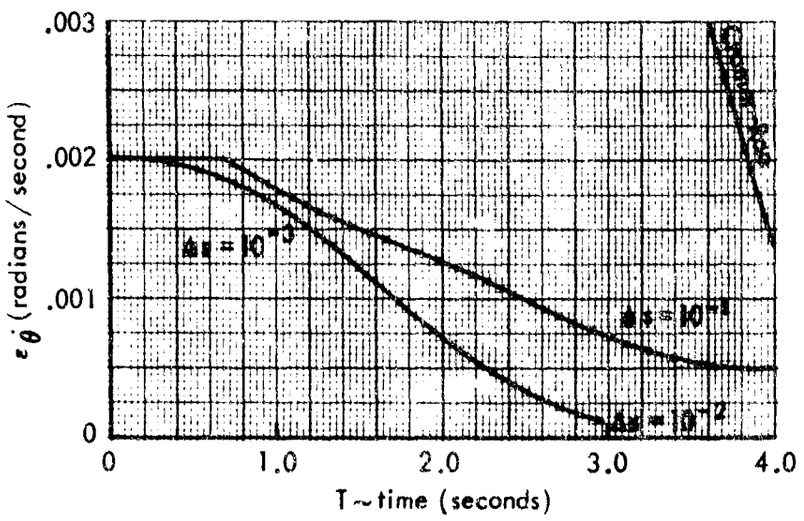

Fig.6d —Angle-rate error for differential-equation/Newtonian solution, and Cramer-Rao term, for various Δs and varying T

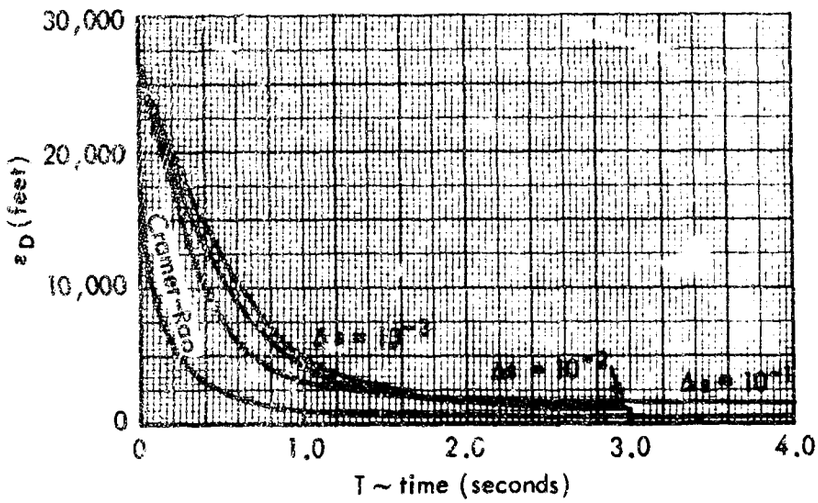

Fig.7a—Miss-distance error for differential-equation/Newtonian solution, and Cramer-Rao terms, for various Δs and varying T

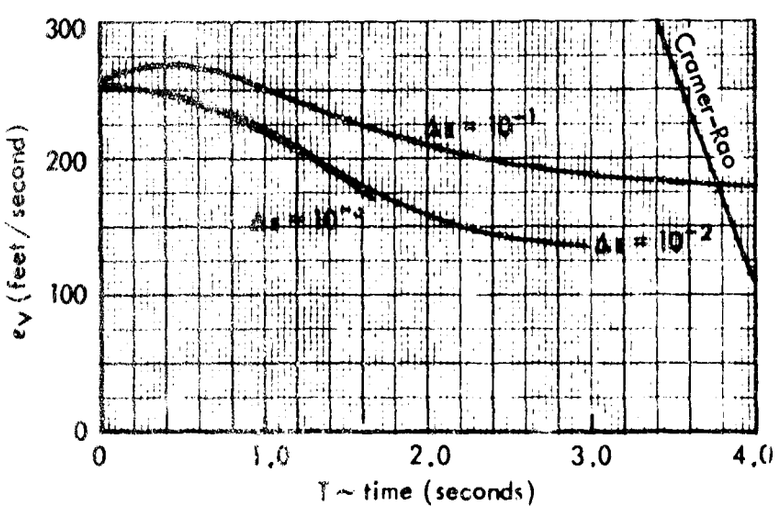

Fig.7b—Velocity error for differential-equation/Newtonian solution, and Cramer-Rao terms, for various Δs and varying T

components converge even at $\Delta s = 10^{-1}$ seconds, and ϵ_D and ϵ_V converge for all values of Δs employed (10^{-1}, 10^{-2}, and 10^{-3} seconds). The apparent slight divergence of ϵ_r and $\epsilon_{\dot{r}}$ after their initial convergence again results from an overall improvement in the estimation, as is borne out by reference to ϵ_D and ϵ_V in Fig. 7. Again, $\epsilon_{\dot{r}}$ and ϵ_V do not converge toward zero as rapidly as the other errors.

Also shown in Figs. 6 and 7 are the corresponding unbiased Crame Rao lower bounds on the conditional RMS error. It is clear that the estimator converges with increasing T to a solution and that the solution is a useful <u>overall</u> estimate of X_0 of the order of the Cramer-Rao bound. In particular, the estimation of $\dot{\theta}_0$, for which there are no direct radar observables, is good. It is this excellent estimate of $\dot{\theta}_0$ (and of θ_0) which makes these results useful even though $\dot{r}_0$ is not handled as well. For unless all coordinates are estimated reasonably well, such accuracy in particular ones is impossible. But the estimator has exchanged accuracy between coordinates, thereby reducing $\epsilon_{\dot{\theta}}$, for which there is no direct measurement, to less than 10^{-4} radians/second, or less than 5 percent of its original magnitude; ϵ_θ is reduced from 0.347 radians to less than 10^{-3} radians, or less than 0.3 percent of its original magnitude; and ϵ_r goes from 2000 feet to less than 200 feet, or less than 10 percent of its original size. It is this overall estimate of X_0, and especially of the coordinate $\dot{\theta}_0$, for which no direct radar measurements are available, that makes these results of interest.

