NASA CONTRACTOR REPORT

SEQUENTIAL PROCESSING TECHNIQUES FOR TRAJECTORY ESTIMATION

by C. G. Pfeiffer, D. D. Morrison, R. E. Mortensen, J. V. Breakwell, W. H. Berry, and M. H. Merel

Prepared by
TRW SYSTEMS GROUP
Redondo Beach, Calif.
for Ames Research Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • OCTOBER 1969

SEQUENTIAL PROCESSING TECHNIQUES FOR TRAJECTORY ESTIMATION

By C. G. Pfeiffer, D. D. Morrison, R. E. Mortensen, J. V. Breakwell, W. H. Berry, and M. H. Merel

Prepared under Contract No. NAS 2-4553 by
TRW SYSTEMS GROUP
Redondo Beach, Calif.

for Ames Research Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

FINAL REPORT

SEQUENTIAL PROCESSING TECHNIQUES FOR TRAJECTORY ESTIMATION

by

C. G. Pfeiffer
D. D. Morrison
R. E. Mortensen
J. V. Breakwell
W. H. Berry
M. H. Merel

13 December 1968

Prepared under Contract No. NAS2-4553

by

TRW Systems Group
Redondo Beach, California

for

AMES RESEARCH CENTER

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

CONTENTS

CONTENTS (Continued)

NOMENCLATURE

A	matrix of partial derivatives of the observations with respect to the state
B	matrix of partial derivatives of the observations with respect to the systematic errors
E	expectation operator
G	matrix relating $\dot{w}$ to $\dot{x}$
H,h	noise-free observation
I	identity matrix
L	linear operator
L(x)	likelihood function
N	tracking normal matrix
n	data noise
p	probability density function
$R_{\oplus}$	magnitude of the station radius vector
r	data residual
s	systematic error
T	final time; duration of a tracking interval
t	time
U	state transition matrix
u	system control
W	weighting matrix
w	state noise
x	system state
y	data vector; the integral of the data vector
z	data vector
α	derivative of state noise; unknown acceleration
Γ	covariance matrix of the data noise
$\delta(s)$	Dirac delta function ($\delta(s) = 0$ if $s \neq 0$, $\delta(0) = 1$)
ε	error
Λ	covariance matrix of x
λ	Lagrange multiplier vector
μ	E(x)
ρ	integral of the data vector; data residual
Σ	covariance matrix
σ	standard deviation
τ	time interval

Subscripts

c	Computed
I	Ito calculus
LS, WLS	Weighted least squares
M	Measured
MV	Minimum variance
o	Initial
T	Final

Superscripts

^, *	Estimate
T	Transpose
$\bullet$	$\frac{d}{dt}$

SUMMARY

This report describes the state of the art in sequential processing techniques for trajectory estimation as seen by the authors. Some of the material presented is a tutorial interpretation and summary of well known results; other material is new. A major topic discussed is the effect of nonlinearities, with the aim of developing and justifying practical algorithms for treating the problem. Since nonlinear, maximum likelihood estimation theory has been successfully used in orbit determination for some years, this approach has been emphasized over the more recently suggested nonlinear minimum variance technique.

Section 1 is the Introduction, and outlines the technical problems to be discussed. Section 2 describes the minimum variance estimation technique and some of the approximations which have been suggested in order to make the algorithm practical. Section 3 discusses maximum likelihood estimation, control, and error analysis of the form presently employed for orbit determination. Section 4 derives the maximum likelihood estimator for unknown acceleration (state noise), and develops a practical algorithm for solving the problem. Section 5 discusses some nonlinear and numerical properties of sequential orbit determination algorithms, and derives a theorem which is fundamental to some forms of sequential estimation. Section 6 analyzes estimation in the presence of a slight nonlinearity, and compares several reasonable estimators. Section 7 treats orbit parameters which, for some reason such as computer limitation, have not been included in the estimation process (systematic errors). Section 8 discusses the treatment of correlated data which does not have the usually assumed Markoff property.

This report describes the work of six individuals. Sections 1 and 3 were contributed by C. G. Pfeiffer; Head, Mathematical Physics Section, Guidance and Analysis Dept., TRW Systems Group. Section 2 by R. E. Mortensen; Assistant Professor of Engineering, UCLA, and consultant, TRW Systems Group. Section 4 by C. G. Pfeiffer and D. D. Morrison; Staff Engineer, Analytical Research Operations, TRW Systems Group. Section 5 by D. D. Morrison. Section 6 by J. V. Breakwell; Professor of Engineering, Standford University, and consultant, TRW Systems Group. Section 7 by

W. H. Berry; Member of the Technical Staff, Guidance and Analysis Dept., TRW Systems Group, and C. G. Pfeiffer. Section 8 by M. H. Merel; Member of the Technical Staff, Guidance and Analysis Dept., TRW Systems Group. Attempting to maintain continuity between the rather diverse subjects treated, compilation and minor revision was done by C. G. Pfeiffer. Editing and typing was performed by D. A. Henderson.

In order to avoid duplicate publication, not all of the analysis completed under this contract has been included here. Thus to complete the discussion on sequential processing techniques the reader should refer to "Mathematical Problems of Modeling Stochastic Nonlinear Dynamic Systems" by R. E. Mortensen, issued as NASA Report CR-1168, and "On the Identification of Observable Orbit Parameters, with Application to Lunar Orbiter Tracking" by C. G. Pfeiffer, which is scheduled to appear in the January-February 1969 issue of the Journal of the Astronautical Sciences. (The latter work was also partially supported by NASA contract NAS 9-4810, administered by Manned Spacecraft Center, Houston, Texas). In addition, two other papers not covered by this contract should be considered as part of this discussion: "Maximum Likelihood Recursive Nonlinear Filtering" by R. E. Mortensen, published in Journal of Optimization Theory and Applications (Vol. 2, No. 6, 1968, pp 386-394), and "Methods for Nonlinear Least Squares Problems and Convergence Proofs" by D. D. Morrison, published in the Jet Propulsion Laboratory Tracking and Orbit Determination Seminar Proceedings (February 23-26, 1960, Pasadena, California).

1. INTRODUCTION

In the orbit determination problem we are given a nonlinear dynamic system of the form

$$\frac{dx}{dt} = f(x(t), u(t), t) + \left(\frac{dw}{dt}\right) \qquad 0 \le t \le T \qquad (1.1)$$

where x is the state vector, consisting of position and velocity of the spacecraft plus any parameters which affect the acceleration of the vehicle (in which case $\frac{dx_i}{dt} = 0$, x_i = constant); u(t) is a control (guidance) to be applied, such as one or more acceleration impulses to be supplied by the spacecraft engine; and $\frac{dw}{dt}$ is state noise, such as random solar wind or leakage of the spacecraft attitude control system. We know that sometime during the flight we will be given a sequence of data points

$$z(N) = [z(t_o), z(t_1) \cdots z(t_N)] \qquad (1.2)$$

and that this data will be used in some way in real time to predict (estimate) the state at the final time T, x(T), or at some earlier time. It is feasible to do this, for if (1.1) is thought of as a discrete equation, so that there are a finite number of state disturbances $dw(t_i)$, then we can model the data as

$$z(N) = H_N(x(0), dw) + n(N) \qquad (1.3)$$

where x(0) is the initial condition, dw is a vector of state disturbances up to t_N, and, n(N) is data noise. Thus, via (1.1), (1.3) and the assumed apriori statistics of x(0), dw, and n(N), a statistical relationship between z(N) and x(T) can be derived, and, given z(N), a meaningful estimate of x(T) can be obtained. If such an estimate is made at $t_K < t_N$, based upon a partial set of data z(K), and the estimate is later revised at $t_J > t_K$, based upon the larger data set z(J), the estimation procedure is said to be sequential. Sequential estimation is almost always required for space missions, for the guidance corrections must be applied before all the data has been obtained.

The ultimate sequential estimation procedure is to revise the estimate of $x(T)$ after each new data point $z(t_i)$ is obtained. It is in general neither necessary nor practical to do this, but, for the special case of linear systems with Gaussian, white disturbances n and dw, Kalman has shown [1],[2] an elegant way to do this. His work was generalized in [3] to obtain for an arbitrary stochastic process the necessary and sufficient conditions which must apply (i.e., the wide sense Markoff property) if the procedure is to yield the minimum variance estimate. It was then pointed out in [4] that, if the necessary and sufficient conditions apply, the stochastic process can always be modeled as a linear dynamic system with white Gaussian disturbances.

Thus in the nonlinear case it is not at all clear how to do sequential estimation. Several approaches have been tried or suggested:

(1) apply the linear Kalman filter, hoping that enough data will be obtained to cause the estimate to converge to a reasonable answer

(2) apply a modified form of the linear Kalman filter which includes nonlinear correction terms

(3) calculate the true minimum variance estimate, which is the first moment of the conditional probability density function of the state, given the data

(4) derive some approximation to the true minimum variance estimate

(5) calculate the true maximum likelihood estimate by iterating over all the available data to find the root of the differentiated likelihood function

(6) derive an approximation to the true maximum likelihood estimate

Methods (1) and (2) sometimes work, but theoretical justification is lacking and counterexamples can be constructed where such an approach is poor. Method (3) is elegant and intuitively satisfying, but is at present not practical because a multidimensional, nonlinear partial differential equation must be solved. Method (4) is another way of arriving at Method (2), and suffers the same limitations. Method (5) is presently used in orbit determination work, where a sequential estimate is needed at only a few times during the mission. This approach has been highly successful, but has the potential disadvantage for future applications that some time and computer

capability is required in order to find the estimate. Method (6) has received little attention, mainly because Method (5) has been available and the computational load and time lag have not been an important factor.

In this report we discuss certain questions arising in nonlinear sequential orbit determination. In so doing we will have in mind the following problem areas:

(1) stochastic modeling - It is not at all clear what meaning can be given to a solution of the nonlinear differential equation (1.1) when state noise ($\frac{dw}{dt}$) is present. This question, which requires that the analyst take a position on the Ito vs. Stratonovich calculus, has been treated under this contract and is discussed in [5]. We shall not repeat that work here; suffice it to say that the Stratonovich interpretation will be assumed unless it is stated otherwise .

(2) minimum variance vs. maximum likelihood estimation - Either of these well known estimators could be employed, depending upon factors such as ease of mechanization for a particular problem and personal taste. These two forms will be discussed in Sections 2 and 3.

(3) combined estimation and control - Although this report is primarily devoted to the estimation problem, it is well known that the combined estimation and control problem leads to some new and puzzling considerations. For example, the partial differential equation of Method (3) above should be modified so as to obtain the equivalent of the dynamic programming solution. The combined problem is discussed in Section 3 from the maximum likelihood point of view in an attempt to justify the presently employed procedure in orbit determination and guidance work.

(4) the effect of nonlinearities - This question, which is central to this report, is discussed from various points of view in Sections 2 - 6.

(5) numerical considerations - It is well known that even in linear systems numerical roundoff error can cause the estimate to diverge from the true value. This question is discussed in Section 5.

(6) convergence properties - If one is willing to iterate over some or all of the data in order to get an estimate, it needs to be shown that the chosen algorithm will converge and that the converged answer will be unique. This question is discussed in Section 5 and in Reference [6].

(7) selection of observable orbit parameters - Whatever algorithm is employed, it is clear that the number of orbit parameters (i.e., initial condition components and acceleration parameters) should be restricted to the minimum required to model the data, that is, only the "observable" parameters should be estimated. This question was treated under this contract, and is discussed in Reference [7].

(8) consideration of systematic errors - As a practical matter, it is usually true that some of the observable parameters are left out of the model because of, say, computer limitations. The treatment of these unestimated parameters, which are called systematic errors, is discussed in Section 7.

(9) treatment of correlated data - Derivations of orbit determination algorithms usually assume uncorrelated (white) data noise, supposing that correlated data could be easily incorporated as part of the model of the dynamic system. This is not necessarily true, especially if the data noise is not a first order Markoff process. This question is discussed in Section 7.

2. MINIMUM VARIANCE ESTIMATION

2.1 INTRODUCTION

The various techniques of sequential stochastic estimation have their origins in the early least-squares differential correction schemes for orbit determination. One justification for these early schemes is the theory of maximum likelihood estimation to be discussed in Section 3. An account of the development of orbit determination methods is given in the paper by Mowery [8]. The nonlinear minimum variance sequential estimator follows from the suggestion of Stratonovich [9] that the fundamental entity in sequential estimation is the conditional probability density function $p(\xi,t|y_{[t_o,t]})$ for the current state (ξ) being estimated, given the record of the observational data ($y_{[t_o,t]}$). In Sections 2.3 and 2.4 the partial differential equations for this function are discussed. The basic problem is to make inferences concerning the behavior of a certain Markov stochastic process, using only the information to be gleaned from a knowledge of a model for generating the process together with the past history of observations of a related process.

In other words, we are given a model which generates two related stochastic processes. The problem is to compute the conditional probability distribution of the current state of the first process, given the past history of observations of the second process. Stratonovich [20] attacked this problem in accordance with the standard approach used in the study of Markov processes, namely, to find a partial differential equation for the transition probability density function. Stratonovich derived a nonlinear stochastic partial integro-differential equation which he asserted is obeyed by the conditional probability density for the current state of the first process given the record of observations of the second process. Somewhat later, the same equation was rederived by Kashyap [10].

Still later, Kushner published a paper [11] in which he claimed Stratonovich's results were in error because of a failure to take into account certain second order terms. This issue arises because the observed process contains additive white noise. As a consequence, the partial differential equation for the conditional density effectively contains a white noise forcing term. This means that certain mathematical pathologies arise

in dealing with this equation. In a revised version[12] of his paper, Kushner clarified this issue somewhat and effectively recognized that the discrepancy between his results and those of Stratonovich is related to the divergence between the Ito and Stratonovich stochastic calculi[5]. In a correspondence item, Bucy[13] emphasized the importance of the Ito calculus and presented some results which are equivalent to some of those of Kushner.

The subject of nonlinear filtering has been the topic of a number of reports and papers (See References 14-27) since Bucy's note[13]. The purpose of this Section is to present a tutorial discussion of the main results which have been obtained so far, attempting to resolve some questions which have arisen concerning these works, and to discuss some suggested approximations intended to simplify the calculation of the nonlinear minimum variance estimate.

2.2 THE MODEL AND THE PROBLEM

In this Section, the problem and the main results will be stated first in Stratonovich form, in order to be as intelligible as possible to those not familiar with the Ito Calculus. Following that, the same material will be presented in Ito form. See Reference [5] for an explanation of the Stratonovich and the Ito stochastic calculus. For simplicity, only the case of scalar-valued random processes will be treated in detail, since the generalization to the vector-valued case is quite direct.

Consider a plant described by a nonlinear differential equation with a white noise forcing term:

$$\dot{x}(t) = f\,(x(t),\ t) + n_1\,(t) \tag{2.1}$$

The state $x(t)$ is not directly observable. Rather, in general one observes a nonlinear time-varying function of $x(t)$ which is corrupted by more white noise:

$$y(t) = h\,(x(t),\ t) + n_2\,(t) \tag{2.2}$$

At the initial time t_o, the initial state has a probability distribution which has a density function denoted as $p_o(\xi)$. One knows the function $p_o(\xi)$, and also one knows the history of $y(\tau)$ over the interval $t_o \le \tau \le t$. Denote the record of this function over the whole interval as $y_{[t_o,\ t]}$.

In Reference [5], it is explained that the Markov stochastic process generated by equation (2.1) is characterized by its transition density function $p(\xi,t|\eta,s)$. This density function is obtained as a solution to the Fokker-Planck-Kolmogorov partial differential equation.

The problem is to find out how this density function is modified when one conditions only on the information contained in the record of observations $y_{[t_o,t]}$ rather than on any exact knowledge of the state at some time prior to t.

Define the conditional density function $p(\xi,t|y_{[t_o,t]})$ by

$$p(\xi,t|y_{[t_o,t]})d\xi = \Pr\left\{\xi \leq x(t) < \xi + d\xi \,|\, y_{[t_o,t]}\right\}. \qquad (2.3)$$

The problem, then, is to compute this function.

2.3 THE PARTIAL DIFFERENTIAL EQUATION FOR THE CONDITIONAL DENSITY

We assume explicitly that the noise terms n_1 and n_2 in (2.1) and (2.2) are gaussian and independent, and that

$$E\left\{n_1(t)\right\} = E\left\{n_2(t)\right\} = 0 \;;\; E\left\{n_1(t)n_2(\tau)\right\} = 0 \;\forall t, \tau;$$

$$E\left\{n_1(t)n_1(\tau)\right\} = \delta(t-\tau) \;;\; E\left\{n_2(t)n_2(\tau)\right\} = \delta(t-\tau) \qquad (2.4)$$

Let $p(\xi,t|\eta,t_o)$ be the transition function for the Markov process generated by (2.1), and let p_o be the probability density for the initial state. For convenience, define the function $q(\xi,t)$ by

$$q(\xi,t) = \int_{-\infty}^{\infty} p(\xi,t|\eta,t_o)\, p_o(\eta)\, d\eta \qquad (2.5)$$

Let E_x denote mathematical expectation taken over the space of sample paths of the x(t) process alone, i.e., E_x denotes averaging over only the ensemble of realizations of solutions of (2.1). Let $\beta(x_{[t_o,t]}, y_{[t_o,t]})$

denote the quantity

$$\beta(x_{[t_o,t]}, y_{[t_o,t]}) = \exp\left\{\int_{t_o}^{t}\left[h(x(\tau),\tau)y(\tau) - \frac{1}{2}h^2(x(\tau),\tau)\right]d\tau\right\} \quad (2.6)$$

Let $r(\xi, y_{[t_o,t]}, t)$ denote the quantity

$$r(\xi, y_{[t_o,t]}, t) = E_x\left\{\beta\left(x_{[t_o,t]}, y_{[t_o,t]}\right) \middle| x(t) = \xi\right\} q(\xi,t) \quad (2.7)$$

It is shown in References [14] and [19] that the sought-for quantity in (2.3) may be written

$$p(\xi,t|y_{[t_o,t]}) = \frac{r(\xi, y_{[t_o,t]}, t)}{\int_{-\infty}^{\infty} r(\xi, y_{[t_o,t]}, t)\, d\xi} \quad (2.8)$$

So far, all the equations we have written have the correct form regardless of whether the Ito or the Stratonovich calculus is to be used. However, from this point on it makes a difference whether the integral in the exponent in (2.6) is interpreted as an Ito integral or a Stratonovich integral. We will proceed from this point following the Stratonovich approach.

It is shown in Reference [19] that $r(\xi, y_{[t_o,t]}, t)$ satisfies the forward equation

$$\frac{\partial r}{\partial t} = -\frac{\partial}{\partial \xi}\left[f(\xi,t)r\right] + \frac{1}{2}\frac{\partial^2 r}{\partial \xi^2} + \left[h(\xi,t)y(t) - \frac{1}{2}h^2(\xi,t)\right]r \quad (2.9)$$

Now define

$$\psi(t) = \int_{-\infty}^{\infty} r(\xi, y_{[t_o,t]}, t)\, d\xi \quad (2.10)$$

Equation (2.8) becomes

$$p(\xi,t|y_{[t_o,t]}) = \psi^{-1}(t)\, r(\xi,y_{[t_o,t]},\, t) \tag{2.11}$$

Employing the usual rules of calculus, one has

$$\frac{\partial p}{\partial t} = \psi^{-1}(t)\ \frac{\partial r}{\partial t} - \psi^{-2}(t)\ \dot{\psi}(t)\ \ r \tag{2.12}$$

After some manipulation and the use of (2.8) - (2.12), one arrives at the following forward equation for $p(\xi,t|y_{[t_o,t]})$:

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial \xi}\left[f(\xi,t)p\right] + \frac{1}{2}\frac{\partial^2 p}{\partial \xi^2} + \left[h(\xi,t)y(t) - \frac{1}{2}h^2\ (\xi,t)\right] p$$

$$- p\int_{-\infty}^{\infty}\left[h(\xi,t)y(t) - \frac{1}{2}h^2(\xi,t)\right] p(\xi,t|y_{[t_o,t]})\, d\xi \tag{2.13}$$

In general, the solution of the nonlinear filtering problem requires that (2.13) be solved in real time, as the data y(t) is received. The boundary conditions on (2.13) are

$$\lim_{t\to t_o} p(\xi,t|y_{[t_o,t]}) = p_o(\xi) \tag{2.14}$$

where $p_o(\xi)$ is the apriori density for the initial state at time t_o, and

$$\lim_{|\xi|\to\infty} p(\xi,t|y_{[t_o,t]}) = 0 \tag{2.15}$$

The minimum mean-squared error Bayes estimate of the current state is given by

$$\hat{x}(t) = E\left\{x(t)|y_{[t_o,t]}\right\} = \int_{-\infty}^{\infty} \xi\ p(\xi,t|y_{[t_o,t]})\ d\xi \tag{2.16}$$

In general, it is not possible to find an ordinary differential equation obeyed by $\hat{x}(t)$ and forced by $y(t)$ which can be solved directly without first finding $p(\xi,t|y_{[t_o,t]})$, because direct differentiation of (2.16) leads to an ordinary differential equation for $\hat{x}(t)$ which involves the unknown higher moments of $p(\xi,t|y_{[t_o,t]})$. In the special case when both $f(\xi,t)$ and $h(\xi,t)$ are <u>linear</u> in ξ, then $p(\xi,t|y_{[t_o,t]})$ will be gaussian. In this case, one may assume

$$p(\xi,t|y_{[t_o,t]}) = \left[2\pi\sigma^2(t)\right]^{-1/2} \exp\left\{-\frac{[\xi - \hat{x}(t)]^2}{2\sigma^2(t)}\right\} \tag{2.17}$$

Substitution of (2.17) into (2.13) and matching coefficients of powers of $(\xi-\hat{x})$ leads eventually to the well-known Kalman-Bucy filter equations for $\hat{x}(t)$ and $\sigma^2(t)$. The same approach applies when $\hat{x}(t)$ is a vector and $\sigma^2(t)$ is a matrix. See Reference [15] for a typical derivation of the Kalman filter using this approach. In carrying out this approach, of course, the differentiation of (2.17) with respect to t and substitution in (2.13) is all done according to the rules of ordinary calculus, i.e., $\hat{x}(t)$ is treated just as if it were a deterministic function of t rather than a random process.

2.4 THE ITO FORMULATION

In order to recast these results in Ito form, it is necessary to begin by rewriting equation (2.3). Let

$$w_2(t) = \int_{t_o}^{t} n_2(\tau)d\tau \quad ; \quad z(t) = \int_{t_o}^{t} y(\tau)\, d\tau \quad , \tag{2.18}$$

and rewrite (2.2) as

$$dz(t) = h(x(t),t)\, dt + dw_2(t) \quad . \tag{2.19}$$

The segment $z_{[t_o,t]}$ contains exactly the same information as $y_{[t_o,t]}$, so now

one conditions on $z_{[t_o,t]}$ rather than $y_{[t_o,t]}$. In place of the quantity defined in (2.3), we write $p_I(\xi,t|z_{[t_o,t]})$. The subscript I stands for Ito.

The quantity defined in equation (2.6) is now replaced by

$$\beta_I(x_{[t_o,t]}, z_{[t_o,t]}) = \exp\left\{\int_{t_o}^{t} h(x(\tau),\tau)dz(\tau) - \frac{1}{2}\int_{t_o}^{t} h^2(x(\tau),\tau)d\tau\right\} \tag{2.20}$$

In (2.20), the first integral in the exponent is to be interpreted as an Ito integral.