Extreme Excursion of $(X_0 - \mu)$

The above example used a value of X_0 differing from its expected value, μ, by about two, twenty, twenty, and two standard deviations in its four coordinates, respectively. To explore the initial convergence of the estimator, a μ differing from X_0 in each coordinate by three times the above differences was employed. Convergence of the differential-equation/Newtonian solution for this case is shown in Fig. 8, where it is seen that by time T = 0.25 seconds the solution has converged to that corresponding to the smaller-deviation case. At T = 3 seconds these solutions correspond to errors in distance (D) and velocity (V) which are about 1 percent and 25 percent of the initial error at time T = 0. This performance for large initial errors is another reason that relatively poor performance in one coordinate ($\dot{r}_0$) for the more nominal case of Figs. 4, 5, 6, and 7 may be acceptable, as discussed earlier.

Excursions in Λ

As discussed above, ΔT must be sized so as to match Λ and R if initiation and convergence are to occur successfully. The preceding variation of ΔT over two orders of magnitude indicates the range of insensitivity of this variation for the particular example. Experience with the routine illustrates that the R and Λ-to-ΔT relationship is a very sensitive one.

Excursions in R and Estimated-to-Actual Noise

The secondary R-matrix given on page 48 was used to explore its effect upon initiation and convergence. Neither the pure differential-

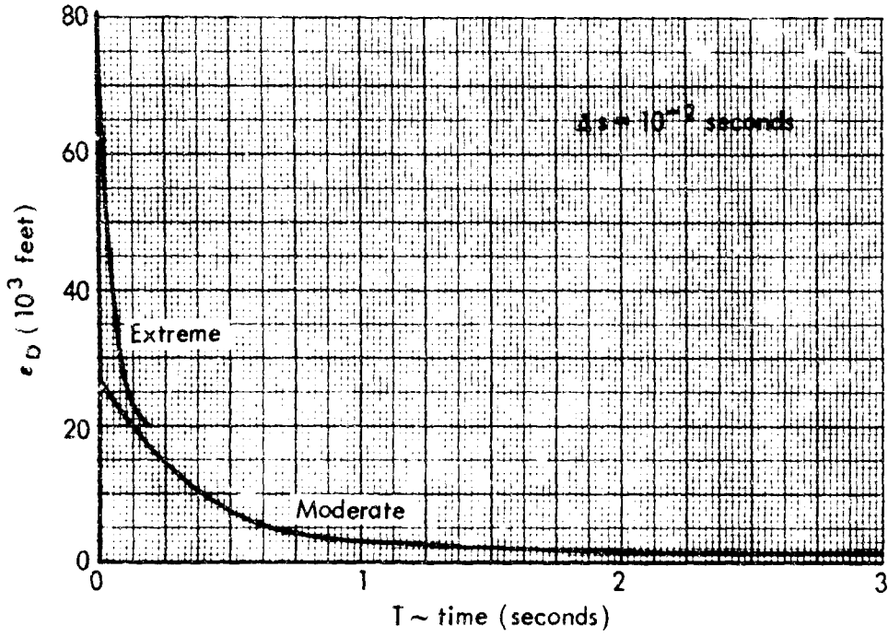

Fig.8a —Miss-distance error for differential-equation/Newtonian solution for moderate and extreme initial conditions

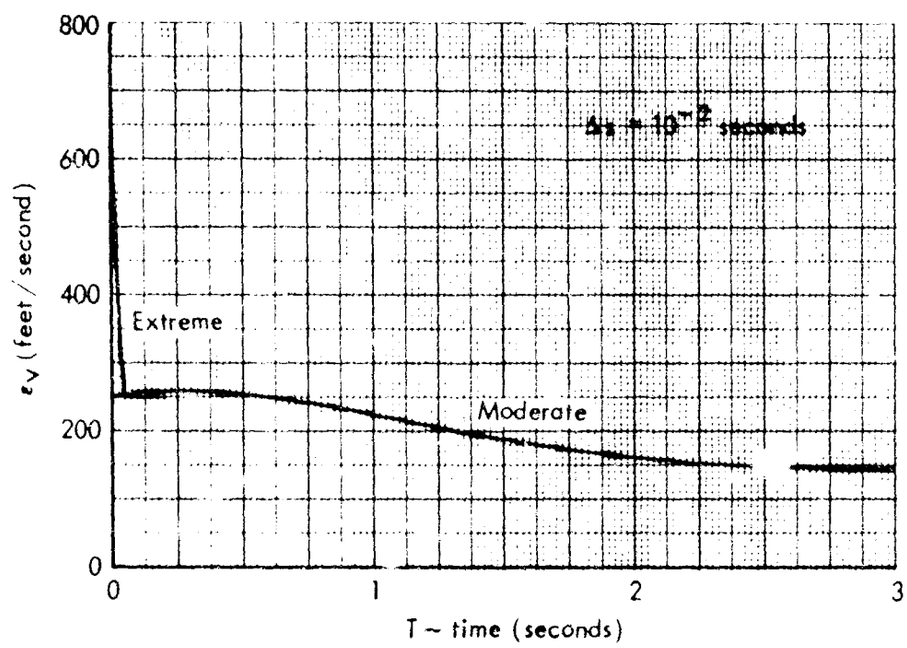

Fig.8b — Velocity error for differential-equation/Newtonian solution for moderate and extreme initial conditions

equation nor the differential-equation/Newtonian estimator would converge appropriately for Δs values of 10^{-1} and 10^{-2}. For $\Delta s = 10^{-3}$ the pure differential-equation estimator converged well, as shown in Fig. 9.

It is also of interest to explore convergence of the estimator if the noise does not correspond exactly to the R-matrix employed. This is shown for $\Delta s = 10^{-2}$ seconds in Fig. 9, where the terminology is defined as follows:

- o No noise--large R: primary R-matrix was used in equations, but there was no noise added to radar observations.
- o Small noise--large R: primary R-matrix was used in equations, but noise added to radar observations corresponded to secondary R-matrix,
- o Large noise--large R: primary R-matrix was used in equations, and noise added to radar observation corresponded to primary R-matrix.

The T range shown is great enough just to show how rapidly all three cases converge to the same solution, independent of the estimated-to-actual noise relationship. It is clear that as long as R is "larger" than or equivalent to the actual noise, convergence takes place. The contrary case--noise "larger" than R--should also converge for appropriately small ΔT. However, its calculation requires considerably more machine time and was not done here.

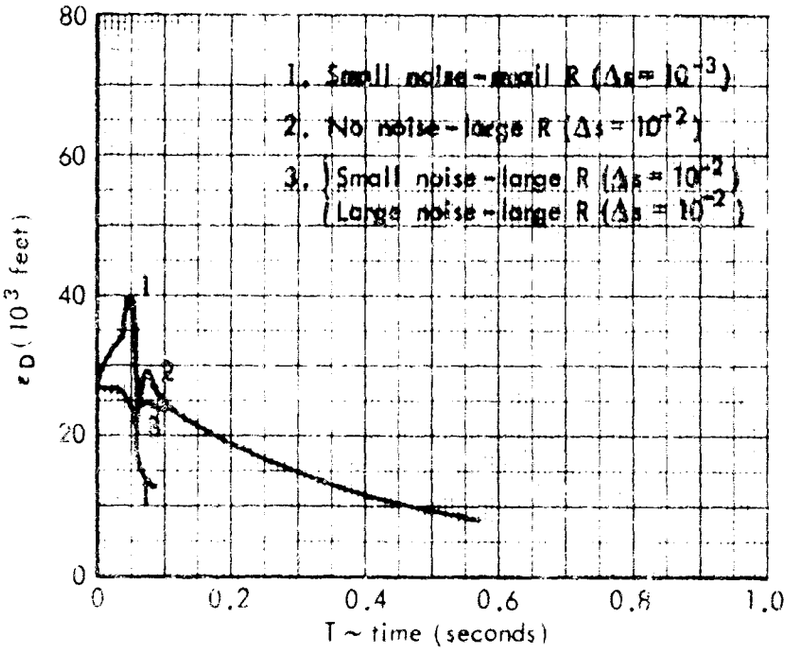

Fig. 9a — Miss distance error for pure differential-equation ($\Delta s = 10^{-3}$) and for differential-equation Newtonian solution ($\Delta s = 10^{-2}$) for several estimated-to-actual noise cases

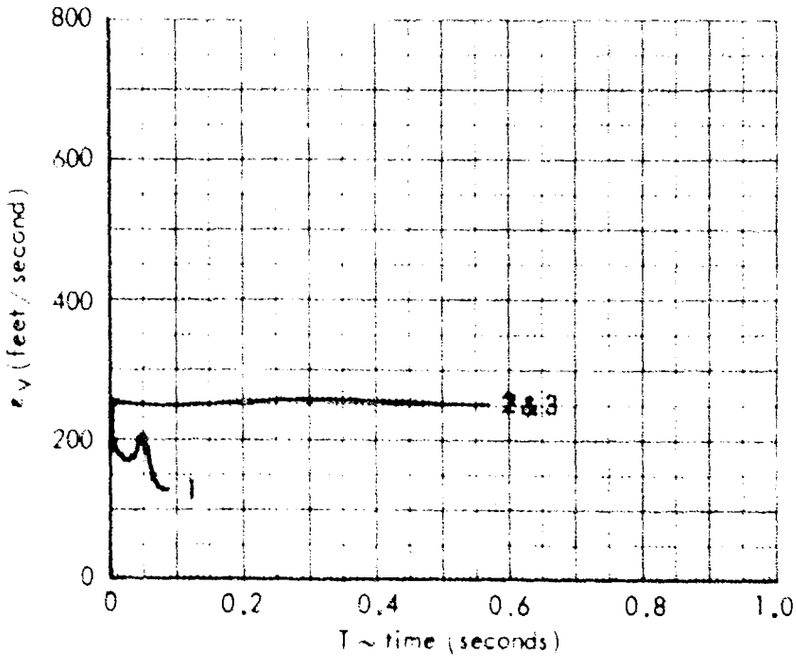

Fig. 9b — Velocity error for pure differential-equation ($\Delta s = 10^{-3}$) and for differential-equation Newtonian solution ($\Delta s = 10^{-2}$) for several estimated-to-actual noise cases

V. CONCLUSIONS

These calculations did not compare MLE performance to other estimators. Since computational results by other researchers (see Ref. 13) had indicated the greater accuracy of the MLE in the scalar case, the calculations here were designed to test the feasibility of applying MLE techniques to complicated vector situations. To that end, they examined convergence, limit of convergence, effect of initial error, noise magnitude, accuracy of noise estimation, and ability to estimate unknown parameters.

THEORETICAL RESULTS

1. The likelihood expression $p(C|Z(s), 0 \leq s \leq T)$, upon which the maximum-likelihood estimate for the initial condition is based, was derived and then specialized to the white-noise case. The desired maximum-likelihood estimate is that initial condition C which maximizes $p(C|Z(s), 0 \leq s \leq T)$.

2. The maximum-likelihood estimate was shown to satisfy (as a function of the upper observation limit T) a differential equation (27) of relatively simple form; it was also shown that the resulting computational algorithm can be usefully applied.

3. The conditional Cramer-Rao bound for the maximum-likelihood estimator was developed and an approximation derived.

4. The maximum-likelihood estimator for a specific problem was then derived from the preceding theoretical results and applied to a white-noise case; similarly for the Cramer-Rao bound.