In analogy to (2.7), define

$$r_I(\xi,z_{[t_o,t]},t) = E_x\left\{\beta_I(x_{[t_o,t]}, z_{[t_o,t]}) \mid x(t) = \xi\right\} q(\xi,t) \tag{2.21}$$

In analogy to (2.8), we now have

$$p_I(\xi,t|z_{[t_o,t]}) = \frac{r_I(\xi,z_{[t_o,t]},t)}{\int_{-\infty}^{\infty} r_I(\xi,z_{[t_o,t]},t)d\xi} \tag{2.22}$$

It is shown in Reference [19] that by applying the rules of Ito calculus to (2.20) and (2.21), it is possible to obtain the following forward equation for $r_I(\xi,z_{[t_o,t]},t)$:

$$d_t r_I = \left\{-\frac{\partial}{\partial\xi}\left[f(\xi,t)r_I\right] + \frac{1}{2}\frac{\partial^2 r_I}{\partial\xi^2}\right\}dt + h(\xi,t)r\,dz(t) \tag{2.23}$$

It should be noted that equation (2.23) is not obtained from equation (2.9) just by making the formal change of variables indicated by (2.18) and (2.19). The two equations are different because the partial differential with respect to t in (2.23) is an Ito differential, whereas the partial derivative with

respect to t in (2.9) is understood in the Stratonovich sense. Now define

$$\psi_I(t) = \int_{-\infty}^{\infty} r_I(\xi, z_{[t_o,t]}, t)d\xi \tag{2.24}$$

Equation (2.22) becomes

$$p_I(\xi, t\, z_{[t_o,t]}) = \psi_I^{-1}(t)\, r_I(\xi, z_{[t_o,t]}, t) \tag{2.25}$$

Employing the Ito stochastic differential rule along with (2.24) and (2.25), one obtains

$$d_t p_I = \psi_I^{-1} d_t r - \psi_I^{-2} r_I d_t \psi_I + \left[\psi_I^{-3} r\left(\int_{-\infty}^{\infty} h(\xi,t) r d\xi\right)^2 - \psi_I^{-2} hr \int_{-\infty}^{\infty} h(\xi,t) r\, d\xi\right] dt \tag{2.26}$$

Define

$$\hat{h}(t) = \int_{-\infty}^{\infty} h(\xi,t)\, p_I(\xi,t|z_{t_o,t})d\xi = \psi^{-1}(t) \int_{-\infty}^{\infty} h(\xi,t)\, r(\xi, z_{[t_o,t]}, t)d\xi \tag{2.27}$$

Equation (2.26) may be rewritten

$$d_t p_I = \psi_I^{-1} d_t r - \psi_I^{-2} r_I d_t \psi_I + p_I \hat{h}(t) \left[\hat{h}(t) - h(\xi,t)\right] dt \tag{2.28}$$

Equation (2.28) should be compared with the corresponding Stratonovich form, which is equation (2.12). After some manipulation and the use of (2.23) - (2.25) and (2.28), one arrives at the Ito form of the forward equation for $p_I(\xi,t|z_{[t_o,t]})$:

$$d_t p_I = \left\{ - \frac{\partial}{\partial \xi}\left[f(\xi,t) p_I\right] + \frac{1}{2} \frac{\partial^2 p_I}{\partial \xi^2} \right\} dt + \left[h(\xi,t) - \hat{h}(t)\right] p_I \left[dz(t) - \hat{h}(t)\, dt\right] \tag{2.29}$$

Again in the special case when both $f(\xi,t)$ and $h(\xi,t)$ are linear in ξ, in exact analogy to (2.17) one may assume

$$p_I(\xi,t|z_{[t_o,t]}) = \left[2\pi\sigma^2(t)\right]^{-1/2} \exp\left\{-\frac{[\xi - \hat{x}(t)]^2}{2\sigma^2(t)}\right\} \tag{2.30}$$

Substitution of (2.30) into (2.29) and matching coefficients of powers of $(\xi-\hat{x})$ will again lead to exactly the same well-known Kalman-Bucy filter equations for $x(t)$ and $\sigma^2(t)$, <u>provided</u> one uses the <u>Ito calculus</u> for computing the stochastic differential with respect to t. This is a point which can lead to controversy (for example, see Reference [15]). For examples of such a derivation in the case where $\hat{x}$ is a vector and σ^2 is a matrix, see References [24], [25], and [26]. The extension of this derivation to the nonlinear case is also discussed in these references.

2.5 THE ONE-DIMENSIONAL LINEAR PROBLEM

In the linear time-invariant case, equations (2.1) and (2.2) become

$$\dot{x}(t) = a\,x(t) + n_1(t) \tag{2.31}$$

$$y(t) = c\,x(t) + n_2(t) \tag{2.32}$$

where a and c are constants.

The Stratonovich form of the partial differential equation for the conditional density is equation (2.13). In the present case, this equation becomes

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial \xi}(a\,\xi p) + \frac{1}{2}\frac{\partial^2 p}{\partial \xi^2} + [c\xi y(t) - \frac{1}{2}c^2\xi^2]\,p - p\int_{-\infty}^{\infty}[c\xi y(t) - \frac{1}{2}c^2\xi^2]\,p\,d\xi \tag{2.33}$$

As suggested in Section 2.3, we assume a solution of the form of equation (2.17), and substitute it into the partial differential equation in order to obtain ordinary differential equations for $\hat{x}(t)$ and $\sigma^2(t)$. When this is done, the resulting equations are

$$\frac{d\hat{x}(t)}{dt} = a\hat{x}(t) + c\sigma^2(t)\,[y(t) - c\hat{x}(t)] \tag{2.34}$$

$$\frac{d}{dt}\,[\sigma^2(t)] = 1 + 2a\sigma^2(t) - c^2\sigma^4(t) \tag{2.35}$$

A boundary condition must be imposed on the solution of the partial differential equation, as given in equation (2.14). In the present case, let us take

$$p_o(\xi) = \delta(\xi - x_o) \tag{2.36}$$

i.e., the initial state $x(t_o) = x_o$ is non-random and known exactly.

In order to satisfy this boundary condition, we must apply appropriate corresponding initial conditions to equations (2.34) and (2.35) above. The correct initial conditions are

$$\hat{x}(t_o) = x_o \tag{2.37}$$

$$\sigma^2(t_o) = 0 \tag{2.38}$$

Equations (2.34) and (2.35), subject to (2.37) and (2.38), may be solved by standard methods. The solutions are found to be

$$\sigma^2(t) = \frac{\sinh \lambda\,(t - t_o)}{\lambda \cosh \lambda(t - t_o) - a \sinh \lambda(t - t_o)} \tag{2.39}$$

$$\hat{x}(t) = \frac{\lambda x_o + c \int_{t_o}^{t} \sinh \lambda(\tau - t_o)\, y\,(\tau)d\tau}{\lambda \cosh \lambda(t - t_o) - a \sinh \lambda(t - t_o)} \tag{2.40}$$

where

$$\lambda = \sqrt{a^2 + c^2} \tag{2.41}$$

Equation (2.40) makes it clear that a sufficient statistic on the past history of the data is the linear functional

$$\int_{t_o}^{t} \sinh \lambda(\tau - t_o)\, y(\tau)d\tau$$

2.6 A ONE-DIMENSIONAL NON-LINEAR EXAMPLE

In order to understand the nonlinear case, it is convenient to have at hand a nonlinear problem for which the exact solution is known. Using the preceding results, we can manufacture such a problem. Our idea is to construct a nonlinear problem which is actually equivalent to the preceding linear one by making a suitable nonlinear transformation of the state variable in both the plant and the observation equations.

Consider the nonlinear change of variables

$$v(t) = \sinh [x(t)] \tag{2.42}$$

This particular nonlinearity has already been discussed in pp. 14-20 of Reference [5], so we expect that its introduction is going to cause a divergence between the Stratonovich and Ito forms of the nonlinear filtering equations. Although this makes the problem more subtle, no actual paradoxes will arise so long as we remember the content of Reference [5].

Let us continue using the Stratonovich form of the equations. The plant equation (2.1) now becomes

$$\dot{v}(t) = a\sqrt{1 + v^2(t)}\ \sinh^{-1}[v(t)] + \sqrt{1 + v^2(t)}\ n_1(t) \tag{2.43}$$

The observation equation (2.2) now becomes

$$y(t) = c \sinh^{-1}[v(t)] + n_2(t) \tag{2.44}$$

Thus, equations (2.43) and (2.44) along with the nonlinear change of variables

$$x(t) = \sinh^{-1}[v(t)] \tag{2.45}$$

are completely equivalent to equations (2.31) and (2.32). However, we can pretend that we were given <u>only</u> eqs. (2.43) and (2.44), and asked to find the optimal nonlinear filter for this model.

Let us write down the exact answer. Since the minimum variance Bayes estimate is the conditional mean, we have

$$\hat{v}(t) = E\left\{v(t) \mid y_{[t_o,\, t]}\right\}$$

$$= E\left\{\sinh\left[x(t)\right] \mid y_{[t_o,\, t]}\right\}$$

$$= \int_{-\infty}^{\infty} \sinh \xi \; p(\xi, t \mid y_{[t_o,\, t]})\, d\xi$$

$$= \left[2\pi\, \sigma^2(t)\right]^{-1/2} \int_{-\infty}^{\infty} \sinh \xi \; e^{-\frac{\left[\xi - \hat{x}(t)\right]^2}{2\sigma^2(t)}}\, d\xi \qquad (2.46)$$

In (2.46), $\sigma^2(t)$ and $\hat{x}(t)$ are of course still given by (2.39) and (2.40) respectively.

Making use of a known result, we have at once from (2.46) that

$$\left[2\pi\, \sigma^2(t)\right]^{-1/2} \int_{-\infty}^{\infty} \sinh \xi \; e^{-\frac{\left[\xi - \hat{x}(t)\right]^2}{2\sigma^2(t)}}\, d\xi$$

$$= e^{1/2\, \sigma^2(t)} \sinh \hat{x}(t) \qquad (2.47)$$

For emphasis, let us review what we have done up to this point. We know the optimal estimate $\hat{x}(t)$ for the linear filtering problem defined by (2.31) and (2.32). We make the nonlinear change of variables (2.42). By (2.46) and (2.47), we therefore find that the optimal estimate $\hat{v}(t)$ is

$$\hat{v}(t) = e^{1/2\, \sigma^2(t)} \sinh\left[\hat{x}(t)\right] \qquad (2.48)$$

This is, therefore, the exact solution to the nonlinear filtering problem defined by (2.43) and (2.44).

2.7 THE FILTER EQUATIONS

We now wish to explore whether this optimal estimate $\hat{v}(t)$ can be obtained as the solution to some nonlinear filtering equations analogous to the linear filtering equations (2.34) and (2.35) for $\hat{x}(t)$. In the present case because of the fact that the quantity

$$\int_{t_o}^{t} \sinh \lambda(\tau - t_o)\, y(\tau)d\tau$$

is a sufficient statistic for computing $\hat{x}(t)$, it is also a sufficient statistic for computing $\hat{v}(t)$. Consequently, in the present case there will exist some exact nonlinear filtering differential equations whose solution is $\hat{v}(t)$.

This is to be contrasted with the general nonlinear filtering problem in which there does not exist a sufficient statistic, and consequently there are no exact (finite dimensional) nonlinear filtering equations. In the general case, the only exact equation would be the partial differential equation for the conditional probability density itself.

The most direct way to find the exact equations for $\hat{v}(t)$ is simply to differentiate (2.48) directly, making use of (2.34) and (2.35). In doing this differentiation, we must be careful to recall the discussion in Reference [5]. If we differentiate (2.48) according to the rules of the Ito stochastic calculus, then we will get an Ito stochastic differential equation which must be solved by the rules of the Ito calculus in order to get the right answer. On the other hand, if we differentiate (2.48) according to the rules of ordinary calculus, the resulting differential equation will be in Stratonovich form and must be solved accordingly.

We emphasize that we really are free to follow either route at this point. The correct form for the conditional density for x(t) is gaussian, with variance $\sigma^2(t)$ and mean $\hat{x}(t)$, as given by (2.39) and (2.40) respectively. This gaussian conditional density is simultaneously the Stratonovich solution of equation (2.13) and the Ito solution of equation (2.29), keeping in mind that, in the present case,

$$f(\xi,t) = a\xi \quad ; \quad h(\xi,t) = c\xi \tag{2.49}$$

Furthermore, in the present case the Ito and the Stratonovich forms of the equations obeyed by $\hat{x}(t)$ and $\sigma^2(t)$ coincide; either way the correct equations are (2.34) and (2.35). Strictly speaking, of course, the change of notation represented by equations (2.18) and (2.19) should be made in order to cast equation (2.34) into the proper Ito form, but this does not alter the truth

of the statements we have just made.

Let us continue, using the Stratonovich calculus for purposes of exposition.

Differentiating equation (2.48) by the rules of ordinary calculus, we have

$$\frac{d\hat{v}(t)}{dt} = e^{\frac{1}{2}\sigma^2(t)} \sinh[\hat{x}(t)] \frac{d}{dt}[\tfrac{1}{2}\sigma^2(t)] + e^{\frac{1}{2}\sigma^2(t)} \cosh[\hat{x}(t)] \frac{d\hat{x}(t)}{dt} \tag{2.50}$$

Substitution of (2.34) and (2.35) into (2.50), simplification, and use of the inverse of (2.48), namely

$$\hat{x}(t) = \sinh^{-1}\left[e^{-\frac{1}{2}\sigma^2(t)}\hat{v}(t)\right] \tag{2.51}$$

yields finally

$$\frac{d\hat{v}(t)}{dt} = \sqrt{e^{\sigma^2(t)} + \hat{v}^2(t)} \left\{ a \sinh^{-1}\left[e^{-\frac{1}{2}\sigma^2(t)}\hat{v}(t)\right] + c\sigma^2(t)\left(y(t) - c \sinh^{-1}\left[e^{-\frac{1}{2}\sigma^2(t)}\hat{v}(t)\right]\right)\right\} + \frac{1}{2}\hat{v}(t)\,[1 + 2a\,\sigma^2(t) - c^2\sigma^4(t)] \tag{2.52}$$

Equation (2.52) along with equation (2.35) thus constitutes the Stratonovich form of the optimal nonlinear filter for the nonlinear filtering problem represented by the Stratonovich eqs. (2.43) and (2.44).

2.8 APPROXIMATIONS OF THE ESTIMATION EQUATIONS

The solution of the general minimum variance nonlinear estimation problem is very difficult to mechanize. Consequently, it appears necessary to search for sub-optimal estimators which are easier to compute, and which still possess near-minimum estimation error. Since the exact minimum variance sequential estimator is known for the linear problem with gaussian nois

namely the Kalman-Bucy filter, one approach to sub-optimal estimation is to attempt to find correction terms for the Kalman-Bucy filter for the case of slight non-linearity. The remainder of Section 2 will be devoted to a discussion of this approach to sub-optimal filtering.

The literature on this problem may be classified in several different ways, according to the basic approach adopted for the estimation procedure and the underlying assumptions about the model. The approach to the problem may be either statistical or non-statistical. In statistical procedures, the existence of certain underlying probability distributions must be postulated, and statistical-type estimates such as maximum likelihood or minimum-variance Bayes are sought. In non-statistical procedures, the problem is viewed as one of choosing an estimate which achieves the optimum curve fit to the actual data according to some criterion such as least-squares or Chebyshev. The underlying model may be either static or dynamic. For dynamic systems, the state space may be either continuous or discrete, and time may be taken either as a continuous variable or as a discrete variable.

The compilation of a bibliography which is exhaustive in all of these categories would be an enormous task, and consequently will not be attempted here. A few isolated references to the literature will be given to illustrate some of the categories. Most of the cited references themselves contain fairly extensive bibliographies, so that the reader interested in a particular area should at least be able to find a lead.

It should be pointed out that much more of the published literature is devoted to theoretical formulations than to actual numerical solutions of particular problems. Consequently, it is not really possible to make a meaningful comparison among the various possible approaches and models which might be applied in a given situation. Therefore, there are only rather vague guidelines available to aid the uninitiated engineer in choosing a method of attack for his particular problem. Adding to the bewilderment is the fact that established workers in the field tend to have developed strong preferences for a particular approach, to the exclusion of all others, which they can justify only on a subjective basis.

Any book on statistical estimation theory should serve for an initial

orientation for static problems. One starting point might be the chapter by Balakrishnan in Reference [28].

References [8] and [6] treat the subject as a least squares curve fitting problem in discrete time. Section 6 is an interesting discussion of one aspect of the subject from this same standpoint.

Reference [29] treats the subject as a problem of computing maximum likelihood estimates in continuous time, although it is also possible to interpret the approach taken there as continuous time least squares curve fitting.

Reference [21] discusses a problem of continuous-time minimum-variance Bayes estimation of the state of a system with discrete state space. This reference shows particularly clearly how the discrepancy between the Ito and Stratonovich calculi appears in continuous-time stochastic estimation problems, a difficulty which does not arise in discrete-time problems.

Finally, References [22]-[23] provide good starting points for investigating the problem of finding minimum-variance Bayes estimates of the state of continuous-time nonlinear dynamic systems with continuous state spaces.

2.9 EXAMPLES OF SIMPLIFYING APPROXIMATIONS

A close study of the simple example discussed in Section 2.7 should enable us to resolve some questions about the validity of various plausible approximations, at least in this special case.

The crudest approximation is complete linearization. This means we expand all nonlinearities in the model, eqs. (2.43) and (2.44), in Taylor series, and retain only first order terms. The effect of this is to yield approximate equations for the model which look exactly like the original linear equations (2.31) and (2.32), with $x(t)$ replaced by $v(t)$. Naturally, the exact Kalman filter for this approximate linearized model turns out to be given by equations (2.34) and (2.35) with $\hat{x}(t)$ replaced by $\hat{v}(t)$.

This approximation is tantamount to making the approximation

$$\sinh [x(t)] \approx x(t) \tag{2.53}$$

in (2.42) and so naturally we come out with the approximate estimate

$$\hat{v}(t) \approx \hat{x}(t) \tag{2.54}$$

It should be pointed out that the above procedure is not equivalent to first deriving the exact nonlinear filter (2.52) and then linearizing it. This can be seen by comparing eqs. (2.42) and (2.48). Whereas the linearization of (2.42) yields (2.54), the linearization of (2.48) yields

$$\hat{v}(t) \approx e^{\frac{1}{2}\sigma^2(t)}\,\hat{x}(t) \tag{2.55}$$

Consequently, linearization of (2.52) must lead to (2.55) rather than (2.54).

Another approximation frequently suggested is the so-called maximum likelihood filter (e.g., see Reference [29]). Although this filter can be obtained through a formal derivation from basic assumptions, it amounts to just retaining the form of the Kalman filter, eqs. (2.34) and (2.35), but replacing the linear terms in eq. (2.34) by the corresponding model nonlinearities from (2.43) and (2.44). In the present case, the result is of the form

$$\frac{d\hat{v}(t)}{dt} = a\sqrt{1+\hat{v}^2(t)}\,\sinh^{-1}[\hat{v}(t)]$$

$$+ c\sigma^2(t)\left\{y(t) - c\sinh^{-1}[\hat{v}(t)]\right\} \tag{2.56}$$

Although this resembles (2.52) somewhat, it is simpler.

Incidentally, it should be pointed out that according to maximum likelihood estimation theory, if it is known that $\hat{x}(t)$ is the maximum likelihood estimate of $x(t)$, and if it is known that equation (2.42) connects $v(t)$ and $x(t)$, then necessarily the maximum likelihood estimate of $v(t)$ is given by

$$\hat{v}(t) = \sinh[\hat{x}(t)] \tag{2.57}$$

Thus, we may compare equations (2.48) and (2.57) to see the theoretical difference between the minimum variance Bayes estimate and the maximum likelihood estimate.

This comparison is meaningful in the present case only because of the fact that, for the linear, gaussian problem to which the Kalman filter applies

exactly, the minimum variance and maximum likelihood estimates coincide. Thus, the $\hat{x}(t)$ given by (2.40) is simultaneously the minimum variance and maximum likelihood estimate of the $x(t)$ generated by (2.31).

Incidentally, equation (2.56) should not be taken too literally. It is only intended to indicate the form of the filter equations which results when the procedure described in Reference [29] is applied to the nonlinear system (2.43). However, this procedure encounters difficulties which do not occur in Reference [29], because of the coefficient $\sqrt{1 + v^2(t)}$ multiplying $n_1(t)$ in (2.43). As pointed out in Reference [5], the presence of a state dependent coefficient multiplying white noise immediately leads to the Stratonovich-Ito divergence. As a matter of fact, the correct form of the likelihood functional, given by eq. (5) of Reference [29], is no longer clear when there are state dependent nonlinearities. Therefore we should say only that the procedure described in Reference [29], when extrapolated in a plausible way, appears to lead to a filter equation having the form represented by eq. (2.56).

For the same reason, the coefficient $\sigma^2(t)$ in equation (2.56) is probably not the same function given by eq. (2.39) and obtained as a solution of eq. (2.35). The procedure described in Reference [29] can be made to lead to a differential equation of Ricatti type for this coefficient, but it will contain terms which depend on $\hat{v}(t)$, thus coupling the $\sigma^2(t)$ and $\hat{v}(t)$ equations, because of the presence of the factor $\sqrt{1 + v^2(t)}$ multiplying $n_1(t)$ in (2.43), as just mentioned.

For reasons given earlier, it seems correct to believe that the $\hat{v}(t)$ given by (2.57) is the correct maximum likelihood estimate of $v(t)$. However, direct differentiation of (2.57) and use of (2.34) does not yield (2.56). Again, there is a discrepancy caused by the factor $\sqrt{1 + \hat{v}^2(t)}$.

A detailed investigation of the nonlinear filtering problem when the noise enters with a state-dependent coefficient appears difficult. Such an investigation was felt to exceed the scope of the present effort. Consequently, we content ourselves here with pointing out some of the questions which arise in such a case.

3. MAXIMUM LIKELIHOOD ESTIMATION, CONTROL AND ERROR ANALYSIS

3.1 INTRODUCTION

Sequential orbit determination must be carried out in real time for space missions so that control (guidance) corrections can be applied. The method presently employed is straightforward. A model of the motion of the spacecraft is constructed, and probability density functions describing the random behavior of initial conditions, trajectory disturbances and data noise are postulated. Then given a data record at time t_k, consisting of all data obtained up to t_k, estimates of the initial conditions and disturbances are calculated by applying an iterative Newton-Raphson technique to find the root of the differentiated likelihood function. The control (guidance) is then determined by treating these estimates as though they were the true values for a deterministic system, and error analysis is performed by linearizing the equations of motion about the estimated trajectory. This intuitively reasonable procedure has been used with much success. It is the purpose of this Section to discuss the theoretical justification for the presently employed estimation, control, and error analysis algorithms. The results will not apply to the most general form of the stochastic control problem, and an exception will be discussed in Section 3.2.

It should be noted that a sequential estimation procedure of this type does not permit the control to be applied immediately after receiving the last data point, because some time is required in order to process the total data record. This consideration has thus far not been important in orbit determination and guidance problems, where efficient numerical techniques have been developed. Considering the capability of present-day computers, and the many simplifications which can be introduced if one is willing to accept some approximation error, there is no reason to believe that the computational time lag will be a limiting factor for future missions. In any case, this is a practical problem which can only be discussed by referring to a specific application, recognizing that the alternatives to the maximum likelihood algorithm are the numerical solution of a nonlinear partial differential equation, or some approximation of that solution, or dynamic programming.

3.2 OPEN LOOP MAXIMUM LIKELIHOOD CONTROL

The combined estimation and control problem might be best understood by first considering a relatively simple example of open-loop maximum likelihood control, where the control correction is based upon only apriori information rather than data gathered during the mission.

Suppose that at time t_o a single correction impulse vector u is to be applied so as to cause the "most likely" state at the final time T to be a desired value. Let the equations of motion be

$$\frac{dx}{dt} = f(x, t) \qquad t_o < t \leq T \tag{3.1}$$

where x is the state vector. The initial condition at $t_o^{(+)}$ is x_o, where x_o is a Gaussian vector with covariance matrix Λ_o and mean equal to $[\mu + Ku]$. The K is supposed to be a given matrix. Thus the control u impulsively changes the state at t_o according to

$$x_o = x(t_o^{(+)}) = x(t_o^{(-)}) + Ku \tag{3.2}$$

where $x(t_o^{(-)})$ is a Gaussian vector with mean equal to μ and covariance Λ_o. Assuming a one-to-one mapping of x_o to the final state $x_T = x(T)$ of the form $x_T = g(x_o)$, the probability density function of x_T is

$$p_T(x_T) = c \exp\left\{ - \frac{1}{2} [q(x_T)]^T \Lambda_o^{-1} [q(x_T)] \right\} \left| \frac{\partial x_o}{\partial x_T} \right| \tag{3.3}$$

where c is the coefficient of the Gaussian density function of x_o, and

$$q(x_T) = g^{-1}(x_T) - (Ku + \mu) \tag{3.4}$$

$$\left| \frac{\partial x_o}{\partial x_T} \right| = \text{determinant } \left| \frac{\partial g^{-1}(x_T)}{\partial x_T} \right| \tag{3.5}$$

Define the likelihood function

$$L(x_T) = - \ln p_T(x_T) = \frac{1}{2} q(x_T)^T \Lambda_o^{-1} q(x_T) - \ln \left| \frac{\partial x_o}{\partial x_T} \right| - \ln c \tag{3.6}$$

But $|\frac{\partial x_o}{\partial x_T}|$ is the inverse determinant of the state transition matrix $|\frac{\partial x_T}{\partial x_o}|$, and it can be shown that (see [30], pp. 28)

$$|\frac{\partial x_o}{\partial x_T}| = \exp\left\{-\int_{t_o}^{T} \text{trace } [\frac{\partial f}{\partial x}(t;\, x_o(x_T))]dt\right\} \tag{3.7}$$

Then the most likely value of x_T is that value $\hat{x}_T$ which satisfies

$$\frac{\partial L(x_T)}{\partial x_T} = 0 = \Lambda_o^{-1}\, q(x_T) + \frac{\partial}{\partial x_T} \int_{t_o}^{T} \text{trace } [\frac{\partial f}{\partial x}]\, dt \tag{3.8}$$

But for hamiltonian systems where the state transition matrix is symplectic (see [31], pp. 306) it can be shown that $|\frac{\partial x_T}{\partial x_o}| = 1$ and hence trace $[\frac{\partial f}{\partial x}] = 0$. In this case we have $q(\hat{x}_T) = 0$, and

$$Ku = g^{-1}(\hat{x}_T) - \mu \tag{3.9}$$

In other words, for any desired final state $\hat{x}_T$ we can find the control u from (3.9), just as though the initial conditions were not random. (i.e., hamiltonian systems can be treated as though they were deterministic) Although this property is also assumed in present orbit determination work, where data is processed and an estimate of x_o is obtained before the control is applied, it is not at all clear that such a procedure is theoretically justified. One purpose of the discussion to follow is to show that under certain conditions, such a separation of the estimation and control problems is indeed legitimate.

Clearly the simple result (3.9) does not apply when trace $[\frac{\partial f}{\partial x}] \neq 0$, in which case a statistical "bias" must be introduced. This should not be an unexpected conclusion, for a similar result applies if one sought to control the expected value $E[x_T]$. For example, in the simple case

$$\frac{dx_1}{dt} = \alpha = f_1$$

$$\frac{dx_2}{dt} = 0 = f_2$$

with $x_2(0) = x_1^2(0)$

$$x_1(T) = \alpha T + x_1(0)$$

$$x_2(T) = x_2^2(0)$$

and

$$E(x_2(T)) = \sigma_1^2$$

where σ_1^2 is the apriori variance of $x_1(0)$. Note that we have

$$\text{trace } \left(\frac{\partial f}{\partial x}\right) = \text{trace } \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0$$

and even so the statistical bias σ_1^2 occurs. One should not apply the maximum likelihood algorithm in this case, however, for the assumption of a one-to-one mapping between x_o and x_T is violated.

3.3 A MAXIMUM LIKELIHOOD ESTIMATION AND CONTROL TECHNIQUE

In this Section we shall formulate a general maximum likelihood estimation and control technique which leads to a algorithm which is consistent with present practice in orbit determination and guidance. Consider a discrete, nonlinear dynamic system described by the vector differential equation

$$dx(i) = f(x(i), u_i, t_i)\, dt + G(t_i)\, dw_i \qquad 0 \le t_i \le T \qquad (3.10)$$

where t_i is one of a closely spaced set of discrete times $\{0, t_1, t_2 \cdots t_i, \cdot\cdot t_{N-1}, t_N = T\}$, x(i) is the dynamic state at time t_i, dx(i) = x(i+1)-x(i), u_i is a constant control applied in the interval dt, and dw_i is the state disturbance applied in the interval dt.† To sidestep mathematical difficulties we shall assume that dx(i) is <u>defined</u> by (3.10), where dw(i) is state noise in the Stratonovich sense. At each time t_i data will be obtained which is of the form

$$z_i = h(x(i), t_i) + n_i \tag{3.11}$$

where n_i is data noise. For simplicity we shall assume z_i is a scalar. We assume that the control u_i applied at each time t_i is a function of the entire data record up to t_i, where this function is supposed to be specified apriori. Defining the data vector

$$z^T(i) = [z_o, z_1, z_2 \cdot\cdot z_{i-1}, z_i] \tag{3.12}$$

we have

$$u^T(z) = [u_o(z(o)), u_1(z(1)), u_2(z(2)) \cdots u_N(z(N))] \tag{3.13}$$

Let the total state of the system be

$$y_o^T = [x_o, dw_o, dw_1, dw_2, \cdots dw_N] \tag{3.14}$$

Since x(i) depends upon y_o and $\{u_o, u_1, \cdot\cdot u_{i-1}\}$, it follows that, for any specified function u(z), the u and z are implicitly complicated functions of y_o and $n^T = [n_o, n_1, \cdot\cdot n_N]$. In particular, if z(i) is written in the form

$$z(i) = H_i[y_o, u(i-1)] + n(i) \tag{3.15}$$

where H_i is a vector function and

†The state may include certain constant "acceleration" parameters x_k, in which case $dx_k = 0$.

$$n^T(i) = [n_o, n_1, n_2, \cdots n_i]$$

$$u^T(i-1) = [u_o, u_1, u_2, \cdots u_{i-1}] \tag{3.16}$$

then the partial derivatives of z(i) with respect to y_o are obtained by the recursive relation

$$\left[\frac{\partial z(i)}{\partial y_o}\right] = \left[\frac{\partial H_i}{\partial y_o}\right] + \left[\frac{\partial H_i}{\partial u(i-1)}\right]\left[\frac{\partial u(i-1)}{\partial z(i-1)}\right]\left[\frac{\partial z(i-1)}{\partial y_o}\right] \tag{3.17}$$

where

$$\left[\frac{\partial z(o)}{\partial y_o}\right] = \left[\frac{\partial H_o}{\partial y_o}\right] \tag{3.18}$$

Similarly,

$$\left[\frac{\partial z(i)}{\partial n_i}\right] = \left[\frac{\partial H_i}{\partial u(i-1)}\right]\left[\frac{\partial u(i-1)}{\partial z(i-1)}\right]\left[\frac{\partial z(i-1)}{\partial n_i}\right] + \left[\frac{\partial n(i)}{\partial n_i}\right] \tag{3.19}$$

where n_i is one component of n, and

$$\left[\frac{\partial z(o)}{\partial n_i}\right] = \begin{cases} 1 & \text{if } i = 0 \\ 0 & \text{if } i \neq 0 \end{cases}$$

From these expressions $\left[\frac{\partial u(i)}{\partial y_o}\right]$ and $\left[\frac{\partial u(i)}{\partial n_i}\right]$ can be calculated.

Suppose that an apriori probability density function of y_o and the total noise vector n = n(N) is available, denoted by $p_o(y_o, n)$. Let the final state be y_T^T, = $[x(T), dw_o, \cdots dw_N]$, and assume that there is a one-to-one mapping of (y_o, n) to (y_T, z), where z is the total data vector z(N). Then

$$y_T = g(y_o, u)$$

where u is the total vector u(N), which is a function of y_o and n.

Applying the inverse relationships, the apriori probability density function of the data and final state is

$$p_T(y_T,z) = p_o(y_o(y_T, z), n(y_T, z)) \left| \frac{\partial(y_o, n)}{\partial(y_T, z)} \right|$$

where $|..|$ is the determinant of the inverse Jacobian matrix

$$\begin{bmatrix} \left(\frac{\partial y_o}{\partial y_T}\right) & \left(\frac{\partial y_o}{\partial z}\right) \\ \left(\frac{\partial n}{\partial y_T}\right) & \left(\frac{\partial n}{\partial z}\right) \end{bmatrix} = \left\{ \begin{bmatrix} \left(\frac{\partial g}{\partial y_o}\right) & \left(\frac{\partial g}{\partial u}\right) \left(\frac{\partial u}{\partial z}\right) \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ \left(\frac{\partial z}{\partial y_o}\right) & \left(\frac{\partial z}{\partial n}\right) \end{bmatrix} \right\}^{-1}$$

$$= \begin{bmatrix} \left(\frac{\partial g}{\partial y_o}\right)^{-1} & -\left(\frac{\partial g}{\partial y_o}\right)^{-1}\left(\frac{\partial g}{\partial u}\right)\left(\frac{\partial u}{\partial z}\right) \\ -\left(\frac{\partial z}{\partial n}\right)^{-1}\left(\frac{\partial z}{\partial y_o}\right)\left(\frac{\partial g}{\partial y_o}\right)^{-1}\left(\frac{\partial z}{\partial n}\right) & \left(\frac{\partial z}{\partial n}\right)^{-1}\left[I + \left(\frac{\partial z}{\partial y_o}\right)\left(\frac{\partial g}{\partial y_o}\right)^{-1}\left(\frac{\partial g}{\partial u}\right)\left(\frac{\partial u}{\partial z}\right)\right] \end{bmatrix} \quad (3.20)$$

where I is the identity matrix. Then

$$\left| \frac{\partial(y_o, n)}{\partial(y_T, z)} \right| = \left| \frac{\partial g}{\partial y_o} \right|^{-1} \left| \frac{\partial z}{\partial n} \right|^{-1} \quad (3.21)$$

But from (3.19)

$$\left[\frac{\partial z}{\partial n}\right] = I + \left[\left(\frac{\partial H}{\partial u}\right)\left(\frac{\partial u}{\partial z}\right)\left(\frac{\partial z}{\partial n}\right)\right] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & . & . \\ x & 1 & 0 & 0 & 0 & . & . \\ x & x & 1 & 0 & 0 & . & . \\ x & x & x & 1 & 0 & . & . \\ . & . & . & . & . & . & . \end{bmatrix} \quad (3.22)$$

since $\left(\frac{\partial H}{\partial u}\right)\left(\frac{\partial u}{\partial z}\right)\left(\frac{\partial z}{\partial n}\right)$ can be shown to be a lower triangular matrix with zeros on the main diagonal. Then $\left|\frac{\partial z}{\partial n}\right| = 1$ and (3.21) becomes $\left|\frac{\partial g}{\partial y_o}\right|^{-1}$.

This is the inverse of the determinant of the well known deterministic state transition matrix (including the dw variables), evaluated for a given $y_o(y_T, z)$ and $u(z)$.

Suppose that at some time t_K we are given the data vector $z(K)$ and the previously applied control $u(K-1)$. The probability density conditioned upon $z(K)$ is

$$p_K(y_T, z(N-K)\,|\,z(K)) = \left[\frac{p_T(y_T, z)}{p_{zK}(z(K))}\right] \tag{3.23}$$

where

$$p_{zK}(z(K)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_T(y_T, z)\, dy_T\, dz(N-K) \tag{3.24}$$

and $z(N-K)$ is the data to be obtained after t_K. Let the superscript $\hat{}$ denote the maximum likelihood estimate. Define the likelihood function

$$L(K) = \ln\,[p_K(y_T, z(N-K)\,|\,z(K))] = \ln p_T(y_T, z) - \ln p_{zK}(z(K)) \tag{3.25}$$

Then $\hat{y}_T(K)$, the anticipated data $\hat{z}(N-K)$, and the present plus anticipated control $u(N-K)$ are obtained as the values which simultaneously satisfy

$$0 = \left[\frac{\partial L(K)}{\partial y_T(K)}\right] = \left[\frac{\partial(\ln p_T)}{\partial y_T(K)}\right] \tag{3.26}$$

$$0 = \left[\frac{\partial L(K)}{\partial z(N-K)}\right] = \left[\frac{\partial(\ln p_T)}{\partial z(N-K)}\right] \tag{3.27}$$

where $u(N-K) = [u_K, u_{K+1}, \ldots u_N] = u(z(N-K))$ (The $P_{zK}(z(K))$ does not enter into (3.26) or (3.27) because $z(K)$ is given.) The instantaneous $u(K)$ is then the first element of $u(N-K)$. This algorithm is supposed to be applied at all times t_K where data is introduced.

Note that there is another maximum likelihood estimation procedure which can be devised. Suppose that the subset of components of $\hat{y}_T$ which appear in the performance index P are denoted by $\hat{y}_{T_1}$, where $y_T^T = [y_{T_1}^T, y_{T_2}^T]$.

Define the marginal distribution

$$p_K^*(y_{T_1},\ z(N-K)\,|\,z(K)) = \int_{-\infty}^{\infty} p_K(y_T,\ z(N-K)\,|\,z(K))\,dy_{T2} \tag{3.28}$$

Then equations similar to (3.26) and (3.27) are developed but with p_K^* replacing p_K and $\hat{y}_{T_1}$ replacing $\hat{y}_T$. In this case the uninteresting components y_{T_2} are integrated out prior to finding the estimates and the control, and we are controlling the "marginal mode" of the trajectory. If the entire y_T vector is estimated and the components $\hat{y}_{T_1}$ are selected, which occurs when p_K is used, we are controlling the "modal trajectory". It can be shown that these two techniques yield identical results when the system is linear, otherwise they do not. Control of the marginal mode has the most esthetic appeal, but for computational convenience we shall work with the modal trajectory. Essentially, we wish to avoid calculating a probability density function in real time.

3.4 ON THE SEPARABILITY OF ESTIMATION AND CONTROL

The combined estimation and control algorithm developed in Section 3.3 might not be practical to apply. Analogous to the open loop control problem discussed in Section 3.2, where the solution can be divorced from statistical considerations, we seek to introduce assumptions which allow the estimation and control problems to be separated. That is, we wish to justify a procedure whereby the initial condition y_o can be estimated at t_K, given data up to t_K, and this estimate can be treated as though it were the true state of a deterministic system for the purpose of computing the control.

Suppose that dt in (3.10) is sufficiently small so that the state transition matrix $[\frac{\partial g}{\partial y_o}]$ can be represented by its continuous analogue. Then, as in Section 3.2,

$$\left|\frac{\partial g}{\partial y_o}\right|^{-1} = \exp\left\{-\int_0^T \text{trace}\,[\frac{\partial f}{\partial x}\,(x(t;\,y_o;\,u),t]dt\right\} \tag{3.29}$$

For many applications it is true that trace $[\frac{\partial f}{\partial x}] \equiv 0$, and hence $|\frac{\partial g}{\partial y_o}| \equiv 1$.

For example, this condition applies when the continuous form of (3.10) is

a Hamiltonian system. Then let us assume that

$$\text{(A)} \qquad \left|\frac{\partial(y_o,n)}{\partial(y_T,z)}\right| = \left|\frac{\partial g}{\partial y_o}\right|^{-1} = 1 \tag{3.30}$$

(B) the apriori probability density function $p_o(y_o, n)$ is of the form

$$p_o(y_o, n) = p_y(y_o)\ p_n(n) \tag{3.31}$$

where y_o and n are both Gaussian variables with zero mean.

(C) the n_i are uncorrelated, i.e.

$$E[n\ n^T] = \Gamma = \text{diagonal}$$

Then, using (3.15),

$$-\ln p_T(y_T,z) = \frac{1}{2}\left\{y_o^T\ \Lambda^{-1}\ y_o + (z-H)^T\ \Gamma^{-1}\ (z-H)\right\} \tag{3.32}$$

where Λ and Γ are, respectively, the apriori variances of y_o, and n. The y_o and $H(y_o, u)$ are considered to be implicit functions of y_T and z. Because of the form of (3.32), the likelihood equations (3.26) and (3.27) become

$$0 = \left\{y_o^T\ \Lambda^{-1} - [z-H]^T\ [\Gamma^{-1}]\left[\left(\frac{\partial H}{\partial y_o}\right) + \left(\frac{\partial H}{\partial u(N-K)}\right)\left(\frac{\partial u(N-K)}{\partial y_o}\right)\right]\right\}\left[\frac{\partial y_o}{\partial y_T}\right] \tag{3.33}$$

$$0 = \left\{y_o^T\ \Lambda^{-1} - [z-H]^T\ [\Gamma^{-1}]\left[\left(\frac{\partial H}{\partial y_\sigma}\right) + \left(\frac{\partial H}{\partial u(N-K)}\right)\left(\frac{\partial u(N-K)}{\partial y_o}\right)\right]\right\}\left[\frac{\partial y_o}{\partial z(N-K)}\right] + [z-H^T]^T\ [\Gamma^{-1}]\left[\begin{matrix}0\\ \cdots\\ I(N-K)\end{matrix}\right] \tag{3.34}$$

where I(N-K) is the (N-K) dimensional identity matrix. But

$$\left[\frac{\partial H_K}{\partial u(N-K)}\right] = 0 \tag{3.35}$$

Let $\hat{y}_o$ be a solution of

$$0 = y_o^T \Lambda^{-1} - [z(K) - H_K]^T \Gamma^{-1} \left[\frac{\partial H_K}{\partial y_o}\right] \tag{3.36}$$

Then, because of (3.35) and assumption (C), it follows that (3.33) and (3.34) are satisfied by this value of $\hat{y}_o$ along with

$$\hat{z}(N-K) = H_{N-K}\ (\hat{y}_o,\ u(K),\ \hat{u}(N-K)) \tag{3.37}$$

where $\hat{u}(N-K)$ the anticipated control, which depends upon the control law $u(z)$, and $u(K)$ is the previously determined (fixed) control.

Since we have as yet made no assumptions about the apriori specified function $u(z)$, it is now legitimate to define the anticipated control $\hat{u}(N-K)$ as a function of $\hat{y}_T(K)$, where $\hat{y}_T(K)$ is the solution of (3.10) with $\hat{y}_o(K)$ used to define the initial conditions x_o and the disturbances dw_i. The $u(N-K)$ might be determined, say, by minimizing some performance index $J(\hat{y}_T(K),\ u(N-K))$. Then

$$\frac{\partial J}{\partial u(N-K)} = 0 \tag{3.38}$$

defines the control. For example, (3.38) would be the Euler-Lagrange equation if the control were continuous and the system were deterministic. This $\hat{u}(N-K)$ has the property that at each time t_K some function of the anticipated maximum likelihood estimate of the final state is minimized. Since u_K at every time t_K is in this way a function of $\hat{y}_T(K)$, which in turn is a function of $z(K)$, it follows that $u(z)$ is a well defined function. Then the combined estimation and control procedure is as follows:

(a) find $\hat{y}_o$ from (3.36)

(b) using this value in (3.10) and (3.38), find the anticipated control $\hat{u}(N-K)$ and hence $u(t_K)$ as the first element of $\hat{u}(t_K)$

Equation (3.37) can be used to determine $\hat{z}(N-K)$, but this is not required to define $\hat{u}(N-K)$. Note that $\hat{z}(N-K)$ is the data which would be realized if $\hat{y}_o$ did describe the true values of x_o and the past and future disturbances dw_j, and if the future data noise were to be zero. Thus $\hat{z}(N-K)$ is the data calculated at times $t_K < t_i \leq t_N$ on the modal trajectory defined by $\hat{y}_o(K)$ and $\hat{u}(N-K)$.

The estimation and control equations developed here might be simpler to implement in continuous, rather than discrete, form. As pointed out above, the continuous form of (3.38) is the Euler-Lagrange equation. The continuous form of the estimation equation (3.36) including the effects of continuous state noise, will be developed in Section 5.

3.5 THE MOMENTS OF THE CONDITIONAL ESTIMATION ERROR

For error analysis purposes it may be necessary to compute at time t_K certain statistics of the estimation error $\varepsilon_K = y_o - \hat{y}_o(K)$ associated with the estimate $\hat{y}_o$. To avoid the necessity of deriving a complicated error probability density function, an approximate expression for the moments of ε_K can be developed, which becomes exact under certain asymptotic conditions. For convenience we will henceforth drop the subscripts K and O, and ε will be understood to mean $\varepsilon = [y_o - \hat{y}_o(K) \,|\, z(K)]$.

Suppose that the estimation equation (3.36) is written

$$0 = \Lambda^{-1} \hat{v} - A^T(\hat{y})\, \Gamma^{-1} [H(y) + n - H(\hat{y})] \tag{3.39}$$

where $H(y)$ is understood to mean $H_K(y, u(K-1))$, with $u(K-1)$ fixed at its predetermined values, and

$$A(\hat{y}) = \left[\frac{\partial H}{\partial \hat{y}} (\hat{y}) \right] \tag{3.40}$$

Then (3.40) implicitly defines the random variable $\hat{y}(y,n)$, which can be expanded about an arbitrary point $y = a$, $n = b$ in the form

$$\hat{y}\ (y,n) = \hat{y}\ (a,b) + \left[\frac{\partial \hat{y}}{\partial y}\ (a,b)\right]\ (y-a) + \left[\frac{\partial \hat{y}}{\partial n}\ (a,b)\right]\ (n-b) + \text{higher order terms} \tag{3.41}$$

Applying the implicit function theorem, (3.41) yields

$$\left[\frac{\partial \hat{y}}{\partial y}\right] = \left[\Lambda^{-1} + A^T\ (\hat{y})\ \Gamma^{-1}\ A(\hat{y}) - B(y,n,\hat{y})\right]^{-1} \left[A^T\ (\hat{y})\ \Gamma^{-1}\ A(y)\right] \tag{3.42}$$

$$\left[\frac{\partial \hat{y}}{\partial n}\right] = \left[\Lambda^{-1} + A^T(\hat{y})\ \Gamma^{-1}\ A(\hat{y}) - B(y,n,\hat{y})\right]^{-1} \left[A^T(\hat{y})\ \Gamma^{-1}\right] \tag{3.43}$$

where, assuming Γ is diagonal, B is the symmetric matrix with elements

$$B_{ij} = \sum_k r_k\ (y,n,\hat{y})\ \Gamma_{kk}^{-1} \left[\frac{\partial^2 H_k\ (\hat{y})}{\partial \hat{y}_i\ \partial \hat{y}_j}\right] \tag{3.44}$$

and r is the residual vector

$$r(y,n,\hat{y}) = H(y) + n - H(\hat{y}) \tag{3.45}$$

The higher order terms can be evaluated by repeated differentiation of (3.42) and (3.43).