NUMERICAL RESULTS

1. For appropriate values of $\Delta s = \Delta T$ the estimator converges to a useful solution for X_0, even given extreme initial errors. The conditional error appears to be of the same order as the Cramer-Rao lower bound.

2. In the application presented, the variable $\dot{\theta}$ cannot be measured by the radar, and so $\dot{\theta}$ must be estimated entirely in terms of the other observables. Therefore, $\dot{\theta}_0$ can be considered as an unknown parameter (with known a priori statistics). The estimator approaches the true value of $\dot{\theta}_0$, as well as those of r_0 and θ_0, with considerable accuracy. This is done at the expense of relatively slower convergence of $\dot{r}_0$.

3. As a function of T, the individual estimates of r_0, $\dot{r}_0$, θ_0, and $\dot{\theta}_0$ may not converge monotonically but may vary so as to improve the overall estimate of X_0.

4. Convergence the pure differential-equation solution estimator depends critically upon the value of ΔT used. This is to be expected, since ΔT is the factor by which the derivative of $\hat{X}(T)$ is weighted in forward projections in time. The nature of this dependence is influenced by Λ and R.

5. Convergence of the differential-equation/Newtonian estimator is less dependent upon the value of ΔT; both estimators have basic accuracy limitations determined by the size of Δs, which introduces a granularity into the surface $f(X,T)$.

6. Initiation difficulties presented by large values for $X_0 - \mu$ can be overcome by the use of sufficiently small values for ΔT. A

variable ΔT, increasing with T, could perhaps be used in future programs.

7. For the case at hand, 3 to 4 seconds of real track time sufficed for essential convergence, at $\Delta a = 10^{-3}$ seconds.

8. Computer time required was several orders of magnitude greater than the real-time observation processed. With specially written programs and special-purpose machines this should be reducible to on-line proportions. Given the apparent accuracy advantages of MLE over certain approximations evaluated heretofore and its computational feasibility (i.e., solution at and continuing from T = 0), MLE techniques promise a useful approach to complicated nonlinear problems of the kind examined here.

FUTURE DEVELOPMENTS

1. An important remaining problem is that of determining the relationship between the maximum-likelihood estimate and the minimum RMS (error) estimate.

2. The bias of the maximum-likelihood estimate is not understood and needs to be further examined.

3. The error evaluation made here needs to be extended to include estimates of the RMS error.

4. More efficient programs need to be devised, now that it has been demonstrated that the algorithm developed here is useful in the vector case. This will assist in the evaluation in 3 just above.

Appendix A

DERIVATION OF CONDITIONAL PROBABILITY $p(C|Z(s),\ 0 \le s \le T)$

Defining

$$X(s;C) = X(s), \quad \text{given that } X(0) = C, \tag{A1a}$$

and

$$X(0) = X_0, \tag{A1b}$$

then for given $X_0 = C$ we have

$$E[Z(s)|C] = E[H(s)X(s;C) + N(s)] = H(s)X(s;C) + E[N(s)]$$

$$= H(s)X(s;C), \tag{A2}$$

and the process covariance conditioned upon C is

$$R(s,u|C) = E[(Z(s) - E[Z(s)|C])(Z(u) - E[Z(u)|C])^*|C]$$

$$= E[[H(s)X(s;C) + N(s) - H(s)X(s;C)]$$

$$[H(u)X(u;C) + N(u) - H(u)X(u;C)]^*]$$

$$= E[N(s)N^*(u)] = R(s,u).$$

That is, the covariance function of $Z(s)$ is independent of $X_0 = C$ and is the same as the noise covariance function. However, the mean varies with $X_0 = C$ as per (A2).

We will assume for this development that $R(s,u)$ satisfies certain conditions permitting the following manipulations; these conditions will be spelled out later.

Let

$$\{\psi_i^k(s),\ \lambda_i^k\}$$

be the <u>orthonormal</u> eigenvectors and the related eigenvalues of $r_k(s,u)$; that is, $\{\psi_i^k(s),\ \lambda_i^k\}$ satisfy

$$\int_0^T r_k(s,u)\psi_i^k(u)du = \lambda_i^k\psi_i^k(s). \tag{A3}$$

(Note: $\psi_i^k(u)$ is a scalar function.)

Define

$$z_i^k = \int_0^T z_k(s)\psi_i^k(s)ds, \quad \text{and} \tag{A4}$$

$$m_i^k(C) = E[z_i^k|C]. \tag{A5}$$

Then

$$m_i^k(C) = E\left\{\int_0^T z_k(s)\psi_i^k(s)ds\,|C\right\} = \int_0^T E\{z_k(s)|C\}\psi_i^k(s)ds$$

or, from (A2),

$$m_i^k(C) = \int_0^T H_k(s)^{\prime}(s;C)\psi_i^k(s)ds, \tag{A6}$$

where

$$H(s) = \begin{pmatrix} H_1(s) \\ \\ H_p(s) \end{pmatrix} \quad ; \text{ i.e., } H_k(s) = (h_{k1}(s) \ \ldots \ h_{km}(s)).$$

Then, defining

$$r_{ij}^{k}(C) = E\{[z_i^k - E(z_i^k|C)][z_j^k - E(z_j^k|C)]|C\}, \tag{A7}$$

one gets

$$r_{ij}^{k}(C) = E\left\{\int_0^T \int_0^T n_k(s)n_k(u)\psi_i^k(s)\psi_j^k(u)dsdu|C\right\} \tag{A8}$$

$$= \int_0^T \int_0^T E[n_k(s)n_k(u)|C]\psi_i^k(s)\psi_j^k(u)dsdu;$$

$$r_{ij}^{k}(C) = \int_0^T \int_0^T r_k(s,u)\psi_i^k(s)\psi_j^k(u)dsdu = r_{ij}^{k} \tag{A9}$$

independent of C, so (A3) gives

$$r_{ij}^{k} = \int_0^T \lambda_j\psi_i^k(s)\psi_j^k(s)ds = \lambda_j^k\delta_{ij}, \tag{A10}$$

where δ_{ij} is the Kronecker delta:

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}. \tag{A11}$$

It is also clear by analogy with (A8) that z_i^k and z_j^v are independent for $k \neq v$, since $n_k(s)$ and $n_v(u)$ are independent noise processes. We thus can define the random vector series $\{Z(i), i = 1,2, \ldots\}$, where

$$Z(i) = \begin{pmatrix} z_i^1 \\ \vdots \\ z_i^p \end{pmatrix}. \tag{A12}$$

and we know from above that $Z(i|C)$ is gaussian with mean

$$M(i;C) = \begin{pmatrix} m_i^1(C) \\ \vdots \\ m_i^p(C) \end{pmatrix} \tag{A13}$$

and covariance matrix

$$R(i) = \begin{pmatrix} \lambda_i^1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \lambda_i^p \end{pmatrix} \text{ (independent of C)}, \tag{A14}$$

where $m_i^k(C)$ is given by (A6):

$$m_i^k(C) = \int_0^T H_k(s)X(s;C)\psi_i^k(s)ds. \tag{A6}$$

Letting $p[C,Z(1), \ldots, Z(\ell)]$ be the probability density for $[C,Z(1), \ldots, Z(\ell)]$, we can write

$$\begin{aligned} p[C,Z(1), \ldots, Z(\ell)] &= p(C)p(Z(1), \ldots, Z(\ell)|C) \\ &= p[Z(1), \ldots, Z(\ell)]p[C|Z(1), \ldots, Z(\ell)], \end{aligned} \tag{A15}$$

and so

$$\begin{aligned} p[C|Z(1), \ldots, Z(\ell)] &= \frac{p(C)p(Z(1), \ldots, Z(\ell)|C)}{p[Z(1), \ldots, Z(\ell)]} \\ &= K_1 p(C)p(Z(1), \ldots, Z(\ell)|C). \end{aligned} \tag{A16}$$

Thus, $p(C|Z(s),\ 0 \le s \le T) = p(C|Z(1),\ Z(2),\ \ldots)$ is given by

$$\lim_{\ell\to\infty} p[C|Z(1),\ \ldots,\ Z(\ell)]$$

$$= K_2 \exp\left[-\frac{1}{2}\left\{(C-\mu)^* \Lambda^{-1}(C-\mu) + \sum_{i=1}^{\infty} [Z(i) - M(i,C)]^* R^{-1}(i)[Z(i) - M(i,C)]\right\}\right]. \tag{A17}$$

Then, $\sum_{i=1}^{\infty} [Z(i) - M(i,C)]^* R^{-1}(i)[Z(i) - M(i,C)]$

$$= \sum_{i=1}^{\infty} \sum_{k=1}^{p} \frac{1}{\lambda_i^k} [z_i^k - m_i^k(C)]^2$$

$$= \sum_{i=1}^{\infty} \sum_{k=1}^{p} \frac{1}{\lambda_i^k} [\{z_i^k\}^2 - 2z_i^k\, m_i^k(C) + \{m_i^k(C)\}^2]$$

$$= \sum_{k=1}^{p} \sum_{i=1}^{\infty} \left[\frac{1}{\lambda_i^k}\{z_i^k\}^2 - 2\frac{z_i^k\, m_i^k(C)}{\lambda_i^k} + \frac{\{m_i^k(C)\}^2}{\lambda_i^k}\right]$$

$$= \sum_{k=1}^{p} \sum_{i=1}^{\infty} \frac{1}{\lambda_i^k}\{z_i^k\}^2 - 2\sum_{k=1}^{p} \mu^k(C) + \sum_{k=1}^{p} \{\nu^k(C)\}^2, \text{ with}$$

$$\mu^k(C) = \sum_{i=1}^{\infty} \frac{z_i^k\, m_i^k(C)}{\lambda_i^k}, \tag{A18a}$$

$$[\nu^k(C)]^2 = \sum_{i=1}^{\infty} \frac{\{m_i^k(C)\}^2}{\lambda_i^k}, \tag{A18b}$$

which gives

$$p[C|Z(s),\ 0 \le s < T] =$$

$$K_3 \exp\left[-\frac{1}{2}\left\{(C-\mu)^*\Lambda^{-1}(C-\mu) - 2\sum_{k=1}^{p}\mu^k(C) + \sum_{k=1}^{p}[\gamma^k(C)]^2\right\}\right]. \tag{A19}$$

If we define

$$g^k(s;C) = \sum_{i=1}^{\infty}\frac{m_i^k(C)}{\lambda_i^k}\cdot\psi_i^k(s), \tag{A20}$$

then we have from (A6),

$$[\gamma^k(C)]^2 = \int_0^T H_k(s)X(s;C)g^k(s;C)ds. \tag{A21}$$

From (A20), provided the $\{\psi_i^k(s)\}$ are complete, $g^k(s;C)$ satisfies

$$\int_0^T g^k(s;C)\tau_k(s,u)ds = H_k(u)X(u;C), \tag{A22}$$

and from

$$\mu^k(C) = \sum_{i=1}^{\infty}\frac{z_i^k\, m_i^k(C)}{\lambda_i^k} = \int_0^T z_k(s)g^k(s;C)ds, \tag{A23}$$

we have

$p[C|Z(s),\ 0 < s < T] =$

$$K_3 \exp\left[-\frac{1}{2}\left\{(C-u)^{*}\Lambda^{-1}(C-u) - 2\sum_{k=1}^{p}\int_0^T z_k(s)g^k(s;C)ds\right.\right.$$

$$\left.\left. + \sum_{k=1}^{p}\int_0^T H_k(s)X(s;C)g^k(s;C)ds\right\}\right], \qquad \text{(A24a)}$$

where $g^k(s;C)$ satisfies

$$\int_0^T g^k(s,C)r_k(s;u)ds = H_k(u)X(u;C). \qquad \text{(A24b)}$$

If we define

$$\zeta_k(s) = \sum_{i=1}^{\infty}\frac{z_i^k}{\lambda_i^k}\psi_i^k(s), \qquad \text{(A25a)}$$