Suppose we are given a data vector z which yields the estimate $\hat{y}$. Choose $a = \hat{y}$ and $b = z - h(a)$ in (3.41), which implies that $\hat{y}(a,b) = a$. This value of b is the best estimate of n, given z. Holding z fixed, so that $\hat{y} = a$ is fixed, the estimation error conditioned upon z is $\varepsilon = y - \hat{y} = y - a$. Then from (3.41) we have

$$0 = \left[\frac{\partial \hat{y}}{\partial y}(a,b)\right] \varepsilon + \left[\frac{\partial \hat{y}}{\partial n}(a,b)\right] (n-b) + \text{higher order terms.} \tag{3.46}$$

The series (3.46) implicitly defines ε (n-b). Since $\varepsilon(o) = 0$, we can apply (3.46) to obtain

$$\begin{aligned} \varepsilon &= \left[\frac{\partial \varepsilon}{\partial n}\right] \left[n-b\right] + \textit{higher order terms in } (n-b) \\ &= \left[A^T \Gamma^{-1} A\right]^{-1} \left[A^T \Gamma^{-1}\right] (n-b) + \ldots \end{aligned} \tag{3.47}$$

The moment generating function for ε is

$$\begin{aligned} M_\varepsilon(\theta) &= \int_{-\infty}^{\infty} \exp\left[\theta^T \varepsilon(n-b)\right] \; p_n(n)\, dn \\ &= c \int_{-\infty}^{\infty} \exp\left[\theta^T \varepsilon(n-b) - n^T \frac{\Gamma^{-1}}{2} n\right] dn \end{aligned} \tag{3.48}$$

where θ is a parameter vector and c is the coefficient of the Gaussian density function p_n. The p^{th} moment of the i^{th} component of ε is then given by

$$E\left[\varepsilon_i^p\right] = \left(\frac{d^p M_\varepsilon(\theta)}{d\theta_i^{\,p}}\right)_{\theta = 0} \tag{3.49}$$

Alternatively, one could evaluate the moments of ε directly from (3.47), since n is a zero mean, Gaussian variable with known covariance. In either case it would be necessary in practice to delete some of the higher order terms in (3.47). This can be justified if terms of the form

$$E\left[\left(\frac{\partial^p \varepsilon}{\partial n^p}\right)\frac{(n-b)^p}{p!}\right]$$

are negligibly small for large p, which is implied if the variance of ε is small.

Note that the first moment $E[\varepsilon]$ is the difference between the maximum likelihood and minimum variance estimates, which is defined by the property that $E[\varepsilon] = 0$. Thus the minimum variance estimate is $\{a + E[\varepsilon]\}$. If only the linear term in (3.47) is significant, as would be the case if the asymptotic conditions to be discussed in Section 3.6 apply, then both estimates are (approximately) the same, and both produce a Gaussian estimation error with variance $(A^T \Gamma^{-1} A)$. This expression will be developed in Section 3.6.

3.6 THE ASYMPTOTIC MOMENTS OF THE CONDITIONAL ESTIMATION ERROR

A simplified (truncated) form of (3.47) can be devised under certain "asymptotic" conditions, which can apply if state noise is not present. Assume that the system (3.10) is observable and asymptotically well conditioned, where

Definition: A dynamic system is said to be completely observable at time t_K if the normal matrix (information matrix) has full rank for all $\hat{y}$, where

$$\text{normal matrix} = [A_K^T(\hat{y})\ \Gamma^{-1} A_K(\hat{y})] \equiv N_K(\hat{y}) \tag{3.50}$$

A system is said to be asymptotically well conditioned if the eigenvalues of $N_K(\hat{y})$ go to infinity as t_K goes to infinity.

Definition: The asymptotic form of (3.47) is the series obtained by deleting terms of order $|N^{-1}|^2$.

The linear terms of the asymptotic form are as given by (3.42) and (3.43). The higher order terms can be evaluated by noting that

$$\left[\frac{\partial \hat{y}}{\partial y}(a,b)\right] = \left[I + N^{-1}\Gamma^{-1}\ N^{-1} B\right]^{-1} \approx I \tag{3.51}$$

Then if

$$\left\{\frac{\partial N^{-1}}{\partial(y,n)}\right\}_{ij} = -\left\{N^{-1}\left[\frac{\partial N}{\partial(y,n)}\right] N^{-1}\right\}_{ij} = \text{terms of order } |N^{-1}|^2$$

it follows that (symbolically)

$$\left[\frac{\partial^2 \hat{y}}{\partial y^2}(a,b)\right] \approx -N^{-1}\left(\frac{\partial B}{\partial \hat{y}}\right) \tag{3.52}$$

where $\left(\frac{\partial B}{\partial \hat{y}}\right)$ involves terms of the type $\left(\frac{\partial^3 H}{\partial y^3}\right)$, since

$$\left[\frac{\partial r}{\partial y}(a,b)\right] = \left[\left(\frac{\partial H}{\partial \check{y}}\right) - \left(\frac{\partial H}{\partial \hat{y}}\right)\left(\frac{\partial \check{y}}{\partial y}\right)\right] \approx 0 \tag{3.53}$$

Similarly,

$$\left[\frac{\partial \hat{y}}{\partial n}(a,b)\right] = \left[\Lambda^{-1} + N - B\right]^{-1} A \Gamma^{-1} \approx N^{-1} A \Gamma^{-1} \tag{3.54}$$

so that (symbolically)

$$\left[\frac{\partial^2 \hat{y}}{\partial n^2}(a,b)\right] \sim \text{terms of order } (N^{-1})^2 \approx 0 \tag{3.55}$$

$$\left[\frac{\partial^2 \hat{y}}{\partial n \partial y}(a,b)\right] = N^{-1}\left[\frac{\partial A^T}{\partial y}\right]\Gamma^{-1} \tag{3.56}$$

Thus to second order (3.46) is approximately linear in (n-b) and quadratic in ε. If the cubic terms in (3.52) are negligible, and if $\left|\frac{\partial A^T}{\partial y}\right|$ is of order N^{-1}, then only the linear terms of (3.47) are significant and

$$\varepsilon = N^{-1} A^T \Gamma^{-1} (n-b) \tag{3.57}$$

The estimation error is then (approximately) a Gaussian variable, with

$$E[\varepsilon] = - N^{-1} A^T \Gamma^{-1} b = - N^{-1} \Lambda^{-1} a \tag{3.58}$$

since $A^T \Gamma^{-1} b = \Lambda^{-1} a$. This is the correction term which makes the estimate for a linear system unbiased over the ensemble of all data realizations (Recall that ε is the error obtained holding the data fixed). The variance is

$$E[\varepsilon - N^{-1} \Lambda^{-1} a]^2 = N^{-1} A^T \Gamma^{-1} A^T N^{-1} = N^{-1} \tag{3.59}$$

Assuming the linear approximation is reasonable, the asymptotic mean of ε could be obtained without inverting the series (3.46) by taking the expectation of (3.46) and substituting $E[\varepsilon^p]$ as obtained from the Gaussian approximation for all $p > 1$.

3.7 THE ASYMPTOTIC STATISTICS OF THE ESTIMATION ERROR FOR PARTIALLY OBSERVABLE SYSTEMS

An asymptotic form of the error statistics can be developed for some components of the state of a partially observable system, where the normal matrix does not have full rank. This situation is to be expected when state noise is present.

Suppose the state vector is decomposed into two parts according to $y^T = [y_1^T, y_2^T]$, and let (3.42) and (3.43) be written

$$\begin{bmatrix} \frac{\partial \hat{y}_1}{\partial y_1} & \frac{\partial \hat{y}_1}{\partial y_2} \\ \frac{\partial \hat{y}_2}{\partial y_1} & \frac{\partial \hat{y}_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}^{-1} \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \tag{3.60}$$

$$\begin{bmatrix} \frac{\partial \hat{y}_1}{\partial n} \\ \frac{\partial \hat{y}_2}{\partial n} \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_1^T \\ A_2^T \end{bmatrix} \Gamma^{-1} \tag{3.61}$$

$$A_1 = \frac{\partial H}{\partial y_1} \tag{3.62}$$

$$A_2 = \frac{\partial H}{\partial y_2} \tag{3.63}$$

$$N = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} = \begin{bmatrix} (A_1^T \Gamma^{-1} A_1) & (A_1^T \Gamma^{-1} A_2) \\ (A_2^T \Gamma^{-1} A_1) & (A_2^T \Gamma^{-1} A_2) \end{bmatrix} \tag{3.64}$$

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = \begin{bmatrix} (\Lambda_{11}^{-1} + N_{11} - B_{11}) & (N_{12} - B_{12}) \\ (N_{21} - B_{21}) & (\Lambda_{22}^{-1} + N_{22} - B_{22}) \end{bmatrix} \tag{3.65}$$

In (3.65) we have assumed that y_1 and y_2 are apriori uncoupled, so that $\Lambda_{12} = 0$ and $\Lambda_{21} = 0$. From (3.60) and 3.61) we have

$$\left(\frac{\partial \hat{y}_1}{\partial y_1}\right) = M_{11}^{-1} N_{11} + K (M_{21}^{-1} M_{11} N_{11} - N_{21}) \tag{3.66}$$

$$\left(\frac{\partial \hat{y}_1}{\partial y_2}\right) = M_{11}^{-1} N_{12} + K (M_{21} M_{11}^{-1} N_{12} - N_{22}) \tag{3.67}$$

$$\left(\frac{\partial \hat{y}_1}{\partial n}\right) = M_{11}^{-1} A_1^T \Gamma^{-1} + K (M_{21} M_{11}^{-1} A_1^T - A_2^T) \Gamma^{-1} \tag{3.68}$$

where

$$K = M_{11}^{-1} M_{12} (M_{22} - M_{21}^T M_{11}^{-1} M_{12})^{-1} \tag{3.69}$$

Suppose that there is a weak coupling between y_1 and y_2, in the sense that $N_{12} \approx 0$, $B_{12} \approx 0$, and hence $M_{12} \approx 0$. Then

$$\left(\frac{\partial \hat{y}_1}{\partial y_1}\right) \approx M_{11}^{-1} N_{11} \tag{3.70}$$

$$\left(\frac{\partial y_1}{\partial y_2}\right) \approx 0 \tag{3.71}$$

$$\left(\frac{\partial \hat{y}_1}{\partial n}\right) \approx M_{11}^{-1} A_1^T \Gamma^{-1} \tag{3.72}$$

In particular, this is the case if y_2 is the unobservable portion of the state vector, for then $A_2 = 0$. Thus if N_{11} asymptotically goes to infinity while N_{22} remains small, we assert that the analysis of Section 3.4 can be applied essentially unchanged to the observable components of the state. This conclusion is important when there are many system parameters to be considered, e.g., when state noise is present.

3.8 APPROXIMATE DIFFERENTIAL EQUATIONS FOR THE ESTIMATION ERROR BIAS

In Section 3.6 it was shown that the conditional estimation error becomes asymptotically Gaussian, and it was suggested that the error bias (mean) in the non-asymptotic case (finite normal matrix) might be approximated by using the Gaussian assumption to calculate the higher moments in the (3.46). Consistent with this approach, it is possible to derive an approximate differential equation for the estimation error bias. The method can only be rigorously justified for observable systems with no state noise, but the effect of state noise will be included for completeness.

Let $\hat{x}_o$ and $d\hat{w}(t)$ be the estimates of the initial conditions, determined from data $z(K)$, where $t \leq t_K$. Define

$$\varepsilon(t) = x(t) - \hat{x}(t)$$

$$d[\delta w(t)] = d[w(t) - \hat{w}(t)]$$

Expanding the right hand side of (3.1) about $[\hat{x}(t), d\hat{w}(t)]$, and (for simplicity) deleting the control, we have

$$\frac{d\varepsilon_i}{dt} = \sum_j \left[\frac{\partial f_i}{\partial x_j}(\hat{x},t)\right] \varepsilon_j(t) + \frac{1}{2}\sum_{j,k}\left[\frac{\partial^2 f_i}{\partial x_j \partial x_k}(\hat{x},t)\right]\varepsilon_j(t)\ \varepsilon_k(t)$$

$$+ \text{higher order terms} + \frac{d}{dt}[\delta w_i(t)] \qquad i,j,h=1,..n \tag{3.73}$$

Let $E[\varepsilon_i(t)\varepsilon_j(t)] = p_{ij}(t)$. Then introducing the Gaussian assumption for the purpose of calculating moments on the right hand side of the differential equation (e.g., first moment is zero, etc.), we have

$$E\left[\frac{d\varepsilon_i}{dt}\right] = \frac{1}{2}\sum_{j,k}\left(\frac{\partial^2 f_i}{\partial x_j \partial x_k}\right)\left(p_{jk}\right)$$

$$+ \frac{1}{4!}\sum_{j,h,lm}\left(\frac{\partial^4 f_i}{\partial x_j \partial x_k \partial x_l \partial x_m}\right)\left(p_{jk}\ p_{lm}\right)$$

$$+ \text{terms of order } (p_{jk})^4 \qquad i,j\cdot\cdot=1,\cdot\cdot n \tag{3.74}$$

The $p_{ij}(t)$ can be calculated from

$$\frac{dp_{ij}}{dt} = E\frac{d}{dt}[\varepsilon_i(t)\varepsilon_j(t)] = E\left[\frac{d\varepsilon_i(t)}{dt}\varepsilon_j(t)\right] + E\left[\varepsilon_i(t)\frac{d\varepsilon_j(t)}{dt}\right]$$

$$= \sum_k \left(\frac{\partial f_i}{\partial x_k}\right)(p_{kj}) + \left(\frac{\partial f_j}{\partial x_k}\right)(p_{ki})$$

$$+ \frac{1}{3!}\sum_{k,l,m}\left(\frac{\partial^3 f_i}{\partial x_k \partial x_l \partial x_m}\right)(p_{kl}\ p_{mj})$$

$$+ \frac{1}{3!}\sum_{k,l,m}\left(\frac{\partial^3 f_j}{\partial x_k \partial x_l \partial x_m}\right)(p_{kl}\ p_{mi})$$

$$+ \text{terms of order } (p_{ij})^4 + \gamma_{ij} \qquad i,j,\cdot\cdot=1,\cdots,n \tag{3.75}$$

where

$$\gamma_{ij}(t)\ dt = E[\{d\delta w_i(t)\}\ \{d\delta w_j(t)\}]$$

Consistent with the Gaussian assumption, the initial conditions for (3.74) and(3.75)are, respectively, $E[\varepsilon_i(0)] = 0$ and $E[p_{ij}(0)] = N_{xo}^{-1}$ where N_{xo} is that portion of the normal matrix of(3.50)which describes the initial conditions. Similarly, the $\gamma_{ij}(t)$ would be calculated as though the system were linear. This should be done by finding the error covariance associated with the continuous estimator of $(\frac{d\hat{w}}{dt})$, to be discussed in Section 6. Alternatively, one could find an approximate $\gamma_{ij}(t)$ from the well known Kalman filter equations.

Note that this method of finding the estimation error statistics differs from other approximation techniques employing Gaussian moments in that the modal trajectory $\hat{x}(t)$ is used as the reference, rather than the apriori trajectory. Recalling the discussion of the asymptotic properties of the estimation error of Section 3.6, it appears that moment truncation can only be justified if one expands about the modal trajectory. This will be the basis for the linearized analysis to be presented in Sections 7 and 8.

4. ESTIMATION OF UNKNOWN ACCELERATION

4.1 INTRODUCTION

The maximum likelihood estimation algorithm developed in Section 3 requires that the initial conditions and the random acceleration function be simultaneously estimated. The resulting solution includes the "smoothed" (rather than "filtered") estimate of the acceleration function at every instant in time, based upon the entire data record. This property, plus the fact that linearity is not assumed, distinguishes this approach from the treatment of state noise in Kalman filtering. The Kalman form of the estimator can be extended to treat the linear data smoothing problem, however, as is shown in [32].

This type of problem arises in many applications, such as the reconstruction of lift/drag histories for re-entry vehicles, or the analysis of trajectories perturbed by random thrust acceleration, or the determination of orbits which are perturbed by model errors of unknown origin. Probably the most important application is the latter. In this case one assumes that an unknown acceleration is acting, pretends that it is white noise with some hypothetical variance, and recovers an estimate of the unknown function which is hopefully a reasonable approximation of physical reality.

In this Section we shall develop a continuous expression for a more general form of the estimation algorithm presented in Section (3.5), by using a variational approach (see also [33]). A practical numerical algorithm for solving the problem will be developed, as well as a successive approximation technique which may be adequate for some applications. A simple example will be discussed, and the associated estimation using the algorithm to estimate non-white acceleration will be analyzed.

4.2 FORMULATION OF THE PROBLEM

Let the equations of motion be

$$dx = f(x, t)dt + G(x, t)\, dw \tag{4.1}$$

where x is the n-dimensional the state vector, composed of the position and velocity coordinates plus all parameters which affect the tracking problem

($f_i = 0$ if x_i is a constant parameter); the m-dimensional vector $\alpha(t) = [\frac{dw}{dt}(t)]$ is the random (unknown) acceleration function, and G(x, t) is a matrix which depends upon the state and time. For example, if x is a six dimensional vector, where (x_1, x_2, x_3) are velocity components and (x_4, x_5, x_6) are position components, then a possible form of G is $[\frac{I}{0}]$, where I is the 3 x 3 identity matrix. In this case n = 6 and m = 3. Let the differential of the $i^{\underline{th}}$ data type (e.g., the incremental change of the doppler integral) be given by

$$d\rho_i = h_i(x(t), t)dt + dn_i(t) \qquad i = 1, \ldots k \tag{4.2}$$

where k is the number of data types, $h_i(x(t), t)$ is some nonlinear function of x and t, and $dn_i(t)$ is data noise. Suppose that $w_i(t)$ and $n_i(t)$ are uncorrelated, zero mean, Gaussian stochastic processes, where, at each time t,

$$E\left[dw_i\right]^2 = \left[\sigma_{w_i}^2\right]dt \qquad i = 1, \ldots m$$

$$E\left[dn_i\right]^2 = \left[\sigma_{n_i}^2\right]dt \qquad i = 1, \ldots k$$

Thus $w_i(t)$ and $n_i(t)$ are Wiener processes. Furthermore, suppose that the apriori distribution of the initial state x_o is Gaussian, with mean μ and covariance matrix Λ. Considering the vectors dw and dn composed, respectively, of the $dw(t_i)$ and $dn(t_i)$ at the discrete times $\{t_o, t_1, \ldots t_i, \ldots t_N\}$, the joint apriori probability density function (p.d.f.) for $\{x_o, dw, dn\}$ is of the form

$$[\text{p.d.f. } (x_o, dw, dn)] = [\text{p.d.f. } (x_o)]\cdot[\text{p.d.f.}(dw(t_o))]\cdots$$

$$\cdots[\text{p.d.f.}(dw(t_N)]\cdot[\text{p.d.f.}(dn(t_o))]\cdots[\text{p.d.f.}(dn(t_N))] \tag{4.3}$$

where each [p.d.f.] is Gaussian. Let $z_i = (\frac{d\rho i}{dt})$. Taking the negative of the logarithm of (4.3), substituting $[z_i - h_i(x, t)]\,dt$ for $dn_i(t)$, and passing to the limit as $dt \to 0$, the likelihood function for $\{x_o, \dot{n}, \alpha\}$ is

$$p = \frac{1}{2}(x_o - \mu)^T \Lambda^{-1}(x_o - \mu)$$

$$+ \frac{1}{2}\sum_{i=1}^{k}\left[\sigma_{n_i}^2(t)\right]^{-1}\int_o^T \left[z_i - h_i(x, t)\right]^2 dt$$

$$+ \frac{1}{2}\sum_{i=1}^{m}\left[\sigma_{w_i}^2(t)\right]^{-1}\int_o^T \alpha_i^2(t)dt \qquad (4.4)$$

where $t_o = 0$ and T is the duration of tracking. Given data $z(t)$, the problem is to choose estimates of the initial condition, $\hat{x}_o$, and acceleration function, $\hat{\alpha}(t)$, which maximizes the likelihood function p.

4.3 THE FORM OF THE ESTIMATION EQUATIONS

The estimation equations can be obtained as the solution of a calculus of variations problem of the Mayer type. Define the additional state vector component p(t), with

$$p(o) = \frac{1}{2}(x_o - \mu)^T \Lambda^{-1}(x_o - \mu) \qquad (4.5)$$

and

$$\frac{dp(t)}{dt} = \frac{1}{2}\sum_{i=1}^{k}\left[\sigma_{n_i}^2(t)\right]^{-1}\left[z_i - h_i(x, t)\right]^2$$

$$+ \frac{1}{2}\sum_{i=1}^{m}\left[\sigma_{w_i}^2(t)\right]^{-1}\left[\alpha_i^2(t)\right] \qquad (4.6)$$

Thus p(T) is the likelihood function, as per (4.4). Adjoin the $\dot{p}$ to (4.1) and construct the variational equation describing variations from the best estimate of the trajectory (denoted by $\hat{}$), given by

$$\frac{d}{dt}\begin{bmatrix}\delta x \\ \delta p\end{bmatrix} = \begin{bmatrix}F & 0 \\ H & 0\end{bmatrix}\begin{bmatrix}\delta x \\ \delta p\end{bmatrix} + \begin{bmatrix}G \\ g^T\end{bmatrix}\delta\alpha \qquad (4.7)$$

where $\delta x(t) = \hat{x}(t) - x(t)$, $\delta\alpha(t) = \alpha(t) - \hat{\alpha}(t)$, and

$$g^T(t) = \left\{ \left[\sigma_{w_1}{}^2(t)\right]^{-1} \hat{\alpha}_1(t), \ \ldots, \ \left[\sigma_{w_m}{}^2(t)\right]^{-1} \hat{\alpha}_m(t) \right\} \tag{4.8}$$

$$F(t) = \left[\frac{\partial f(\hat{x}(t))}{\partial x}\right] + [\hat{\alpha}(t)]^T \left[\frac{\partial G^T(\hat{x}(t), t)}{\partial x}\right] \tag{4.9}$$

$$H(t) = - \sum_{i=1}^{k} \left[\sigma_{n_i}{}^2(t)\right]^{-1} \left[\frac{\partial h_i(\hat{x}(t), t)}{\partial x(t)}\right] [z_i - h_i(\hat{x}(t), t)] \tag{4.10}$$

$$G(t) = G(\hat{x}(t), t) \tag{4.11}$$

Introduce the state transition matrix associated with the variational equation (4.7), which is of the form

$$\left[\begin{array}{c|c} \left(\frac{\partial x(t)}{\partial x(s)}\right) & \left(\frac{\partial x(t)}{\partial p(s)}\right) \\ \hline \left(\frac{\partial p(t)}{\partial x(s)}\right) & \left(\frac{\partial p(t)}{\partial p(s)}\right) \end{array}\right] = \left[\begin{array}{c|c} U(t, s) & 0 \\ \hline \lambda^T(t, s) & 1 \end{array}\right] \qquad t \geq s$$

where $U(t, s)$ is the familiar matrix $\left[\frac{\partial x(t)}{\partial x(s)}\right]$, and the row vector $\lambda^T(t, s)$ has the property that, for fixed t,

$$\frac{d}{ds} \lambda^T(t, s) = - [\lambda^T(t, s) F(s) + H(s)] \tag{4.12}$$

Note that $\lambda(T, s)$ is the Lagrange multiplier vector of [33]. The solution of (4.7) can then be written in the form

$$\delta x(t) = U(t, o) \delta x_o + \int_o^t U(t, s) G(s) \alpha(s) ds \tag{4.13}$$

$$\delta p(t) = \left[\lambda(t, o) + \frac{\partial p(o)}{\partial x_o}\right]^T \delta x_o + \int_o^t \left[\lambda^T(T, s) G(s) + \hat{\alpha}(s)^T R^{-1}(s)\right] \delta\alpha(s) ds \tag{4.14}$$

where

$$R(s) = \begin{bmatrix} \sigma_{w_1}^2(s) & & 0 \\ & \ddots & \\ 0 & & \sigma_{w_m}^2(s) \end{bmatrix} \tag{4.15}$$

If $p(T)$ is to be maximized with respect to arbitrary variations in $\hat{x}_o$ and $\hat{\alpha}(t)$, it follows that $\delta p(T)$ must be zero for all variations δx_o and $\delta\alpha(t)$, for if this were not true we could improve the value of $p(T)$ by a first-order change in these quantities. Thus, (taking the transpose)

$$\lambda(T, o) + \Lambda^{-1} (\hat{x}_o - \mu) = 0 \tag{4.16}$$

$$R(s)G^T(s)\lambda(T, s) + \hat{\alpha}(s) = 0 \tag{4.17}$$

where $\left[\frac{\partial p(o)}{\partial x_o}\right]$ was obtained from (4.5). The solution for $\lambda(T, s)$ can be explicitly obtained in terms of $U(s, o)$, for it can be shown that

$$\frac{d}{ds} U(s, o) = F(s)U(s,o) \tag{4.18}$$

Then

$$\frac{d}{ds} [U^{-1}(s, o)] = - [U^{-1}(s, o)] F(s) \tag{4.19}$$

and it is easily verified that

$$\lambda^T(T, s) = \left[\int_s^T H(\tau) U(\tau, o) d\tau \right] U^{-1}(s, o) \tag{4.20}$$

satisfies (4.12). Define the familiar row matrix

$$A_i(t) = \left[\frac{\partial h_i(\hat{x}(t),t)}{\partial x} \right]^T \left[U(t, o) \right] = \left[\frac{\partial h_i(t)}{\partial x_o} \right]^T \tag{4.21}$$

Then, from (4.10), (4.16), (4.17) and (4.20), we have

$$0 = \Lambda^{-1}(\hat{x}_o - \mu) - \int_o^T \sum_{i=1}^{k} \left[\sigma_{n_i}{}^2(t)\right]^{-1} \left[A_i{}^T(t)\right] \left[z_i(t) - h_i(\hat{x}(t), t)\right] dt \tag{4.22}$$

$$0 = \hat{\alpha}(s) - R(s)\ [U^{-1}(s, o)\ G(s)]^T \int_s^T \sum_{i=1}^{k} \left[\sigma_{n_i}{}^2(t)\right]^{-1} \left[A_i{}^T(t)\right] \left[z_i(t) - h_i\left(\hat{x}(t)t\right)\right] dt \tag{4.23}$$

Equations (4.22) and (4.23) are the results we seek. The solution of the estimation problem is the $\hat{x}_o$ and $\hat{\alpha}(s)$ which satisfies these equations.

4.4 A NUMERICAL ALGORITHM FOR SOLVING THE PROBLEM

Suppose that observations are made at discrete times $\{t_i\}$, where the measured observation vector is, according to 4.2, $d\rho(t_i) = \rho(t_i) - \rho(t_{i-1})$. The computed (predicted) observation at t_i, given the estimate of the state at t_{i-1}, is $[h(\hat{x}(t_{i-1}), t_{i-1})]dt$. Assume the data has been normalized so that $\sigma_{n_i}{}^2(t) = 1$ for all $i = 1, 2, \ldots k$, and define the data residual

$$r_i = r(t_i) = [h(\hat{x}(t_{i-1}), t_{i-1})]dt - d\rho(t_i) \tag{4.24}$$

If $\rho(t)$ is a k dimensional vector (usually $k = 1$) then $[\frac{\partial h}{\partial x(t)}]$ is a k by n matrix (usually $n = 6$) and the matrix

$$A_i = A(t_i) = [\frac{\partial h}{\partial x}\ (\hat{x}(t_{i-1}), t_{i-1})]^T [U(t_{i-1})] \tag{4.25}$$

is also a k by n matrix, where U(t) denotes the state transition matrix $U(t, o)$. Suppose that the apriori variance of x_o is infinite ($\Lambda^{-1} = 0$), so that the discrete form of (4.22) is

$$\sum_{j=1}^{k} A_j^T r_j = 0 = \sum_{j=1}^{i} A_j^T r_j + \sum_{j=i+1}^{k} A_j^T r_j \tag{4.26}$$

Setting $U^{-1} = V$, and employing (4.26), the discrete form of (4.23) is

$$\hat{\alpha}(t) = [R(t_i)\ G^T(t_i)\ V^T(t_i)] \sum_{j=1}^{i} A_j^T r_j \qquad t_i \le t \le t_{i+1} \tag{4.27}$$

If we let

$$\beta_i = \sum_{j=1}^{i} A_j^T r_j \tag{4.28}$$

then (for dt sufficiently small) the equations of motion are

$$\dot{x} = f(x, t) + G(x, t)\ \alpha(t) = f + (GRG^T V^T)\beta = f + \phi\beta \tag{4.29}$$

Each time a data point is passed, β is changed simply by $\beta + A^T r \to \beta$. If $\beta = 0$ after the last data point is processed, then the problem is finished. If not, then x_o must be changed. To see how much x_o must be changed, define $N(t) = \left[\frac{\partial \beta(t)}{\partial x_o}\right]$ so that a variation δx_o results in a variation $\delta\beta$ given by $\delta\beta = N\delta x_o$. We also define $M = M(t)$ so that $\delta x(t) = U(t)\ M(t)\ \delta x_o$. Now δx satisfies the differential equation

$$(\delta\dot{x}) = F\delta x + \phi\delta\beta \tag{4.30}$$

If M satisfies the differential equation where $F = F(t)$ is defined by (4.9)

$$\dot{M} = V\ \phi\ N\ , \tag{4.31}$$

then $\delta x = U\ M\ \delta x_o$ will satisfy the required differential equation as may be seen by substitution:

$$(\delta\dot{x}) = \frac{d}{dt}(U\ M\ \delta x_o)^{\cdot}$$

$$= \dot{U}\ M\ \delta x_o + U\ \dot{M}\ \delta x_o$$

$$= F\ U\ M\ \delta x_o + U\ V\ \phi\ N\ \delta x_o$$

$$= F\ \delta x + \phi\ N\ \delta x_o$$

$$= F\ \delta x + \phi\ \delta\ \beta \quad . \tag{4.32}$$

At each data time, the variation in r can be computed

$$\delta r = \left[\frac{\partial z_c}{\partial x_i}\right]\ \delta x_i$$

$$= \left[\frac{\partial z_c}{\partial x_i}\right]\ U(t_i)\ M(t_i)\ \delta x_o$$

$$= A\ M\ \delta x_o \tag{4.33}$$

Now the change in the variation of β at each data time can be computed:

$$\beta + A^T\ r \rightarrow \beta$$

$$\delta\beta + A^T\ \delta r \rightarrow \delta\beta$$

$$N\ \delta x_o + A^T A\ M\ \delta x_o \rightarrow N\ \delta x_o$$

$$N + A^T A\ M \rightarrow N \quad .$$

Finally, if $\beta \neq 0$ after the last data time then a variation of $\delta x_o = [N(t)]^{-1}\ \beta(T)$ should make it near zero.

In summary, we have established the following algorithm:

Initial conditions for integration

$x(t) = x_o$ = approximate initial state vector (6 x 1 vector)

$M(0) = U(0) = I$ (6 x 6 matrices)

$\beta(0) = 0$ (6 x 1 vector)

$N(0) = 0$ (6 x 6 matrix)

Equations to be integrated

$\dot{x} = f(x, t) + \phi\ \beta$, where $\phi = (G\ R\ G^T V^T)$, $V = U^{-1}$

$\dot{U} = FU$

$\dot{M} = V\ \phi\ N$

Data processing (performed each data time)

$r = [h(x)]dt - d\rho$

$\beta + A^T\ r \rightarrow \beta$

$N + A^T A\ M \rightarrow N$

Initial state correction (performed after all data has been processed)

$\delta x_o = -\ [N(t)]^{-1}\ \beta(T)$

If $\delta x_o = 0$, the problem is finished; otherwise $x_o + \delta x_o \rightarrow x_o$, and the integration is restarted. After convergence, the acceleration function is given by $\hat{\alpha}(t) = [R(t)\ G^T(t),\ V^T(t)]\ \beta(t)$.

Note that the inverse of the state transition U(t) can be simply obtained. The standard variational equations are

$$\dot{U} = FU,\ U(t_o) = I,\ U(t) = 6 \times 6 \text{ matrix}$$

where F is the matrix of partial derivatives of components of f with respect to components of q. The inverse of U is called the adjoint matrix : $V = U^{-1}$. This matrix can be computed by $V = J\ U^T\ J^T$, where

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

This is known as the "Jacobi inverse" of U. To prove that this is the inverse of U note first that F has the form

$$F = \begin{pmatrix} 0 & I \\ \Psi & 0 \end{pmatrix}$$

where $\Psi^T = \Psi$. It follows that $F^T J^T + J^T F = 0$. Letting $\phi = VU$, it follows that $\dot{\phi} = \dot{V}U + V\dot{U} = J\dot{U}^T J^T U + JU^T J^T \dot{U} = JU^T F^T J^T U + JU^T J^T FU = JU^T (F^T J^T + J^T F)U = 0$. Since $\phi(t_o) = I$ and $\dot{\phi} = 0$, it follows that $\phi \equiv I$ and $V \equiv U^{-1}$.

4.5 A SUCCESSIVE APPROXIMATION TECHNIQUE FOR SOLVING THE PROBLEM

Existing orbit determination computer programs solve the nonlinear estimation problem when no random acceleration is present by a modified Newton-Raphson method. That is, equation (4.22) is linearized about the $K^{\underline{th}}$ estimate of x_o, the $(K + 1)^{\underline{st}}$ estimate is obtained by solving a set of linear equations, and the procedure is iterated to convergence. This procedure will obviously give an approximately correct estimate of the parameter vector x_o if only small amplitude random acceleration is present. It, therefore, seems reasonable to suppose that the presently employed parameter estimation algorithm could be modified so as to reflect the presence of small applitude acceleration. Such a method will be developed here.

Assume the data has been normalized so that $\sigma_{n_i}^2(t) = 1$ for all i, and let

$$\hat{x}_o^K = \text{the } K^{\underline{th}} \text{ estimate of } x_o$$

$$\hat{\alpha}^K(t) = \text{the } K^{\underline{th}} \text{ estimate of } \alpha(t)$$

Holding $\hat{\alpha}^K(t)$ fixed, expand (4.22) to first order in $\hat{x}_o$ to obtain

$$[\hat{x}_o^{K+1} - \hat{x}_o^K] = [\Lambda^{-1} + N^K]^{-1} \left\{ \int_0^T [A^T(t)]^K [z(t) - h(x^K(t), t)]dt - [\Lambda_o^{-1} (\hat{x}_o^K - \mu)] \right\} \tag{4.34}$$

where $[A(t)]^K$, $[U(t, o)]^K$, and $[G(t)]^K$ are the previously defined matrices, evaluated on the $K^{\underline{th}}$ estimated trajectory, and the normal matrix is

$$N^K = \sum_{i=1}^{k} \int_0^T [A_i^T(t)\ A_i(t)]^K\ dt \tag{4.35}$$

The revised estimate of $\alpha(t)$ is

$$[\hat{\alpha}(s)]^{K+1} = [R(s)\ G^T(s)\ V^T(s)]^K \int_s^T [A^T(t)]^K [z(t) - h(x^{K+1}(t), t)]dt \tag{4.36}$$

This algorithm is to be iterated to convergence, where the partial derivative matrices could be changed at each iteration in order to introduce the effect of nonlinearity. Note that equation (4.34) is identical to the form presently implemented in orbit determination programs, except that the calculated residuals reflect the effect of $\hat{\alpha}(s)$ and successive iterations introduce changes in $\hat{\alpha}(t)$ as an intermediate step.

It is not at all clear that the suggested successive approximation technique will converge. The solution of the linearized problem can be thought of as a successive approximation technique for inverting a matrix. To see this, suppose that there is only one data type which is linear in x_o and $\alpha(t)$, that is

$$z(t) = A(t)\ x_o + \left[\frac{\partial h}{\partial x}\ (t)\right] \int_o^t U(t, s)\ G(s)\ \alpha(s)\ ds + \frac{dn}{dt} \tag{4.37}$$

Suppose that $\mu = 0$ and $\sigma_{w_i}^{\ 2}(t) = \sigma_v^{\ 2}$ for all t, and let equations (4.22) and (4.23) be written in the symbolic form (think of $\alpha(t)$ as the infinite-dimensional vector α)

$$[\Lambda^{-1} + N]\ \hat{x}_o + P\hat{\alpha} = f_1(z) \tag{4.38}$$

$$[I + \sigma_v^{\ 2} M]\ \hat{\alpha} + \sigma_v^{\ 2}\ Q\hat{x}_o = \sigma_v^{\ 2}\ f_2(z) \tag{4.39}$$

where P, M and Q are linear integral operators denoting, respectively, the operations

$$\int_o^T [A^T(t)][\frac{\partial h}{\partial x}(t)] \int_o^t U(t, s)\, G(s)\, \alpha(s)\, ds\, dt$$

$$[U^{-1}(s,o)\, G(s)]^T \int_s^T [A^T(t)][\frac{\partial h}{\partial x}(t)] \int_o^t U(t, \tau)\, G(\tau)\, \alpha(\tau)\, d\tau\, dt$$

$$[U^{-1}(s,o)\, G(s)]^T \int_s^T [A^T(t)\, A(t)]\, dt$$

and

$$f_1(z) = \int_o^T A^T z\, dt$$

$$f_2(z) = [U^{-1}(s,o)\, G(s)]^T \int_s^T A^T z\, dt$$

Using equation (4.39) to eliminate $\hat{\alpha}$ from equation (4.38), we have

$$(\Lambda_o^{-1} + N - \sigma_v^2\, P[I + \sigma_v^2\, M]^{-1}\, Q)\, \hat{x}_o = (f_1(z) - \sigma_v^2\, P[I + \sigma_v^2\, M]^{-1}\, f_2(z)) \tag{4.40}$$

If σ_v^2 is sufficiently small, an approximate solution for $\hat{x}_o$ is

$$\hat{x}_o \cong [\Lambda_o^{-1} + N]^{-1}\, f_1(z) \tag{4.41}$$

and the corresponding approximation for $\hat{\alpha}$ is

$$\hat{\alpha} = \sigma_v^2\, [f_2(z) - Q\, \hat{x}_o] \tag{4.42}$$

This procedure can be iterated, yielding

$$[\hat{x}_o]^{i+1} - [\hat{x}_o]^i = [\Lambda_o^{-1} + N]^{-1}\, \{f_1(z) - [\Lambda_o^{-1} + N][\hat{x}_o]^i - P[\hat{\alpha}]^i\} \tag{4.43}$$

$$[\hat{\alpha}]^{i+1} = \sigma_v^2 \{f_2(z) - M[\hat{\alpha}]^i - Q[\hat{x}]^{i+1}\} \quad (4.44)$$

It can be shown that this process converges to the correct answer if σ_v^2 is sufficiently small. Recalling the meaning of the operators P, M, and Q, equations (4.43)and (4.44) are identical to equations (4.22) and (4.23) for the linear case. Thus it appears that the successive approximation technique will converge for linear systems if σ_v^2 is sufficiently small.

4.6 ON THE ERROR STATISTICS ASSOCIATED WITH ESTIMATING UNKNOWN, NON-WHITE ACCELERATION

The estimation equations for determining $\hat{x}_o$ and $\hat{\alpha}(t)$ depend upon the parameters σ_n^2, which is the apriori variance of the white data noise, and σ_w^2, which is the apriori variance of the white random acceleration. In practice, one might apply such an algorithm to simplify the task of estimating an unmodeled acceleration which is not white; that is, the postulated model is to be thought of as an approximation to physical reality. Given σ_n^2, a sub-optimal estimator is then obtained which depends upon the parameter σ_w^2. Given the true error statistics of the unmodeled acceleration, the analyst must then determine the resulting estimation error statistics as a function of σ_w^2 and choose σ_w^2 so as to achieve an acceptable result. In this subsection such an analysis will be illustrated by treating a simplified trajectory model which describes the motion over a short tracking arc. Considering the inevitable uncertainties in the apriori statistics, it is probably true that the simplified analysis will be adequate for determining a rule for selecting the σ_α^2 parameter.

4.6.1 The Simplified Problem

Suppose the one component of the velocity variation from a perturbed trajectory over a short arc can be described by

$$\dot{v} = \alpha(t) \quad (4.45)$$

where $\dot{v}$ is the velocity variation and $\alpha(t)$ is the unknown acceleration.†

†This model assumes that the gravitational acceleration can be represented as a function of time, so that gravity variations are zero in equation (4.45).

We postulate that $\alpha(t)$ is white acceleration noise with variance σ_α^2, and that the apriori variance of the initial condition v_o is infinite. The data consists of

$$z(t) = v(t) + n(t) = v_o + \int_o^t \alpha(s)ds + n(t) \tag{4.46}$$

where $n(t)$ is white data noise. Let

$$\beta(t) = \int_o^t \alpha(s)ds \tag{4.47}$$

Then, applying Section 4.3, the maximum likelihood estimates (denoted by *) of v_o and $\beta(t)$ are found from

$$v_o^*(T) = \frac{1}{T}\int_0^T \left[z(t) - \beta^*(t) \right] dt \tag{4.48}$$

$$\dot{\beta}^*(t) = -\lambda^2 \int_o^t \left[z(s) - \beta^*(s) - v_o^* \right] ds \tag{4.49}$$

where

$$\lambda = \left(\frac{\sigma_w}{\sigma_n} \right).$$

If $\alpha(t)$ is indeed white acceleration noise the value of λ is determined; otherwise, λ is a parameter to be chosen. In this section the dependence of the estimation error statistics upon λ will be examined.

4.6.2 The Estimation Error Statistics

Let

$$\varepsilon_v = v_0^* - v_0 \tag{4.50}$$

$$\varepsilon_\beta(t) = \beta^*(t) - \beta(t) \tag{4.51}$$

then, substituting equation (4.46),

$$\varepsilon_v = \frac{1}{T}\int_0^T \left[n(t) - \varepsilon_\beta(t) \right] dt \tag{4.52}$$

$$\dot{\varepsilon}_\beta = -\lambda^2 \int_o^t \left[n(s) - \varepsilon_\beta(s) - \varepsilon_v \right] ds - \alpha \tag{4.53}$$

and

$$\ddot{\varepsilon}_\beta - \lambda^2 \varepsilon_\beta = \lambda^2 \left[\varepsilon_v - n - \frac{\dot{\alpha}}{\lambda^2} \right] \tag{4.54}$$

Since

$$\varepsilon_\beta(0) = \beta^*(0) - \beta(0) \equiv 0 \tag{4.55}$$

$$\dot{\varepsilon}_\beta(0) = -\alpha(0) \tag{4.56}$$

the solution of this equation is

$$\varepsilon_\beta(t) = -\left(\frac{\alpha(0)}{\lambda}\right) \sinh \lambda t$$

$$+ \lambda \int_o^t \sinh \lambda(t-s) \left[\varepsilon_v - n(s) - \frac{\dot{\alpha}(s)}{\lambda^2} \right] ds$$

$$= \lambda \int_o^t \sinh \lambda(t-s) \left[\varepsilon_v - n(s) \right] ds$$

$$- \int_o^t \cosh \lambda(t-s)\, \alpha(s) ds \tag{4.57}$$

where we have integrated by parts to eliminate the $\dot{\alpha}(s)$ and $\dot{\alpha}(o)$ terms. Assuming for the moment that $E[\varepsilon_v^2]$ is negligibly small, the integrated acceleration (β) estimation error is described by the autocorrelation function

$$r(t,\tau) = E\left[\varepsilon_\beta(t)\ \varepsilon_\beta(\tau)\right]$$

$$= \lambda^2\sigma_n^2 \int_o^\tau \sinh\lambda(t-s)\ \sinh\lambda(\tau-s)ds$$

$$+ \int_o^t\int_o^\tau \cosh\lambda(t-s_1)\cosh\lambda(\tau-s_2)R(s_1,s_2)ds_1ds_2 \qquad (4.58)$$

where $t \geq \tau$ and

$$E\left[n(s_1)n(s_2)\right] = \sigma_n^2\,\delta(s_1-s_2) \qquad (4.59)$$

$$E\left[\alpha(s_1)\alpha(s_2)\right] = R(s_1, s_2) \qquad (4.60)$$

Note that $E[\varepsilon_\beta^2(t)] = r(t,t)$. The initial condition estimation error is described by (see equation (4.52):

$$\sigma_v^2 = E\left[\varepsilon_v^2\right] = \frac{\sigma_n^2}{T}\left[\frac{2\sinh\lambda T}{\lambda T} - 1\right]$$

$$+ \frac{1}{T^2}\int_o^T\int_o^T r(s_2,s_1)ds_1ds_2\ , \qquad (4.61)$$

since, from equation (4.57),

$$E\left[n(s_1)\varepsilon_\beta(s_2)\right] = -\lambda\sigma_n^2 \int_0^{s_2} \sinh \lambda(s_2 - s)\, \delta(s_1 - s)\, ds$$

$$= \begin{cases} -(\lambda\sigma_n^2) \sinh \lambda(s_2 - s_1) & s_2 \geq s_1 \\ 0 & s_1 > s_2 \end{cases} \tag{4.62}$$

and hence

$$-\left(\frac{2}{T^2}\right) \int_0^T \int_0^T E\left[n(s_1)\varepsilon_\beta(s_2)\right] ds_1 ds_2 = 2\left(\frac{\sigma_n^2}{\lambda T^2}\right)\left(\sinh \lambda T - \lambda T\right) \tag{4.63}$$

We also have

$$\int_0^T \int_0^T E\left[n(s_1)n(s_2)\right] ds_1 ds_2 = \sigma_n^2 T \tag{4.64}$$

Equations (4.58) and (4.61) are the results we seek.

4.6.3 Selection of the Tracking Time and Estimation Parameter

The values of T and λ can be chosen so as to minimize a bound on the error variance σ_v^2. Since $\alpha(t)$ is not white noise there is an M such that

$$R(s_1, s_2) \leq M \tag{4.65}$$

From equation (4.48) we have

$$r(t,\tau) \leq \frac{\lambda^2\sigma_n^2}{2}\int_0^\tau [\cosh \lambda(t+\tau-2s)-\cosh \lambda(t-\tau)]\,ds$$

$$+ M\int_0^t \cosh \lambda(t-s)ds \int_0^\tau \cosh \lambda(\tau-s)ds$$

$$= \frac{\lambda\sigma_n^2}{2}[\cosh \lambda t \sinh \lambda\tau - \lambda\tau \cosh \lambda(t-\tau)]$$

$$+ \frac{M}{\lambda^2} \sinh \lambda t \sinh \lambda\tau \qquad (4.66)$$

Then the last term on the right hand side of equation (4.61) is bounded according to

$$\frac{1}{T^2}\int_0^T\int_0^T r(t,\tau)dt\,d\tau \leq \frac{\sigma_n^2}{2\lambda T^2}(\cosh \lambda T - 1)(\sinh \lambda T - \lambda T)$$

$$+ \frac{M}{\lambda^4 T^2}(\cosh \lambda T - 1)^2$$

and a bound on σ_v^2 has been established. Suppose that we assume that λT is sufficiently large so that the ($\sinh \lambda T \cosh \lambda T$) terms dominate equation (4.61), and define

$$\sigma_v^2(\text{bound}) = \frac{\lambda\sigma_n^2}{2k^2}[\cosh k\,(\sinh k - k) + 3 \sinh k - k] \qquad (4.67)$$

$$+ \frac{M}{\lambda^2 k^2}(\cosh k - 1)^2$$

where $k = \lambda T$. Minimizing equation (4.67) with respect to λ and T yields

$$\lambda = \left(\frac{4M}{\sigma_n^2}\right)^{1/3} \tag{4.68}$$

$$k = \lambda T \approx 1 \tag{4.69}$$

Thus we conclude that λ should be chosen according to equation (4.68) and the optimal tracking time is $T = 1/\lambda$. The corresponding bound on σ_v^2 is then

$$\sigma_v^2 \leq 1.48\ \lambda\ \sigma_n^2 \tag{4.70}$$

Note that we get the intuitively obvious result that $\lambda \to \infty$ and $T \to 0$ as $M \to \infty$, or as $\sigma_n^2 \to 0$. Conversely, $\lambda \to 0$ and $T \to \infty$ as $M \to 0$ or as $\sigma_n^2 \to \infty$. The latter case is usually assumed in orbit determination; that is, it is postulated that random acceleration is negligible and the entire random component of the data residual is attributed to data noise.