there results

$$\int_0^T \zeta_k(s)r_k(s,u)ds = \sum_{i=1}^{\infty} z_i^k\,\psi_i^k(u) = z_k(u), \qquad \text{(A25b)}$$

provided (again) that the $\{\psi_i^k(u)\}$ are complete. That is,

$$\sum_{i=1}^{\infty}\frac{[z_i^k]^2}{\lambda_i^k} = \int_0^T z_k(s)\zeta_k(s)ds,$$

where $\zeta_k(s)$ satisfies

$$\int_0^T \zeta_k(s)r_k(s,u)ds = z_k(u).$$

Since the last is independent of C, we can rewrite (A24) as

$$p(C|Z(s),\ 0 < s \le T) = K\ G(C), \tag{A26a}$$

where

$$G(C) = \exp\left[-\frac{1}{2}\left\{(C-\mu)^{*}\Sigma^{-1}(C-\mu) + \sum_{k=1}^{p}\int_0^T z_k(s)\zeta_k(s)ds\right.\right.$$

$$\left.\left. - 2\sum_{k=1}^{p}\int_0^T z_k(s)g^{k}(s;C)ds + \sum_{k=1}^{p}\int_0^T H_k(s)X(s;C)g^{k}(s;C)ds\right\}\right], \tag{A26b}$$

and $\zeta_k(s)$ and $g^k(s;C)$ satisfy

$$\int_0^T \zeta_k(s)r_k(s,u)ds = z_k(u),\quad \int_0^T g^k(s;C)r_k(s;u)ds = H_k(u)X(u;C), \tag{A26c}$$

and

$$K^{-1} = \int_C G(C)dc\ . \tag{A26d}$$

Given the preceding development (and nomenclature), it is now convenient to state the conditions under which it is admissible.[21] First, if $r_k(s,u)$ is positive definite, then there exists a countable (non-empty) set of orthonormal eigenvectors $\{\phi_i^k(s)\}$ and corresponding

eigenvalues $\{\lambda_i^k\}$ such that the $\{\lambda_i^k\}$ are positive or zero and bounded above; there are no other eigenvectors orthnormal to these $\{\psi_i^k(s)\}$; and the $\{\psi_i^k(s)\}$ are complete over the functions of integrable square on $(0,T)$. The latter means that for any function $g(s)$ of integrable square it can be written as

$$g(s) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{i=1}^{N} g_i \psi_i^k(s),$$

where

$$g_i = \int_0^T g(s)\psi_i^k(s)ds$$

and the meaning of l.i.m. is that

$$\lim_{N\to\infty} \int_0^T \left[g(s) - \sum_{i=1}^{N} g_i \psi_i^k(s)\right]^2 ds = 0.$$

Mercer's Theorem then applies, which states that

$$r_k(s,u) = \sum_{i=1}^{\infty} \lambda_i^k \psi_i^k(s)\psi_i^k(u),$$

where the convergence is uniform in (s,u) for $0 \le s, u \le T$. Further, Picard's Theorem then states that the $g^k(s;C)$ in (A22) exists and is of integrable square over $(0,T)$ if and only if

$$R_k(u)X(u;C) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{i=1}^{N} \gamma_i^k(C) \psi_i^k(u),$$

where

$$\overset{k}{\underset{i}{\beta}}(C) = \int_0^T \phi_k(u)X(u;C)\psi_i^k(u)du$$

and the series

$$\sum_{i=1}^{\infty} \frac{[\overset{k}{\underset{i}{\beta}}(C)]^2}{[\lambda_i^k]^2} < \infty .$$

Picard's Theorem further states that the unique solution is then

$$g^k(s;C) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{i=1}^{N} \frac{\overset{k}{\underset{i}{\beta}}(C)}{\lambda_i^k} \psi_i^k(s),$$

as in (A20) with $\overset{k}{\underset{i}{\beta}}(C) = \overset{k}{\underset{i}{m}}(C)$.

Similar remarks (from Picard's Theorem) apply to $\zeta_k(s)$ of (A25) where now we insist that

$$\sum_{i=1}^{\infty} \frac{[\gamma_i^k]^2}{[\lambda_i^k]^2} < \infty ,$$

where

$$\gamma_i^k = \int_0^T z_k(s)\psi_i^k(s)ds .$$

The unique expression for $\zeta_k(s)$ is then

$$\zeta_k(s) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{i=1}^{N} \frac{\nu_i^k}{\lambda_i^k} \phi_i^k(s),$$

as in (A25a) with $\nu_i^k = z_i^k$.

We thus see that necessary and sufficient conditions for the preceding development are

1. $r_k(s,u)$ positive definite, $k = 1, \ldots, p$,

2. $\displaystyle\sum_{i=1}^{\infty} \frac{[m_i^k(C)]^2}{(\lambda_i^k)^2} < \infty$, $k = 1, \ldots, p$,

3. $\displaystyle\sum_{i=1}^{\infty} \frac{[z_i^k]^2}{(\lambda_i^k)^2} < \infty$, $k = 1, \ldots, p$,

where λ_i^k, z_i^k, and $m_i^k(C)$ are defined in (A3), (A4), and (A6).