4.6.4 An Application to Lunar Orbiter Tracking

As an example of the application of these results, consider the case of tracking a spacecraft in orbit about the moon. Suppose the unmodeled gravitational effects produce a periodic acceleration of the form

†The dimensions of equation (4.68) are compatible, for if M is in units of (ft^2/sec^4) and σ_n^2 is in units of (ft^2/sec), we have λ in units of (1/sec).

$$\alpha(t) = a \cos \omega t \tag{4.71}$$

where $\sigma_a = 10^{-4}$ m/sec^2. The corresponding speed error would be

$$\beta(t) = \int_o^t \alpha(s)ds = \frac{a}{\omega} \sin \omega t \tag{4.72}$$

and the maximum β is described by $(\sigma_a/\omega) = 0.5$ m/sec for $1/\omega = 5{,}000$ sec (approximately 1.4 hours). The acceleration autocorrelation function is

$$R(t,\tau) = \sigma_a^2 \cos \omega t \cos \omega\tau \tag{4.73}$$

and $M = 10^{-8}$ m^2/sec^4. Let the doppler data noise variance be $\sigma_n^2 = 60(10^{-2})$ m^2/sec, corresponding to a one sigma "counted" doppler data error of 0.1 m/sec for a 60 second count. Then

$$\lambda = \left[\frac{4\ (10^{-8})}{60(10^{-2})}\right]^{1/3} = \frac{1}{245} \tag{4.74}$$

$$T = \frac{1}{\lambda} = 245 \text{ sec} \tag{4.75}$$

$$\sigma_v \leq .06 \text{ m/sec} \tag{4.76}$$

Alternatively, if $\sigma_n = 60(10^{-5})$, corresponding to a one sigma counted doppler error of 0.0032 m/sec for a 60 sec count, we have T = 24.5 sec. In this case, only a very short tracking interval is indicated, which implies that a sequential filtering technique should be employed with a dynamic model which includes state noise (a Kalman filter).

5. PROPERTIES OF SEQUENTIAL ORBIT DETERMINATION ALGORITHMS

5.1 INTRODUCTION

It is shown in Reference [6] that the iterated, weighted least squares (maximum likelihood), nonlinear orbit determination algorithm will converge under certain hypotheses. It is not possible, however, to give a similar proof for the particular sequence of computations known as the Kalman filter. In this method the information contained in early data points (or batches of data points) is stored in the form of an estimate and an error covariance matrix, and the revised estimate which incorporates the most recent data point (or batch of data points) is obtained as though the system were linear. In this case the neglected nonlinear effects could lead to divergence, which means that a given data point could cause the estimate to move away from the "true" value. A second source of divergence, which can also occur in the weighted least squares form of the estimate and in linear systems (where the weighted least squares and Kalman forms are theoretically identical), is numerical roundoff error. Both of these sources of divergence will be discussed in this Section, and a theorem which is the basis for sequential estimation will be presented. In Section 6 an analysis of the effect of nonlinearities in certain sequential estimators will be carried out in detail.

5.2 THE LEAST SQUARES ALGORITHM

We measure an N dimensional data vector z_M (that is, we take N observations), and calculate for corresponding times the N dimensional vector $z_C(x)$ which would be observed if the spacecraft were on the trajectory implied by the initial state vector x and the given perturbation model. We take the difference between the z_M and z_C, called the residual, weight it according to the reciprocal of the standard deviation expected for that observation plus any cross correlations between observations. These are generalized in an N x N variance-covariance matrix W^{-1}. If the apriori variance of the initial conditions is infinite ($\Lambda_o^{-1} = 0$), then the orbit determination problem is to determine $x = x_o$ so as to minimize the scalar expression

$$r(x) = \left(z_C(x) - z_M\right)^T W \left(z_C(x) - z_M\right) \tag{5.1}$$

If a decomposition of W can be found of the form $W = R^T R$, and if new variables $g_C(x) = R\, z_C(x)$, $g_M = R\, z_M$ are introduced, (5.1) becomes

$$r(x) = \left(g_C(x) - g_M\right)^T \left(g_C(x) - g_M\right), \quad \text{or}$$

$$= \left\| g_C(x) - g_M \right\|^2 \tag{5.2}$$

A program which minimizes (5.1) is called a minimum variance program, and a program which minimizes (5.2) is called a least squares program. It is sometimes said that the minimum variance formulation is more general than the least squares formulation and that to solve the minimum variance problem it is necessary to invert a very large matrix, W^{-1}. Neither statement is quite true, since, as seen above, any least squares program that permits the above sort of change of variable will solve the minimum variance problem, and, in simplifying, it is only necessary to know the matrix R; it is not necessary to know W. The latter fact is important since R is generally simpler to compute and, in practical cases, is a simpler matrix to handle than W.

We consider that we want to minimize expression (5.2), which we rewrite as

$$r(x) = \left\| h(x) \right\|^2$$

where $h(x) = g_C(x) - g_M$. Let A be the matrix of partial derivatives of the components of g_C (and hence of h) with respect to the components of x. The partial derivatives are evaluated at the point x_o. Let x_o be an approximation. Then, expanding for a first order Taylor series about x_o, we have

$$r(x) = \left\| h(x_o) + A(x - x_o) \right\|^2 \tag{5.3}$$

Let $\Phi(v)$ be defined by

$$\Phi(v) = \|h_o + Av\|^2$$

where $v = x - x_o$ and $h_o = h(x_o)$. Then (5.3) can be written $r(x) = \Phi(v)$. Hence, choosing x to minimize $r(x)$ is approximately the same as choosing v to minimize $\Phi(v)$.

The problem of minimizing $\Phi(v)$ is a linear, least squares curve fitting problem with the solution

$$v_o = - \left(A^T A\right)^{-1} A^T h_o$$

The differential correction process consists of finding the vector v_o, letting $x_o + v_o$ be the next approximation, and repeating the process.

The above abstraction is actually much oversimplified. In the practical problem, for example, $\Phi(v)$ is minimized subject to the side condition that the components of v should be within prescribed bounds. This prevents divergence in many nearly singular and/or nonlinear problems. As the iterations proceed, the bounds are permitted to grow or forced to shrink, depending on whether the iterations are successful.

5.3 SIMPLE SEQUENTIAL PROCESSING

In the following, a subscript of 1 indicates old data, and a subscript of 2 indicates new data. The problem is to minimize $\Phi_1(v) + \Phi_2(v)$, where $\Phi_i(v) = |h_{oi} + A_i v|^2$. Before the new data comes in, one can already find the solution

$$v_1 = - \left(A_1^T A_1\right)^{-1} A_1^T h_{o1}$$

With the new data the solution is

$$v_2 = - \left(A_1^T A_1 + A_2^T A_2\right)^{-1} \left(A_1^T h_{o1} + A_2^T h_{o2}\right) \qquad (5.4)$$

Note that one can find v_2 by adding the "new normal matrix" $A_2^T A_2$ to the "old normal matrix" $A_1^T A_1$, and the new set of coefficients $A_2^T h_{o2}$ to the old set of coefficients $A_1^T h_{o1}$. Hence it is, in general, not necessary to solve the whole problem anew each time new date comes in.

Note that $x_o + v_1$ is the solution for the old data, and $x_o + v_2$ is the solution with all data. Note that both corrections are intended to be added to the same nominal vector. Since the basic problem is nonlinear, after all the data has been processed, the final derived correction should be added to x_o, and the process then repeated with all data collectively. The above process is then the ordinary differential correction process, except that one can point out intermediate results; one simply takes the matrix A^TA and the vector A^Th_o for the data "so far," and prints out the solution

$$\left(A^T A\right)^{-1} A^T h_o$$

for this data. Only when all the data has been processed does one actually use the solution v as a correction to x_o; the intermediate results are only for information.

One can prove the convergence of the general method described above under suitable hypotheses; but this is beyond the present scope. Two things are fairly simple however: (a) for a process to converge, it should have the property that if one starts out with the right answer, one stays there; and (b) if the normal matrix

$$\left(A^T A\right)^{-1}$$

is nonsingular, the method under discussion has property (a). To prove (b), one notes that the vector of partial derivatives of r with respect to the components of x is just

$$A^T h_o + \left(A^T h_o\right)^T$$

Hence at a solution

$$A^T h_o = 0, \quad v = \left(A^T A\right)^{-1} A^T h_o = 0.$$

5.4 SEQUENTIAL PROCESSING WITH SHIFTING NOMINAL TRAJECTORY

In this method, one finds the solution v_1 for the old data, corrects the nominal trajectory immediately by $x_1 = x_o + v_1$, and processes the new data with the new nominal. This results in a slight simplification since the $A_1^T h_{o1}$ term drops out of the right side of (5.4), simply because $A^T h = 0$ is a solution to the problem.

More precisely, let a superscript of (0) mean evaluated at x_o, and a superscript of (1) mean evaluated at x_1. Then the sequential processing method consists of evaluating

$$v_1 = -\left[\left(A_1^{(0)}\right)^T \left(A_1^{(0)}\right)\right]^{-1} \left(A_1^{(0)}\right)^T h_1^{(0)}$$

$$x_1 = x_o + v_1$$

$$v_2 = -\left[\left(A_2^{(1)}\right)^T A_2^{(1)} + \left(A_1^{(0)}\right)^T A_1^{(0)}\right]^{-1} \left(A_2^{(1)}\right)^T h_2^{(1)}$$

$$x_2 = x_1 + v_2$$

As the processing continues, the nominal value of x changes each time new data is added. If the process is not iterated the algorithm is the Kalman filter.

The claim is sometimes made that the above process avoids the need for iterations. A more accurate statement would be that it prevents one from iterating even though one should. The difficulty is that even if one starts out with the correct answer, one does not stay there, and hence the previously mentioned property of converging processes is not met.

To prove the last statement, let $x^{(0)}$ be the true solution. Then

$$\left(A_1^{(0)}\right)^T h_1^{(0)} + \left(A_2^{(0)}\right)^T h_2^{(0)} = 0$$

Now excluding the exceptional case when $x^{(0)}$ is a solution to the old data alone and the new data alone:

$$\left(A_1^{(0)}\right)^T h_1^{(0)} \neq 0$$

Hence $v_1 \neq 0$ and $x_1 \neq x_o$. Expanding h_2 in a Taylor series about x_o gives

$$h_2^{(1)} = h_2^{(0)} + A_2^{(0)} v_1 + \epsilon$$

where in general the remainder term ϵ is nonzero.

Let

$$E = A_2^{(1)} - A_2^{(0)},$$

$$F = \left(A_2^{(1)}\right)^T A_2^{(1)} - \left(A_2^{(0)}\right)^T A_2^{(0)}$$

v_2 satisfies

$$\left[\left(A_1^{(0)}\right)^T A_1^{(0)} + \left(A_2^{(1)}\right)^T A_2^{(1)}\right] v_2 = -\left(A_2^{(1)}\right)^T h_2^{(1)}$$

$$\left[\left(A_1^{(0)}\right)^T A_1^{(0)} + \left(A_2^{(0)}\right)^T A_2^{(0)} + F\right] v_2 = -\left(A_2^{(0)}\right)^T h_2^{(1)} - E^T h_2^{(1)}$$

$$= -\left(A_2^{(0)}\right)^T h_2^{(0)} - \left(A_2^{(0)}\right)^T A_2^{(0)} v_1$$

$$- \left(A_2^{(0)}\right)^T \epsilon - E^T h_2^{(1)}$$

$$= -\left(A_1^{(0)}\right)^T h_1^{(0)} - \left(A_2^{(0)}\right)^T h_2^{(0)}$$

$$- \left(A_1^{(0)}\right)^T A_1^{(0)} v_1$$

$$- \left(A_2^{(0)}\right)^T A_2^{(0)} v_1 - \left(A_2^{(0)}\right)^T \epsilon$$

$$- E^T h_2^{(1)}$$

Hence

$$\left[\left(A_1^{(0)}\right)^T A_1^{(0)} + \left(A_2^{(0)}\right)^T A_2^{(0)}\right] (v_1 + v_2) = -F v_2 - \left(A_2^{(0)}\right)^T \epsilon - E^T h_2^{(1)}$$

This shows that if the problem is linear (that is, if $E = F = \epsilon = 0$) then $v_1 + v_2 = 0$, and we would stay at the correct solution. If the problem is not linear, however, $v_1 + v_2 \neq 0$, and we would not stay at the correct

solution. Iterating on the data obviously will not correct this problem.

5.5 RATE OF NONLINEAR DIVERGENCE OF THE KALMAN FILTER

In this section we will investigate the extent to which the non-iterated sequential processing methods with shifting nominal (the Kalman filter) does not converge. Instead of attempting to prove convergence, we will outline a proof that under certain hypotheses the divergence is not too severe.

The basic problem is to choose x to minimize $|h(x)|^2$. A necessary condition for the minimization of h(x) is that

$$A^T h(x) = 0$$

where the partial derivative matrix A is evaluated at x.

In the Kalman filter method, suppose that the problem has been solved for all previous data, new data has arrived, and one wishes to solve the problem with the old and new data combined. A subscript of 1 will refer to old data and a subscript of 2 will refer to new data. The vector x_1 is the orbital element vector after processing the old data and x_2 is the orbital element vector after including the new data. A superscript of 1 and 2 will refer to where a matrix or vector is evaluated. For example $h_2^{(1)}$ means the residual vector for the new data evaluated at the old orbital element vector x_1.

After having processed the old data, one has an error

$$E_1 = (A_1^{(1)})^T h_1^{(1)}$$

If the old data had been processed exactly we would have $E_1 = 0$.

In the sequential processing method, one computes a correction v by solving the system of equations

$$(N_1^{(1)} + N_2^{(1)})v = - (A_2^{(1)})^T h_2^{(1)}$$

and letting

$$x_2 = x_1 + v \quad .$$

The main problem now is to evaluate

$$E_2 = (A_1^{(2)})^T h_1^{(2)} + (A_2^{(2)})^T h_2^{(2)} \quad .$$

Expanding $h_1^{(2)}$ and $h_2^{(2)}$ in power series about the point x_1 gives

$$h_1^{(2)} = h_1^{(1)} + A_1^{(1)} v + \varepsilon_1$$

$$h_2^{(2)} = h_2^{(1)} + A_2^{(1)} v + \varepsilon_2 \quad .$$

If the functions were linear, the errors ε_1, ε_2 would be zero. Also, the errors

$$P_1 = A_1^{(2)} - A_1^{(1)}$$

$$P_2 = A_2^{(2)} - A_2^{(1)}$$

would be zero if the functions were linear.

$$E_2 = (A_1^{(2)})^T h_1^{(2)} + (A_2^{(2)})^T h_2^{(2)}$$

$$= (A_1^{(1)} + P_1)^T (h_1^{(1)} + A_1^{(1)} v + \varepsilon_1) + (A_2^{(1)} + P_2)^T (h_2^{(1)} + A_2^{(1)} v + \varepsilon_2)$$

$$= (A_1^{(1)})^T h_1^{(1)} + N_1^{(1)} v + (A_1^{(1)})^T \varepsilon_1 + P_1^T h_1^{(2)}$$

$$+ (A_2^{(1)})^T h_2^{(1)} + N_2^{(1)} v + (A_2^{(1)})^T \varepsilon_2 + P_2^T h_2^{(2)}$$

$$= E_1 + [(N_1^{(1)} + N_2^{(2)}) v + (A_2^{(1)})^T h_2^{(1)}]$$

$$+ (A_1^{(1)})^T \varepsilon_1 + (A_2^{(1)})^T \varepsilon_2 + P_1^T h_1^{(2)} + P_2^T h_2^{(2)}$$

$$E_2 = E_1 + (A_1^{(1)})^T \varepsilon_1 + (A_2^{(1)})^T \varepsilon_2 + P_1^T h_1^{(2)} + P_2^T h_2^{(2)}$$

This formula indicates the growth of the error in the Kalman filter processing method. The added error in each step occurs because of the nonlinear terms reflected by ε_1, ε_2, P_1, P_2. (By iterating on the new data it is possible to get rid of the terms involving ε_2 and P_2, as will be discussed in Section 6.)

It is possible to use the mean value theorem to express the errors ε_1, ε_2, P_1, P_2 in terms of second partial derivatives of the residual components with respect to components of x_j and then use the above formula to prove a rigorous theorem about the growth of the error in terms of bounds on these second partial derivatives. Since these bounds are usually not estimated easily it is felt that such a theorem would be of largely academic interest.

5.6 DIVERGENCE DUE TO NUMERICAL ROUNDOFF ERROR

It has been frequently observed that the sequence of computations called a Kalman filter sometimes produces numerically unstable results, even in the linear case. This can be caused by numerical roundoff error. In order to shed more light on this phenomenon, we will examine fairly completely the simplest non trivial problem in which a Kalman filter is applied. Even in this simple case, some of the results are fairly hard to obtain. It is hoped that a complete understanding of the simple case is useful in understanding the more general case.

The results which will be shown are: (a) The "instability" is not really an instability in the usual sense. It is an increase in the variance and a bias in the mean of the estimate, and the increase and bias can be theoretically predicted. (b) The same phenomenon occurs by the "least squares" algorithm, but it shows up later.

The case considered here is the simple problem of finding the mean of a set of number y_i with mean $\bar{y}$ and standard deviation 1.

$$a_k = \frac{1}{k} \sum_{j=1}^{k} y_j$$

We will consider two algorithm for computing y_k sequentially:

$$\text{I}: \quad C_o = 0$$

$$S_o = 0$$

$$\left.\begin{aligned} C_{k+1} &= C_k + 1 \\ S_{k+1} &= S_k + y_k \\ \bar{a}_{k+1} &= \frac{S_{k+1}}{C_{k+1}} \end{aligned}\right\} \quad k = 0, 1, \ldots$$

$$\text{II}: \quad v_1 = 1$$

$$S_1 = y_1$$

$$v_{k+1} = v_k - v_k^2 \, (1 + v_k)^{-1}$$

$$\bar{a}_{k+1} = v_{k+1} \, S_{k+1}$$

These two methods are generally called the least squares method and the Kalman filter method in somewhat harder problems. (Actually method 2 is the "Kalman filter method without shifting nominal". This fact does not affect the present argument.)

If the arithmetic were done exactly, the results of both processes I and II would be the same. However, since the computations are done on a computer with finite word length, the results will be different.

Consider a floating point machine with, say, 3 decimal digits. Suppose that S_k is computed exactly. This is a reasonable assumption if the mean of the random numbers y_k is near zero; in this case there is no appreciable growth of roundoff error in the computation of S_k.

In method 1, there will be no error at first, and one will have $C_k = k$. As soon as $k = 1{,}000$, however, the computation $1000 + 1$ will result in 1000 on the 3 place computer. Successively adding 1 will not change the succeeding values of C_k. Hence the computer will produce for method I

$$\text{I: (computer)} \quad a_k = \frac{1}{k} \sum_1^k y_j \qquad \text{for } k \le 1000$$

$$a_k = \frac{1}{1000} \sum_1^k y_j \qquad \text{for } k > 1000 \quad .$$

The mean and standard deviation of a_k can be easily computed

$$\bar{a}_k = \bar{y},\ \sigma_k = 1/\sqrt{k}\ , \qquad k \le 1000$$

$$\bar{a}_k = \frac{k}{1000}\,\bar{y}\ ,\ \sigma_k = \frac{\sqrt{k}}{1000}\ ;\ k > 1000 \quad .$$

This shows that the computer estimate of the mean will be biased and that the standard deviation of the estimate will grow for $k \ge 1000$.

An analysis of method II can be given in similar fashion. There is a difference in that the computations are not done exactly for early values of k; one can however show that the roundoff errors are stable. Hence assume that the computations in method II are carried out exactly for early value of k; $v_k = \frac{1}{k}$ at first. The computations will break down when $v_k + 1$ is indistinguishable from 1 on the computer. This happens when $v_k = .005$ or $k = 200$. At this point also, v_k^2 is indistinguishable from v_k and we have $v_{k+1} = v_k$ for $k \ge 200$. Hence method II produces the result

$$\text{II: (computer)} \quad a_k = \frac{1}{k} \sum_{j=1}^k y_j\ , \qquad k \le 200$$

$$a_k = \frac{1}{500} \sum_{j=1}^k y_j\ , \qquad k > 200 \quad .$$

The two methods are compared in Figures 5.1 and 5.2 with the results with no machine error. They illustrate that there is in fact a bias and and increasing standard deviation for both methods I and II.

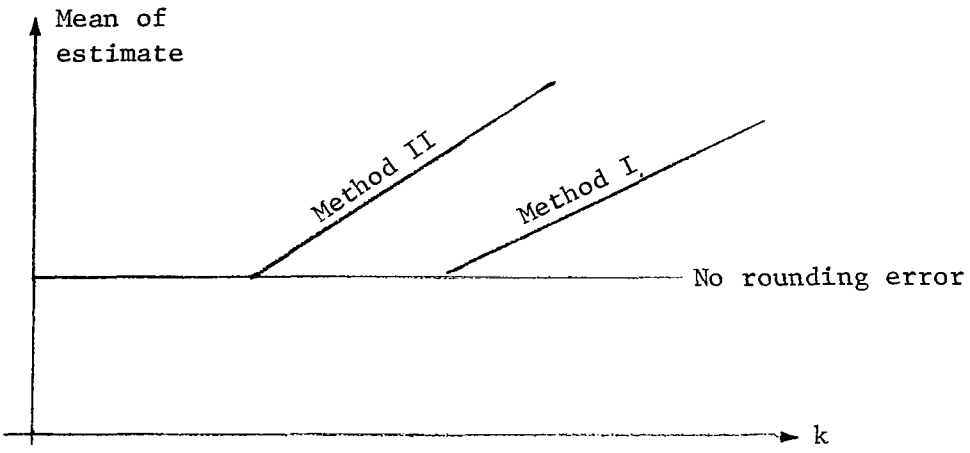

Figure 5.1

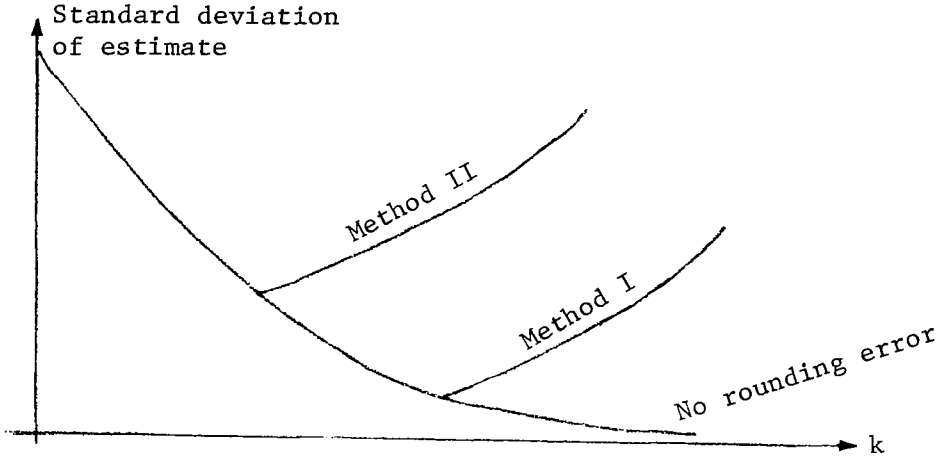

Figure 5.2

One might object to the above simple case as being unrealistic for the reason that before the "instability" would show up on a realistic computer one would have to process an unrealistic number of data points; for example, 10^{38} data points on an IBM 7094, which would require about 10^{24} years. The problem would show up in a realistic time however, if one includes an apriori first estimate of the mean with a small standard deviation. When this happens, the above analysis still holds, except that the instability occurs somewhat sooner.

5.7 A THEOREM FUNDAMENTAL TO SEQUENTIAL DATA PROCESSING

This Section concerns a method of obtaining and understanding a number of identities which are important in sequential orbit determination and other branches of applied mathematics. The idea is that a large number of results come from one simple theorem. It is hoped that this will provide some further insight into the nature of computations which are performed now and also suggest some new methods.

Let m and n be two integers. While the theorem below does not depend on the relative size of m and n, the applications are generally for $m > n$. For example, in the orbit determination problem one typically has $m = 6$, $n = 1$.

Let B, D, and A be fixed matrices. The matrix P is a variable matrix in the theorem below. The dimensions of these matrices are

A: m x m
B: m x n
D: n x m
P: n x n

It is assumed that $(DAB)^{-1}$ exists.

Now define the m x m matrix V_P as follows

$$\text{Definition:} \quad V_P = I + B(DAB)^{-1} (P - I) DA \tag{5.5}$$

Note that the computation of V_P requires the inversion of an n x n matrix.

The basic result is

$$\text{Theorem:} \quad V_P V_Q = V_{PQ} \tag{5.6}$$

The theorem is easily verified by simply multiplying the two matrices V_P, V_Q and collecting terms.

The importance of the theorem is that a certain subset of m x m matrices is isomorphic to the multiplicative group of n x n matrices. In other words, matrices which look like the matrix V_P can be operated on (multiplication, inversion, square root extraction) by doing these operatior

on the smaller matrices P. Since the dimension of P is smaller than the dimension of V_P a savings in computer time can result. This will be made clearer by the examples below.

The following results can be obtained immediately from the theorem

$$V_I = I \tag{5.7}$$

$$V_P^{-1} = V_{P^{-1}} \tag{5.8}$$

$$V_P^{1/2} = V_{P^{1/2}} \tag{5.9}$$

Application 1: The Schur Identity

As a first application, consider, the problem of finding the inverse of a matrix of the form $W^{-1} + TUV$. One can write

$$\begin{aligned}(W^{-1} + TUV)^{-1} &= W\,(I + TUVS^{-1})^{-1} \\ &= W\,V_P^{-1} \\ &= W\,V_{P^{-1}}\end{aligned} \tag{5.10}$$

To identify the matrix $I + TUVS^{-1}$ with the matrix V_P, we let

$$\begin{aligned}T &= B \\ U &= (DAB)^{-1}\,(P - I) \\ V &= D \\ W &= A\end{aligned} \tag{5.11}$$

The matrix P is hence computed from

$$P = I + (VWT)\,U \tag{5.12}$$

Now suppose that W is known and one needs the inverse of $(W^{-1} + TUV)$. One can obtain the inverse of this large (m x m) matrix by forming and inverting the small (n x n) matrix P, finding $V_{P^{-1}}$ and then $W V_{P^{-1}}$.

The above result is a disguised form of the well known matrix identity:

$$(W^{-1} + TUV)^{-1} = W - W\,T(U^{-1} + VWT)^{-1}\,VW \tag{5.13}$$

To prove this note that

$$P^{-1} = U^{-1}\,(U^{-1} + VWT)^{-1} \tag{5.14}$$

and

$$\begin{aligned} V_{P^{-1}} &= I + T(VWT)^{-1}\,(P^{-1} - I)\,VW \\ &= I - T(VWT)^{-1}\,(P - I)\,P^{-1}\,VW \\ &= I - T(VWT)^{-1}\,(VWT)\,VU^{-1}\,(U^{-1} + VWT)^{-1}\,VW \\ &= I - T(U^{-1} + VWT)^{-1}\,VS^{-1} \end{aligned} \tag{5.15}$$

It then follows that $W V_{P^{-1}}$ agrees with the right hand side of (5.13).

Application 2: Orbit Determination - Kalman Filter

The matrix

$$(N + \beta\,\beta')^{-1}$$

appears in orbit determination theory; usually with N and β being 6 x 6 and 6 x 1 matrices respectively. If one knows the inverse of N one can find the inverse of $(N + \beta\,\beta')$ as in the last section.

$$(N + \beta\ \beta')^{-1} = N^{-1}\ (I + \beta\ \beta'\ N^{-1})^{-1}$$

$$= N^{-1}\ V_P^{-1}$$

$$= N^{-1}\ V_{P^{-1}}$$

where

$$N = A^{-1}$$

$$\beta = B$$

$$\beta' = D$$

$$P = 1 + \beta'\ N^{-1}\ \beta \text{ (scalar)} \quad .$$

If one performs the algebra corresponding to equation (5.15) above, the result is the "Kalman filter" equation. This is expected, since the Kalman filter equation is known to be a special case of the Schur identity.

Application 3: Orbit Determination - Square Root Method

Suppose that one knows a matrix R such that

$$R'\ R = C$$

and one wishes to find a matrix R_1 so that

$$(C^{-1} + \beta\ \beta')^{-1} = R_1'\ R_1 \quad .$$

This can be done as follows

$$(C^{-1} + \beta\ \beta')^{-1} = R'\ (I + R\ \beta\ \beta'\ R')^{-1}\ R$$

$$= R'\ V_P\ R$$

$$= R'\ V_{P^{1/2}}\ V_{P^{1/2}}\ R$$

$$= (V_{P^{1/2}}\ R)'\ (V_{P^{1/2}}\ R)$$

Hence $R_1 = V_{P^{1/2}} R$.

In the above derivation we set

$$A = I$$

$$B = R\beta$$

$$D = \beta' R'$$

$$(DB)^{-1} (P - 1) = 1$$

$$P = 1 + |R \beta|^2 \quad .$$

We also used the fact that $V_{P^{1/2}}$ is symmetric.

Application 4: Variations (Partial Derivative)

In orbit determination, one frequently has occasion to compute the derivatives of a function like

$$f(x) = X |X|^P \quad .$$

Where X is a vector; the variation of this expression is given by

$$\delta f(x) = |X|^P V_{P+1} \delta X$$

where

$$V_{P+1} = I + P \frac{XX^T}{|X|^2} \quad .$$

This expression is useful in deriving, for example Taylor series methods for integrating equations. It is also useful in deriving variational equations. For example, if the equations of motion are

$$\ddot{X} = - \mu \frac{X}{|X|^3}$$

one can write the variational equations immediately:

$$\ddot{(\delta X)} = - \frac{\mu}{|X|^{-3}} V_{-2} \, \delta X \quad .$$

6. ESTIMATION WITH A SLIGHT NON-LINEARITY

6.1 INTRODUCTION

The effect of non-linearities has been discussed throughout this Report. In general, little can be said about this source of error, or about any corrective approximation which might be introduced, unless the non-linearities are small.

It will be assumed that the dynamics is noise-free so that only the data is contaminated by noise. The state variables x of the system will here be chosen to be certain constants of the motion, such as orbital elements or initial position and velocity. The only non-linearities arising are then those in the relations between the observations y_i and the state x. These non-linearities will be assumed small. We shall consider, then, the problem of estimating quantities x_α, given a set of observations

$$y_i = a_{i\alpha}\, x_\alpha + \varepsilon_{i\alpha\beta}\, x_\alpha\, x_\beta + n_i \tag{6.1}$$

where the n_i's are independent standard normal variables, and summation is understood over repeated Greek indices. The coefficients $\varepsilon_{i\alpha\beta}$ of the non-linear terms are understood to be small, and their squares and products will be neglected throughout. It may be assumed that the apriori covariance Λ of the x_α is infinite; alternatively, apriori information may be included as a first data point.

Three forms of the estimator will be analyzed and their biases compared. These will be (i) The least squares fit, (ii) The sequential (Kalman) estimate with linearization of the latest residual about the previous estimate, and (iii) An "iterated-sequential" scheme involving iteration through the latest data point designed to reduce the non-linearity in the latest residual.

6.2 THE LEAST SQUARES FIT

This estimator obtains that set of values x_α which cause

$$\sum_i (y_i - a_{i\alpha}\, x_\alpha - \varepsilon_{i\alpha\beta}\, x_\alpha\, x_\beta)^2 = \text{minimum} \tag{6.2}$$

Thus

$$\sum_i (a_{i\alpha} + 2\varepsilon_{i\alpha\gamma}\hat{x}_\gamma)(y_i - a_{i\beta}\hat{x}_\beta - \varepsilon_{i\beta\delta}\hat{x}_\beta\hat{x}_\delta) = 0 \qquad (6.3)$$

That is, if $\bar{N}_{\alpha\beta}$ denotes $\sum_i a_{i\alpha}a_{i\beta}$, then

$$\bar{N}_{\alpha\beta}\hat{x}_\beta = \sum_i a_{i\alpha}y_i - \sum_i a_{i\alpha}\varepsilon_{i\beta\gamma}\hat{x}_\beta\hat{x}_\gamma + 2\sum_i \varepsilon_{i\alpha\beta}\hat{x}_\beta y_i$$

$$- 2\sum_i a_{i\beta}\varepsilon_{i\alpha\gamma}\hat{x}_\beta\hat{x}_\gamma \quad , \qquad (6.4)$$

In the right member of (6.4) (neglecting ε^2), $\hat{x}_\beta$ is replaceable by

$$(\bar{N})^{-1}_{\beta\alpha} \sum_i a_{i\alpha}y_i$$

Substituting for y_i in terms of n_i and x_α, the distribution of $\hat{x}_\alpha - x_\alpha$ may be investigated.
In particular:

$$E\left\{\hat{x}_\alpha - x_\alpha\right\} = -\left(\sum_i a_{i\beta}\varepsilon_{i\gamma\delta}\right)(\bar{N})^{-1}_{\alpha\beta}(\bar{N})^{-1}_{\gamma\delta} \quad , \qquad (6.5)$$

and

$$E\left\{(\hat{x}_\alpha - x_\alpha)(\hat{x}_\beta - x_\beta)\middle| x\right\}$$

$$= (\bar{N})^{-1}_{\alpha\beta} - 2(\bar{N})^{-1}_{\alpha\gamma}(\bar{N})^{-1}_{\beta\gamma}\sum_i (a_{i\gamma}\varepsilon_{i\delta\varepsilon} + a_{i\delta}\varepsilon_{i\gamma\varepsilon})x_\varepsilon \qquad (6.6)$$

6.3 THE SEQUENTIAL FIT

This estimate is obtained successively as follows:

$\hat{x}^{(i)}_\alpha$ is that x_α which causes the sum of previous squares, represented by

$$N_{\alpha\beta}^{(i-1)} \left(x_\alpha - \hat{x}_\alpha^{(i-1)}\right)\left(x_\beta - \hat{x}_\beta^{(i-1)}\right) + \rho_i^2 (x) = \text{minimum}, \quad (6.7)$$

where the first term on the left represents the sum of squares of previous residuals, and where $\rho_i(x)$ is the i^{th} residual linearized around $\hat{x}^{(i-1)}$, i.e.,

$$\rho_i (x) = y_i - a_{i\alpha}\hat{x}_\alpha^{(i-1)} - \varepsilon_{i\alpha\beta}\hat{x}_\beta^{(i-1)}\hat{x}_\alpha^{(i-1)} - \left(a_{i\alpha} + 2\varepsilon_{i\alpha\beta}\hat{x}_\beta^{(i-1)}\right) \times \left(x_\alpha - \hat{x}_\alpha^{(i-1)}\right) \quad (6.8)$$

and

$$N_{\alpha\beta}^{(i)} = N_{\alpha\beta}^{(i-1)} + \left(a_{i\alpha} + 2\varepsilon_{i\alpha\gamma}\hat{x}_\gamma^{(i-1)}\right)\left(a_{i\beta} + 2\varepsilon_{i\beta\delta}\hat{x}_\delta^{(i-1)}\right) . \quad (6.9)$$

Thus

$$N_{\alpha\beta}^{(i)}\left(\hat{x}_\beta^{(i)} - \hat{x}_\beta^{(i-1)}\right) = \rho_i\left(\hat{x}^{(i-1)}\right)\left(a_{i\alpha} + 2\varepsilon_{i\alpha\gamma}\hat{x}_\gamma^{(i-1)}\right) . \quad (6.10)$$

Hence,

$$N_{\alpha\beta}^{(i)}\left(\hat{x}_\beta^{(i)} - x_\beta\right) - N_{\alpha\beta}^{(i-1)}\left(\hat{x}_\beta^{(i-1)} - x_\beta\right)$$

$$= \rho_i\left(\hat{x}^{(i-1)}\right)\left(a_{i\alpha} + 2\varepsilon_{i\alpha\gamma}\hat{x}_\gamma^{(i-1)}\right)$$

$$+ \left(a_{i\alpha} + 2\varepsilon_{i\alpha\gamma}\hat{x}_\gamma^{(i-1)}\right)\left(a_{i\beta} + 2\varepsilon_{i\beta\delta}\hat{x}_\delta^{(i-1)}\right)\left(\hat{x}_\beta^{(i-1)} - x_\beta\right) . \quad (6.11)$$

Substituting for y_i into ρ_i in terms of n_i and x, and summing over i from 1 to m(i.e., m data points):

$$N_{\alpha\beta}^{(m)} \left(\hat{x}_{\beta}^{(m)} - x_{\beta} \right) = \sum_{i=1}^{m} \left[n_i + \varepsilon_{i\beta\delta} \left(\hat{x}_{\beta}^{(i-1)} - x_{\beta} \right) \left(\hat{x}_{\delta}^{(i-1)} - x_{\delta} \right) \right]$$

$$\times \left[a_{i\alpha} + 2\varepsilon_{i\alpha\gamma} \hat{x}_{\gamma}^{(i-1)} \right] \quad , \tag{6.12}$$

In the right member of (6.12) the $\hat{x}_{\beta}^{(i-1)}$ is replaceable by

$$x_{\beta} + \left[\overline{N}^{(i-1)} \right]_{\beta\varepsilon}^{-1} \sum_{j=1}^{i-1} n_j \, a_{j\varepsilon} \quad , \text{ where}$$

$$\overline{N}_{\beta\varepsilon}^{(i-1)} = \sum_{j=1}^{i-1} a_{j\beta} \, a_{j\varepsilon} \quad . \tag{6.13}$$

In order to investigate the distribution of $\hat{x}_{\alpha}^{(m)} - x_{\alpha}$ it is necessary to observe that $N_{\alpha\beta}^{(m)}$ is itself a random variable. In fact,

$$N_{\alpha\beta}^{(m)} = \overline{N}_{\alpha\beta}^{(m)} + 2 \sum_{i=1}^{m} \left(a_{i\alpha} \, \varepsilon_{i\beta\gamma} + a_{i\beta} \, \varepsilon_{i\alpha\gamma} \right) \hat{x}_{\gamma}^{(i-1)} \quad , \tag{6.14}$$

so that

$$\left[N^{(m)} \right]_{\alpha\beta}^{-1} = \left[\overline{N}^{(m)} \right]_{\alpha\beta}^{-1} - 2 \left[\overline{N}^{(m)} \right]_{\alpha\gamma}^{-1} \left[\overline{N}^{(m)} \right]_{\beta\delta}^{-1} \sum_{i=1}^{m} \left(a_{i\gamma} \varepsilon_{i\delta\varepsilon} + a_{i\delta} \varepsilon_{i\gamma\varepsilon} \right) \hat{x}_{\varepsilon}^{(i-1)} . \tag{6.15}$$

We can now obtain:

$$\hat{x}_{\alpha}^{(m)} - x_{\alpha}$$

$$= \left[\overline{N}^{(m)} \right]_{\alpha\beta}^{-1} \sum_{i=1}^{m} \left[n_i \, a_{i\beta} + 2 \, n_i \, \varepsilon_{i\beta\gamma} \hat{x}_{\gamma}^{(i-1)} + a_{i\beta} \, \varepsilon_{i\gamma\delta} \left(\hat{x}_{\gamma}^{(i-1)} - x_{\gamma} \right) \right.$$

$$\times \left(\hat{x}_{\delta}^{(i-1)} - x_{\delta} \right) \Big]$$

$$- 2 \left[\bar{N}^{(m)} \right]^{-1}_{\alpha\gamma} \left[\bar{N}^{(m)} \right]^{-1}_{\beta\delta} \left\{ \sum_{i=1}^{m} \left[\left(a_{i\gamma} \varepsilon_{i\delta\varepsilon} + a_{i\delta} \varepsilon_{i\gamma\varepsilon} \right) \hat{x}_{\varepsilon}^{(i-1)} \right] \right\} \left(\sum_{j=1}^{m} n_j a_{j\beta} \right), \tag{6.16}$$

on the right of (6.16) we may use

$$\hat{x}_{\gamma}^{(i-1)} = x_{\gamma} + \left[\bar{N}^{(i-1)} \right]^{-1}_{\gamma\xi} \sum_{k=1}^{i-1} n_k a_{k\xi} \tag{6.17}$$

The distribution of $\hat{x}_{\alpha}^{(m)} - x_{\alpha}$ is now available. The covariance

$E\left\{ \left(\hat{x}_{\alpha}^{(m)} - x_{\alpha} \right) \left(\hat{x}_{\alpha}^{(m)} - x_{\beta} \right) \Big| \; x \right\}$ is the same as that for the least-squares fit to the same data, so that $\bar{\bar{N}}$ is identical with $\bar{N}^{(m)}$. The bias, however, is more serious:

$$E\left\{ \hat{x}_{\alpha}^{(m)} - x_{\alpha} \right\} = \left[\bar{N}^{(m)} \right]^{-1}_{\alpha\beta} \sum_{i=1}^{m} a_{i\beta} \varepsilon_{i\gamma\delta} \left[\bar{N}^{(i-1)} \right]^{-1}_{\gamma\delta}$$

$$- 2 \left[\bar{N}^{(m)} \right]^{-1}_{\alpha\gamma} \left[\bar{N}^{(m)} \right]^{-1}_{\beta\delta} \sum_{i=1}^{m} \left(a_{i\gamma} \varepsilon_{i\beta\delta} + a_{i\delta} \varepsilon_{i\beta\gamma} \right) . \tag{6.18}$$

6.4 THE ITERATED-SEQUENTIAL FIT

This estimator is obtained as follows:
$\hat{x}_{\alpha}^{*(i)}$ is that x_{α}, obtainable as the limit in an iterative procedure, which causes

$$N_{\alpha\beta}^{*(i-1)} \left(x_\alpha - \hat{x}_\alpha^{*(i-1)}\right) \left(x_\beta - \hat{x}_\beta^{*(i-1)}\right) + \rho_i^{*2}(x) \quad = \text{minimum} \qquad (6.19)$$

where $\rho_i^*(x)$ is the $i^{\underline{th}}$ residual linearized not around $\hat{x}^{*(i-1)}$ but around the desired $\hat{x}^{*(i)}$, and

$$N_{\alpha\beta}^{*(i)} = N_{\alpha\beta}^{*(i-1)} + \left(a_{i\alpha} + 2\,\varepsilon_{i\alpha\gamma}\,\hat{x}_\gamma^{*(i)}\right)\left(a_{i\beta} + 2\,\varepsilon_{i\beta\delta}\,\hat{x}_\delta^{*(i)}\right) \quad . \qquad (6.20)$$

The iterative procedure consists of linearizing ρ_i firstly around $\hat{x}^{*(i-1)}$, obtaining from the minimization a first estimate $\hat{x}^{*(i)1}$, re-linearizing around $\hat{x}^{*(i)1}$ to obtain a second estimate $\hat{x}^{*(i)2}$, etc. Since ε^2 is negligible, the next iteration is optional, and any further iteration pointless. We see that

$$N_{\alpha\beta}^{*(i-1)} \left(\hat{x}_\beta^{*(i)} - \hat{x}_\beta^{*(i-1)}\right) = \rho_i^{*}\left(\hat{x}^{*(i)}\right)\left(a_{i\alpha} + 2\varepsilon_{i\alpha\beta}\,\hat{x}_\beta^{*(i)}\right) \quad , \qquad (6.21)$$

and hence

$$N_{\alpha\beta}^{*(i)} \left(\hat{x}_\beta^{*(i)} - x_\beta\right) - N_{\alpha\beta}^{*(i-1)} \left(\hat{x}_\beta^{*(i-1)} - x_\beta\right)$$

$$= \rho_i \left(\hat{x}^{*(i)}\right)\left(a_{i\alpha} + 2\,\varepsilon_{i\alpha\gamma}\,\hat{x}_\gamma^{*(i)}\right)$$

$$+ \left(a_{i\alpha} + 2\varepsilon_{i\alpha\gamma}\hat{x}_\gamma^{*(i)}\right)\left(a_{i\beta} + 2\varepsilon_{i\beta\delta}\hat{x}_\delta^{*(i)}\right)\left(\hat{x}_\beta^{*(i)} - x_\beta\right) \quad . \qquad (6.22)$$

Substituting for y_i into ρ_i^* in terms of n_i and x, and summing over i from 1 to m,

$$N^{*(m)}_{\alpha\beta}\left(\hat{x}^{*(m)}_{\beta} - x_{\beta}\right) = \sum_{i=1}^{m}\left[n_i + \varepsilon_{i\beta\delta}\left(\hat{x}^{*(i)}_{\beta} - x_{\beta}\right)\left(\hat{x}^{*(i)}_{\delta} - x_{\delta}\right)\right]$$

$$\times \left[a_{i\alpha} + 2\varepsilon_{i\alpha\gamma}\hat{x}^{*(i)}_{\gamma}\right] \qquad (6.23)$$

On the right of (6.23) we may use

$$\hat{x}^{*(i)}_{\gamma} = x_{\gamma} + \left[\bar{N}(i)\right]^{-1}_{\gamma\xi}\sum_{k=1}^{i} n_k\, a_{k\xi} \quad . \qquad (6.24)$$

Again it is necessary to pre-multiply by the inverse of $N^{*(m)}$, given by:

$$\left[N^{*(m)}\right]^{-1}_{\alpha\beta} = \left[\bar{N}(m)\right]^{-1}_{\alpha\beta} - 2\left[\bar{N}(m)\right]^{-1}_{\alpha\gamma}\left[\bar{N}(m)\right]^{-1}_{\beta\delta}\sum_{i=1}^{m}\left(a_{i\gamma}\varepsilon_{i\delta\varepsilon} + a_{i\delta}\varepsilon_{i\gamma\varepsilon}\right)\hat{x}^{*(i)}_{\varepsilon} \quad . \qquad (6.25)$$

The distribution of $\hat{x}^{*(m)}_{\alpha} - x_{\alpha}$ is then available. The covariance

$$E\left\{\left(\hat{x}^{*(m)}_{\alpha} - x_{\alpha}\right)\left(\hat{x}^{*(m)}_{\beta} - x_{\beta}\right)\middle|\, x\right\}$$

is again the same as before. In computing the bias $E\left\{\hat{x}^{(m)}_{\alpha} - x_{\alpha}\right\}$, account must be taken of the correlation between n_i and $\hat{x}^{*(i)}_{\gamma}$. The result is:

$$E\left\{\hat{x}^{*(m)}_{\alpha} - x_{\alpha}\right\} = 3\left[\bar{N}^{(m)}\right]^{-1}_{\alpha\beta} \sum_{i=1}^{m} a_{i\beta}\,\varepsilon_{i\gamma\delta}\left[\bar{N}^{(i)}\right]^{-1}_{\gamma\delta}$$

$$- 2\left[\bar{N}^{(m)}\right]^{-1}_{\alpha\gamma}\left[\bar{N}^{(m)}\right]^{-1}_{\beta\delta} \sum_{i=1}^{m}\left(a_{i\gamma}\varepsilon_{i\beta\delta} + a_{i\delta}\varepsilon_{i\beta\gamma}\right) \qquad (6.26)$$

It is noteworthy that in the sequential estimates of Sections 6.3 and 6.4 the early non-linearities ε_i have much greater effect than the later ones, due to the presence of $\left[\bar{N}^{(i-1)}\right]^{-1}$ or $\left[\bar{N}^{(i)}\right]^{-1}$. The <u>iterated-sequential</u> fit, moreover, replaces $\bar{N}^{(i-1)}$ by $\bar{N}^{(i)}$, which by itself would be an improvement, but introduces also a multiplicative factor 3, which is adverse.