Equation (A26) simplifies greatly when the noise $N(s)$ is such as to permit the above development <u>and</u> the $r_k(s,u)$ can be treated as delta-functions:

$$r_k(s,u) = \phi_k\, \delta(s - u),\ k = 1, \ldots, p.$$

From (A26) we then have $\zeta_k(s)$ and $g^k(s;C)$ satisfying

$$\int_0^T \zeta_k(u)\, \phi_k\, \delta(s - u)du = \phi_k\, \zeta_k(s) = z_k(s)$$

and

$$\int_0^T g^k(s;C)\,\xi_k\,\delta(s-u)\,ds = H_k(u)X(u;C),$$

with

$$\xi_k(s) = \frac{1}{\phi_k}\,z_k(s),$$

$$g^k(s;C) = \frac{1}{\phi_k}\,H_k(s)X(s;C),$$

so that (A26) may be rewritten as

$$p(C|Z(s),\ 0 < s < T) = K\,J(C), \tag{A27a}$$

where

$$J(C) = \exp\left[-\frac{1}{2}\left\{(C-\mu)^* \Lambda^{-1}(C-\mu) + \int_0^T [Z(s) - H(s)X(s;C)]^* R^{-1}[Z(s) - H(s)X(s;C)]\,ds\right\}\right], \tag{A27b}$$

$$R = \begin{pmatrix} \phi_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \phi_p \end{pmatrix}, \tag{A27c}$$

$$K^{-1} = \int_C J(C)\,dc. \tag{A27d}$$

Appendix B

PROGRAM FLOW DIAGRAM

In Fig. 10 is shown a flow diagram depicting the operation of the maximum-likelihood-estimator program. The following describes what takes place at particular points in the program.

1. Start.
2. Input of indices, constants, observation. T is the observation-interval upper limit, $\{T_i\}$ are the points at which estimates are made, $\hat{X}(T_i)$ is the estimate at time T_i.
3. Question leading to use of certain input values if answer is no; leading to update if answer is yes.
4. Update of $\hat{X}(T_i)$ to $\hat{X}(T_i + \Delta T)$ via basic differential equation (27) governing $\hat{X}(T)$ and (60).
5. Update of time from T_i to $T_i + \Delta T$.
6. Setup of initial conditions for integration to get functions tabled over $0 < s < T$ which are needed in Newtonian iteration in 13, as well as needed in 4.
7. Integration to table over $0 < s < T$ functions needed for Newtonian iteration in 13.
8. Print $\hat{X}(T_i)$, $f(\hat{X}(T_i),T_i)$, $\mathrm{Det}[\nabla_{\hat{X}\hat{X}} f(\hat{X}(T_i),T_i)]$, $\nabla_{\hat{X}} f(\hat{X}(T_i),T_i)$, $\nabla_{\hat{X}} X(T_i;\hat{X}[T_i])$.
9. Test of whether $\nabla_{\hat{X}} f(\hat{X}(T_i),T_i)$ is near enough to $\emptyset$.
10. If $\nabla_{\hat{X}} f(\hat{X}(T_i),T_i)$ is sufficiently near $\emptyset$ in 9, [illegible] there has been calculated a previous Newtonian solution $\hat{X}_0$ at T_i.

11. If there was a previous Newtonian solution $\hat{X}_0$ at T_i, compare the difference between it and the current solution to see if there is need to proceed further with Newtonian iterations.

12. If the difference between solutions in 11 is too great, temporarily store the latest solution as the best one.

13. Improve this best solution via another Newtonian iteration. Then go back to 6 and redo this series of tests et al.

14. If $\nabla_{\hat{X}} f(\hat{X}(T_i), T_i)$ is near enough to $\emptyset$ or $\hat{X}_0$ is near enough to $\hat{X}(T_i)$, ask if $T_i = T$. If $T_i \neq T$, go to 4 and update $\hat{X}(T_i)$ to $\hat{X}(T_i + \Delta T)$ via basic differential equations (27) and (60), etc.

15. If in 14 $T_i = T$, stop.

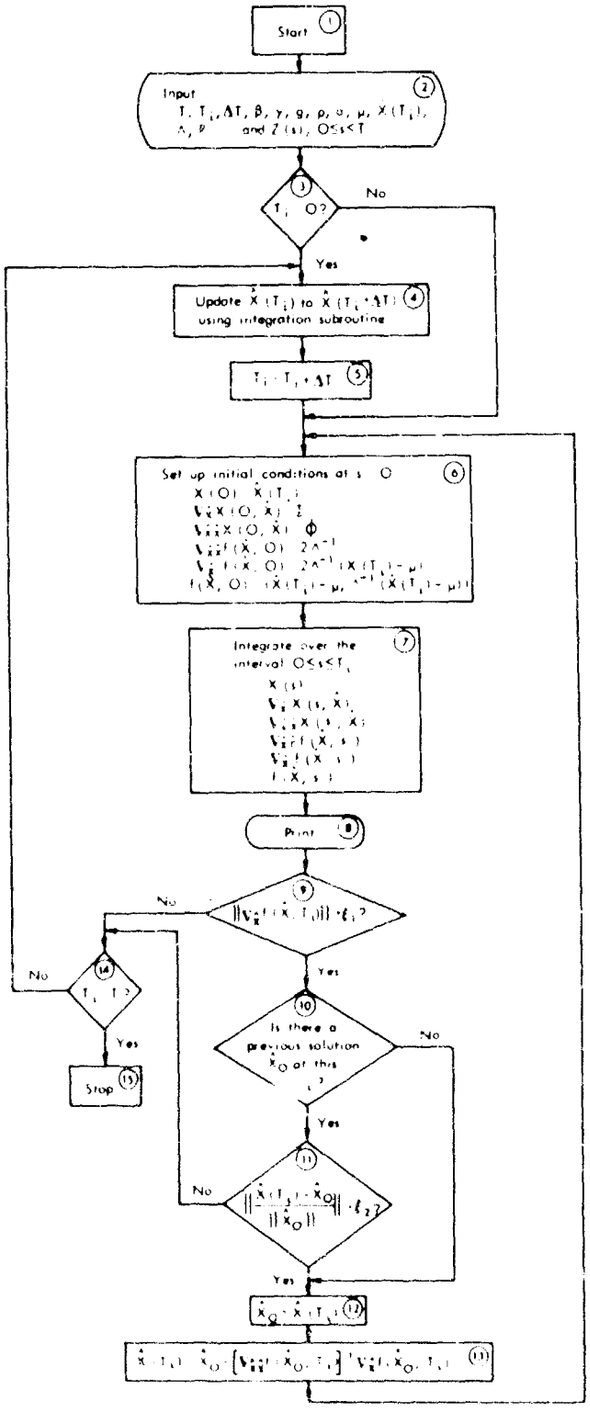

Fig. 10 — Program flow diagram

REFERENCES

1. Kalman, R. E., and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," Journal of Basic Engineering, Vol. 83D, March 1961, pp. 95-108.

2. Doob, J. L., Stochastic Processes, John Wiley & Sons, Inc., New York, 1953.

3. Stratonovich, R. L., "Conditional Markov Processes," Theory of Probability and Its Applications, Vol. 5. No. 2, 1960, pp. 156-178.

4. Wonham, W. M., Some Applications of Stochastic D.E.s to Optimal Nonlinear Filtering, Technical Report 64-3, Research Institute for Advanced Study, Baltimore, Md., February 1964.

5. Balakrishnan, A. V., "A General Theory of Nonlinear Estimation Problems in Control Systems," Journal of Mathematical Analysis and Applications, No. 8, 1964, pp. 4-30.

6. Kushner, H. J., "On the Dynamical Equations of Conditional Probability Density Functions, with Applications to Optimal Stochastic Control Theory," Journal of Mathematical Analysis and Applications, No. 8, 1964, pp. 332-344.

7. Kushner, H. J., "On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes," Journal SIAM on Control, Vol. 2, 1964, pp. 106-119.

8. Bucy, R. S., "Nonlinear Filtering Theory," IEEE Transactions on Automatic Control, Vol. AC-10, No. 2, April 1965, p. 198.

9. Mowery, V. O., "Least Squares Recursive Differential Correction Estimation in Nonlinear Problems," IEEE Transactions on Automatic Control, Vol. AC-10, No. 4, October 1965, pp. 399-407.

10. Bellman, R. E., H. H. Kagiwada, R. E. Kalaba, and R. Sridhar, "Invariant Imbedding and Nonlinear Filtering Theory," The Journal of the Astronautical Sciences, Vol. XIII, No. 3, May-June 1966, pp. 110-115.

11. Friedland, Bernard, and Irwin Bernstein, "Estimation of the State of a Nonlinear Process in the Presence of Nongaussian Noise and Disturbances," Journal of the Franklin Institute, Vol. 281, No. 6, June 1966.

12. Bass, R. W., V. D. Norum, and L. Schwartz, "Optimal Multichannel Nonlinear Filtering," Journal of Mathematical Analysis and Applications, Vol. 16, No. 1, 1966, pp. 152-164.

13. Carney, T. M., and R. M. Goldwyn, "Numerical Experiments with Various Optimal Estimators," Journal of Optimization Theory and Applications, Vol. 1, No. 2, 1967, pp. 113-130.

14. Kushner, H. J., "Approximations to Optimal Nonlinear Filters," Joint Automatic Control Conference Preprint of Papers, June 1967, pp. 613-623.

15. Kushner, H. J., "Nonlinear Filtering: The Exact Dynamical Equations Satisfied by the Conditional Mode," IEEE Transactions on Automatic Control, Vol. AC-12, No. 3, June 1967, pp. 262-267.

16. Balakrishnan, A. V., "Filtering and Prediction Theory," Chap. 3 in A. V. Balakrishnan, et al., Lectures on Communication Theory, McGraw-Hill Book Co., Inc., New York, being published.

17. Balakrishnan, A. V., A New Computing Technique in System Identification, Report No. 6814, Department of Engineering, University of California, Los Angeles, May 1968.

18. Page, L., Introduction to Theoretical Physics, 3d ed., D. Van Nostrand Co., Inc., New York, 1952, pp. 65-66.

19. Boehm, B., ROCKET: RAND's Omnibus Calculator of the Kinematics of Earth Trajectories, The Rand Corporation, RM-3534-PR, July 1963.

20. Saaty, T. L., and J. Bram, Nonlinear Mathematics, McGraw-Hill Book Co., Inc., New York, 1964.

21. Davenport, W. B., Jr., and W. L. Root, Random Signals and Noise, McGraw-Hill Book Co., Inc., New York, 1958, pp. 371-375.

DOCUMENT CONTROL DATA

1. ORIGINATING ACTIVITY THE RAND CORPORATION	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

MAXIMUM-LIKELIHOOD PREDICTION AND ESTIMATION FOR NONLINEAR DYNAMIC SYSTEMS

4. AUTHOR(S) (Last name, first name, initial)

Attaway, L. D.

5. REPORT DATE	6a. TOTAL No. OF PAGES	6b. No. OF REFS.
December 1968	102	21

7. CONTRACT OR GRANT No.	8. ORIGINATOR'S REPORT No.
F44620-67-C-0045	RM-5859-PR

9a. AVAILABILITY / LIMITATION NOTICES	9b. SPONSORING AGENCY
DDC-1	United States Air Force Project RAND

10. ABSTRACT

A method for determining the system state using noise-corrupted observations of a non-linear dynamic invector process, with a numerical application to radar observation of a reentry body. The study examined the feasibility of numerically solving the vector-differential equations satisfied by the maximum-likelihood estimator. The maximum-likelihood estimate is that initial condition which minimizes a certain functional on itself, on the observation, and on the a priori statistics. This functional describes an m plus 1 surface at time T, the upper time limit of observation; the desired estimate corresponds to its minimum. A differential equation is developed governing the evolution of this estimate with time T. Using the differential equation, the algorithm calculates as a function of T that maximum-likelihood solution that evolves over time from the unique solution at time T = 0 (given by the a priori mean vector). Differential-equation solutions with and without Newtonian techniques (for constant T) were used to estimate reentry-vehicle initial conditions for a simulated reentry. One coordinate time rate (initial angle) was not measured and served as an unknown parameter with a priori statics. The estimates converged to within the order of the Cramer-Rao conditional bound. The unknown parameter was handled as successfully as the others.

11. KEY WORDS

Control theory
Detection
Probability
Reentry vehicles
Radar