6.5 AN EXAMPLE

Suppose that we are making repeated measurements of a single quantity x with varying non-linear coefficients ε_i, and that we have an apriori estimate $\hat{x}_o$ of x with variance σ_o^2. Here we put $a_o = \frac{1}{\sigma_o}$, $\varepsilon_o = o$, $a_i = 1$ $(i \geqq 1)$. The least-square bias is

$$E\left\{\hat{x} - x\right\} = -\frac{1}{\left(m + \frac{1}{\sigma_o^2}\right)^2} \sum_{i=1}^{m} \varepsilon_i \qquad (6.27)$$

The sequential bias is

$$E\left\{\hat{x}^{(m)} - x\right\} = +\left(\frac{1}{m + \frac{1}{\sigma_o^2}}\right) \sum_{i=1}^{m}\left(\frac{\varepsilon_i}{i - 1 + \frac{1}{\sigma_o^2}}\right) - \frac{4\sum_{i=1}^{m} \varepsilon_i}{\left(m + \frac{1}{\sigma_o^2}\right)^2} \quad ; \qquad (6.$$

and the iterated-sequential bias is

$$E\left\{\hat{x}^{*(m)} - x\right\} = +\left(\frac{3}{m + \frac{1}{\sigma_o^2}}\right)\sum_{i=1}^{m}\left(\frac{\varepsilon_i}{i + \frac{1}{\sigma_o^2}}\right) - \frac{4\sum_{i=1}^{m}\varepsilon_i}{\left(m + \frac{1}{\sigma_o^2}\right)^2} \qquad (6.29)$$

Except for the first non-linearity, whose effect in $\hat{x}^{*(m)}$ is less than in $\hat{x}^{(m)}$, at least if σ_o is large (i.e. $> > 1$), the early non-linearities, which have the larger effect, are more serious in $\hat{x}^{*(m)}$ than in $\hat{x}^{(m)}$, whereas the most recent non-linearities affect $\hat{x}^{(m)}$ about three times as much as they affect either $\hat{x}$ or $\hat{x}^{*(m)}$. This rather paradoxical feature is not apparent in the simpler two-batch analysis in chapter 5 (see 1st paragraph on p.73). In the case of equal biases ($\varepsilon_i = \varepsilon$) and large m, we have:

$$E\left\{\hat{x} - x\right\} \sim -\frac{\varepsilon}{m} \quad , \quad E\left\{\hat{x}^{(m)} - x\right\} \sim \frac{\varepsilon}{m}\ln m \quad , \qquad (6.30)$$

and

$$E\left\{\hat{x}^{*(m)} - x\right\} \sim \frac{3\varepsilon}{m}\ln m \quad . \qquad (6.31)$$

It is tempting to conclude that a sequential fitting procedure should be iterated only at the outset. (!)

6.6 ANOTHER SEQUENTIAL PROCEDURE

An estimate equivalent (for negligible ε^2) to the least-square fit is available if third as well as second degree terms are carried forward in the expression representing the sum of the previous squares. This "higher-order sequential estimate" $\hat{x}^{\ddagger(i)}$ is defined as that x which minimizes

$$\rho_i^2 + Q_{i-1}(x) = N_{\alpha\beta}^{\dagger\ddagger(i-1)} \left(x_\alpha - \hat{x}_\alpha^{\dagger\ddagger(i-1)}\right)\left(x_\beta - \hat{x}_\beta^{\dagger\ddagger(i-1)}\right)$$

$$+ \frac{1}{3} E_{\alpha\beta\gamma}^{(i-1)} \left(x_\alpha - \hat{x}_\alpha^{\dagger\ddagger(i-1)}\right)\left(x_\beta - \hat{x}_\beta^{\dagger\ddagger(i-1)}\right)\left(x_\gamma - \hat{x}_\gamma^{\dagger\ddagger(i-1)}\right) + \rho_i^2 \quad . \tag{6.32}$$

Again, this is obtained by first linearizing ρ_i around $\hat{x}^{\dagger\ddagger(i-1)}$, expanding $Q_{i-1}(x)$ to 2nd order in $x - \hat{x}^{(i-1)}$, obtaining thus a first estimate $\hat{x}^{\dagger\ddagger(i)}{}_1$. $Q_{i-1}(x)$ and ρ_i are then re-expanded around $\hat{x}^{\dagger\ddagger(i)}{}_1$ to obtain $\hat{x}^{\dagger\ddagger(i)}{}_2$, etc., and a further iteration is optional. The next $N^{\dagger\ddagger}$ and E are:

$$N_{\alpha\beta}^{\dagger\ddagger(i)} = N_{\alpha\beta}^{\dagger\ddagger(i-1)} + E_{\alpha\beta\gamma}^{(i-1)} \left(\hat{x}_\gamma^{\dagger\ddagger(i)} - \hat{x}_\gamma^{\dagger\ddagger(i-1)}\right)$$

$$+ \left(a_{i\alpha} + 2\,\varepsilon_{i\alpha\gamma}\,\hat{x}_\gamma^{\dagger\ddagger(i)}\right)\left(a_{i\beta} + 2\,\varepsilon_{i\beta\delta}\,\hat{x}_\delta^{\dagger\ddagger(i)}\right) \quad , \tag{6.33}$$

and

$$E_{\alpha\beta\gamma}^{(i)} = E_{\alpha\beta\gamma}^{(i-1)} + 2\left(a_{i\alpha}\,\varepsilon_{i\beta\gamma} + a_{i\beta}\,\varepsilon_{i\alpha\gamma} + a_{i\gamma}\,\varepsilon_{i\alpha\beta}\right) \quad . \tag{6.34}$$

This improved sequential estimate required, of course, not only more storage but the calculation of all the second partial derivatives $\varepsilon_{i\alpha\beta}$.

6.7 AN APPLICATION

As a possible application of this analysis, we shall develop an algorithm for computing the effect of non-linearity in the estimate of position of an incoming vehicle (see Figure 6.1) based only on range measurements from a tracking station in the plane of motion.

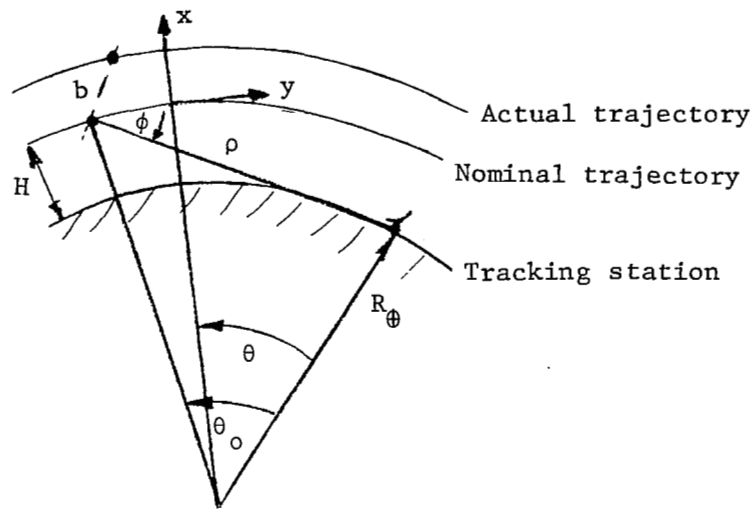

Figure 6.1 Tracking Geometry

Linearize the orbit description about the circular orbit through indicated first observed position, i.e.,

$$x_o = b \cos \phi \tag{6.35}$$

$$y_o = b \sin \phi \tag{6.36}$$

where (x,y) is a rectangular system moving on the nominal circular orbit, b is the magnitude of the initial position error, which may be assumed to be predominately perpendicular to the line of sight since the initial range measurement itself should be fairly accurate. Thus $\phi \cong \theta_o$, and the 3 unknowns b, $\dot{x}_o$, $\dot{y}_o$ may be assumed as zero mean with known variances and zero correlation. Define the departure from the nominal orbit, ignoring the initial range error, by means of

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} b \\ \frac{\dot{x}_o}{n} \\ \frac{\dot{y}_o}{n} \end{bmatrix}, \tag{6.37}$$

where n is mean motion on reference circular orbit.

Then:

$$\begin{pmatrix} x \\ y \end{pmatrix} = B(\theta) \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \tag{6.38}$$

where the 2 x 3 matrix B is

$$B(\theta) = \left(\begin{array}{c|c|c} \dfrac{\cos\phi\left(4 - 3\cos(\theta_o-\theta)\right)}{\sin\phi + 6\cos\phi\left(\sin(\theta_o-\theta)-\theta_o+\theta\right)} & \dfrac{\sin(\theta_o-\theta)}{-2\left(1-\cos(\theta_o-\theta)\right)} & \dfrac{2\left(1-\cos(\theta_o-\theta)\right)}{4\sin(\theta_o-\theta)-3\theta_o+3\theta} \end{array} \right) \tag{6.39}$$

The Measured range is

$$\rho = \sqrt{(R_\oplus + H + x - R_\oplus \cos\theta)^2 + (R_\oplus \sin\theta - y)^2} + \text{random error} \tag{6.40}$$

where the random error has variance σ_ρ^2, and

$$\rho_x = \frac{R_\oplus + H + x - R_\oplus \cos\theta}{\rho} \tag{6.41}$$

$$\rho_y = \frac{-R_\oplus \sin\theta + y}{\rho} \tag{6.42}$$

$$\rho_{xx} = \frac{(R_\oplus \sin\theta - y)^2}{\rho^3} \tag{6.43}$$

$$\rho_{yy} = \frac{(R_\oplus + H + x - R_\oplus \cos\theta)^2}{\rho^3} \tag{6.44}$$

$$\rho_{xy} = \frac{(R_\oplus \sin\theta - y)(R_\oplus + M + x - R_\oplus \cos\theta)}{\rho^3} \tag{6.45}$$

We assume that the non-linearities in expression for ρ in terms of λ_i arise predominantly from non-linearities in ρ as a function of (x,y) rather than in (x,y) as functions of λ_i. [These latter non-linearities have already been ignored in writing equation (6.38)]

Then

$$\rho \simeq \underset{[x=y=0]}{\rho^{(o)}(\theta)} + \underbrace{\rho'(\theta)B(\theta)}_{1\times 3}\lambda + \frac{1}{2}\lambda^T \underbrace{B^T(\theta)\rho''(\theta)B(\theta)}_{3\times 3}\lambda \tag{6.46}$$

where

$$\lambda = \begin{pmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\end{pmatrix}, \qquad \rho'(\theta) = \underset{x=y=0}{[\rho_x,\ \rho_y]}, \qquad \rho''(\theta) = \underset{x=y=0}{\begin{bmatrix}\rho_{xx} & \rho_{xy}\\ \rho_{xy} & \rho_{yy}\end{bmatrix}} \tag{6.47}$$

The previous analysis now applies (λ here is x there), where

$$\underset{\alpha = 1,2,3}{||a_{(i)\alpha}||} = \frac{1}{\sigma_\rho}\underbrace{\rho'(\theta_i)\ B(\theta_i)}_{1\ \times\ 3} \tag{6.48}$$

$$\underset{\alpha,\beta = 1,2,3}{||\varepsilon_{(i)\alpha\beta}||} = \frac{1}{2\sigma_\rho} \underbrace{B^T(\theta_i)\ \rho''(\theta_i)\ B(\theta_i)}_{3 \times 3} \tag{6.49}$$

We may suppose that the 1st, 2nd, ... observations are counted by i = 4, 5, ... and reserve i = 1, 2, 3 for apriori variances:

$$\alpha_{1\alpha} = \left(\frac{1}{\sigma_b}\ ,\ 0,\ 0\right) \tag{6.50}$$

$$\alpha_{2\alpha} = \left(0,\ \frac{n}{\sigma_{\dot{x}_o}}\ ,\ 0\right) \qquad \text{and} \quad \varepsilon_{i\alpha\beta} = 0 \text{ for } i=1,2,3. \tag{6.51}$$

(n: mean motion)

$$\alpha_{3\alpha} = \left(0,\ 0,\ \frac{n}{\sigma_{\dot{y}_o}}\right) \tag{6.52}$$

We can now proceed to compute the biases in $\begin{pmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\end{pmatrix}$, say $\begin{pmatrix}\Delta\lambda_1\\ \Delta\lambda_2\\ \Delta\lambda_3\end{pmatrix}$,

(i) according to (6.5), for the least-squares fit,
(ii) according to (6.18) and (6.26) for the sequential fits.

Finally the biases in extrapolated in-plane position, at say $\theta=\theta^*$, are given by:

$$\begin{pmatrix}\Delta x\\ \\ \Delta y\end{pmatrix} = B(\theta^*) \begin{pmatrix}\Delta\lambda_1\\ \Delta\lambda_2\\ \Delta\lambda_3\end{pmatrix} \tag{6.53}$$

7. CONSIDERATION OF SYSTEMATIC ERRORS

7.1 INTRODUCTION

As a practical matter it is usually not feasible to estimate all of the orbit parameters if the dimension of the state vector is high. For example, it is obviously impossible to estimate all of the coefficients of the spherical harmonics describing the gravitational potential of the central body, even though any or all of them might have a significant effect upon the solution of the orbit determination problem. Ideally, one would like to apply the method described in Reference [7] to determine those linear combinations of parameters which affect the data weakly so that they can be deleted from consideration. In practice, this is done on some intuitive basis, and perhaps some significant parameters are left out. These unestimated parameters are called systematic errors.

In the presence of systematic errors one sometimes resorts to a "consider option," which can take on several forms: (1) weighted least squares - estimate the desired parameters as if the systematic errors were not present, where the inverse data noise covariance is the weighting matrix, but reflect their contribution to the estimation error in the calculation of the estimation error covariance matrix. (2) minimum variance - reflect the presence of the systematic errors in the estimate of the desired orbit parameters and in the calculation of the estimation error covariance matrix, in such a way as to obtain the minimum error in the estimated parameters. This approach is equivalent to estimating all the orbit parameters, including the systematic errors. (3) general form - choose an arbitrary form of the estimator which is computationally convenient and produces an acceptably small error covariance matrix in the presence of systematic errors. The general form of course includes (1) and (2) as special cases.

It is the purpose of this Section to develop a collection of methods for obtaining estimation error covariance matrices when the general form of the consider option is employed. A linearized relation between the data and state will be assumed, which, from the point of view of the nonlinear theory of Section 3, can be thought of as a linear expansion about

the converged modal trajectory. It will be further assumed that no state noise is present, but this case will be treated in Section 4.

7.2 THE CONSIDER OPTION FOR A SINGLE DATA BATCH

Suppose the vector z of observation residuals is given by

$$z = Ax + Bs + n \tag{7.1}$$

where x, s, and n have the usual meanings, i.e. x is the vector to be estimated, s is the vector of system parameter errors, and n is the random vector of observation noise. It will be assumed in Subsections 7.2 and 7.3 that the apriori variance of x is infinite ($\Lambda^{-1} = 0$). A linear estimator L converts z into an estimate $\hat{x}$ of x : $\hat{x} = Lz$. In view of (7.1) this relation becomes

$$\hat{x} = LAx + LBs + Ln \tag{7.2}$$

The estimator L will be called unbiased if $LA = I$. This property defines the general form of the consider option, and will be assumed throughout. Hence

$$\hat{x} - x = LBs + Ln$$

$$= Cs + Dn \tag{7.3}$$

There are two interpretations of the vector s. It may be regarded as a constant but unknown vector. We know only that the best estimate $\hat{s}$ of s is $\hat{s} = 0$, and that our confidence in this estimate is given by a covariance matrix Σ_s. We may then define $E(s) = \hat{s} = 0$, and $E(ss^T) = \Sigma_s$. With these conventions s may be handled mathematically as if it were a random vector with mean 0 and covariance Σ_s. This is the second interpretation. In either case, (7.3) implies

$$\Sigma = \text{cov}(\hat{x} - x) = D\Sigma_n D^T + C\Sigma_s C^T \tag{7.4}$$

where $\Gamma = E(nn^T)$ and it is assumed that $E(sn^T) = 0$.

Two well known estimators are the following:

Weighted Least Squares Estimator This tried and true estimator is given by

$$L_{LS} = (A^T\Sigma_n^{-1}A)^{-1}A^T\Sigma_n^{-1} \quad .$$

The corresponding estimate will be denoted by $\hat{x}_{LS}$; the covariance matrix (7.4) by Σ_{LS}.

Minimum Variance Estimator This estimator provides the minimum covariance matrix (4.4). It is given by

$$L_{MV} = (A^TWA)^{-1}A^TW \quad ,$$

where $W^{-1} = \Gamma + B\Sigma_s B^T$. This is easily shown: Any other estimator may be expressed as $L = K + (A^TWA)^{-1}A^TW$. Then $LA = I$ implies $KA = 0$, and it follows that (7.4) reduces to

$$KW^{-1}K^T + (A^TWA)^{-1} \quad .$$

This is a minimum for $K = 0$. It follows that $\Sigma_{MV} = (A^TWA)^{-1}$.

It is usually impractical to use L_{MV} in making estimates of x alone, since the calculations required are equivalent to those required to estimate s in addition to x. The matrix Σ_{MV} is calculated by using a weighted least squares estimator to estimate the vector $\binom{x}{s}$ from the data

$$\begin{pmatrix} z \\ z_s \end{pmatrix} = \begin{pmatrix} A & B \\ 0 & I \end{pmatrix} \begin{pmatrix} x \\ s \end{pmatrix} + \begin{pmatrix} n \\ n_s \end{pmatrix},$$

where $E(n_s n_s^T) = \Sigma_s$. The estimator in this case is

$$L_{LS} = \left[\begin{pmatrix} A & B \\ 0 & I \end{pmatrix}^T \begin{pmatrix} \Gamma^{-1} & 0 \\ 0 & \Sigma_s^{-1} \end{pmatrix} \begin{pmatrix} A & B \\ 0 & I \end{pmatrix} \right]^{-1} \begin{pmatrix} A & B \\ 0 & I \end{pmatrix}^T \begin{pmatrix} \Gamma^{-1} & 0 \\ 0 & \Sigma_s^{-1} \end{pmatrix}$$

The resulting covariance matrix (7.4) is

$$\begin{pmatrix} A^T\Gamma^{-1}A & A^T\Gamma^{-1}B \\ B^T\Gamma^{-1}A & B^T\Gamma^{-1}B + \Sigma_s^{-1} \end{pmatrix}^{-1} .$$

The upper left portion of this matrix, which is just $\mathrm{cov}(\hat{x} - x)$, is equal to Σ_{MV} as may be shown using the Schur identity.

The main interest in Σ_{MV} is that it gives a lower bound for the obtainable tracking accuracy.

7.3 THE CONSIDER OPTION FOR TWO DATA BATCHES

Suppose two tracking runs result in observations

$$z_1 = A_1x + B_1s + n_1$$

$$z_2 = A_2x + B_2s + n_2 \quad . \tag{7.5}$$

Suppose linear estimators L_1 and L_2 are applied to give estimates

$$\hat{x}_1 = L_1A_1x + L_1B_1s + L_1n_1$$

$$= x + C_1s + D_1n_1$$

$$\hat{x}_2 = L_2A_2x + L_2B_2s + L_2n_2$$

$$= x + C_2s + D_2n_2 \quad .$$

It is desired to combine these estimates to give an estimate $\hat{x}_3$ which is better than either and to compute the covariance of the error $\hat{x}_3 - x$. Assuming a linear unbiased combination, all schemes may be written as

$$\hat{x}_3 = M\hat{x}_1 + N\hat{x}_2 \quad , \tag{7.6}$$

where $M + N = I$.

Thus,

$$\hat{x}_3 = x + (MC_1 + NC_2)s + (MD_1, ND_2)\begin{pmatrix} n_1 \\ n_2 \end{pmatrix} \tag{7.7}$$

and

$$\Sigma_3 = \operatorname{cov}(\hat{x}_3 - x) = MD_1 \Gamma_1 D_1^T M^T + ND_2 \Gamma_2 D_2^T N^T$$

$$+ (MC_1 + NC_2)\, \Sigma_s (MC_1 + NC_2)^T$$

$$= (M,N)\begin{pmatrix} \Sigma_1 & \theta \\ \theta^T & \Sigma_2 \end{pmatrix}\begin{pmatrix} M^T \\ N^T \end{pmatrix} \tag{7.8}$$

where

$\Gamma_i = E(n_i n_i^T)$, $\Sigma_i = \Sigma_{in} + \Sigma_{ic} = D_i \Gamma_i D_i^T + C_i \Sigma_s C_i^T$, $i = 1, 2$, and $\theta = C_1 \Sigma_s C_2^T$.
These definitions will be used frequently in what follows.

The foregoing is quite straight forward but it emphasizes the fact that, once the matrix M is selected (and hence also $N = I - M$), the resulting covariance matrix (7.8) is determined and needs merely to be calculated. Looking at the matter differently, whenever a matrix Σ_3 is purported to be the covariance matrix of a combined estimate, there must be a corresponding choice of M related to Σ_3 by (7.8). Any deviation from this rule may be regarded with suspicion unless accompanied with an estimate of the difference between Σ_3 and a covariance matrix calculated from (7.8).

Two combinative procedures will be discussed.

Weighted Least Squares Combination Using the notation of (7.5), this method consists of letting

$$M_{LS} = (A_1^T W_1 A_1 + A_2^T W_2 A_2)^{-1} (A_1^T W_1 A_1)$$

$$N_{LS} = (A_1^T W_1 A_1 + A_2^T W_2 A_2)^{-1} (A_2^T W_2 A_2)$$

where $W_1 = \Gamma_1^{-1}$, $W_2 = \Gamma_2^{-1}$ and it is assumed that n_1 and n_2 are independent.

If $\hat{x}_1$ and $\hat{x}_2$ were obtained using the weighted least squares estimator described in the previous section, then this choice of M and N results in a simple combination of data batches 1 and 2 into a larger batch which is then processed with a weighted least squares estimator. In other words, the tracking matrices $[A_i^T W_i A_i \mid A_i^T W_i B_i]$, $i = 1, 2$ are combined to form the tracking matrix

$$[A_1^T W_1 A_1 + A_2^T W_2 A_2 \mid A_1^T W_1 B_1 + A_2^T W_2 B_2] \quad .$$

This is no longer valid if $\hat{x}_1$ and $\hat{x}_2$ were formed in any other way.

Minimum Variance Combination If the estimators L_1 and L_2 are fixed, then how can M be chosen to obtain the minimum covariance, Σ_3, of $\hat{x}_3 - x$? Assuming $E(n_1 n_2^T) = 0$, the answer is given by

$$M_{MV} = \left[(\Sigma_1 - \theta)^{-1} + (\Sigma_2 - \theta^T)^{-1} \right]^{-1} (\Sigma_1 - \theta)^{-1}$$

$$N_{MV} = \left[(\Sigma_1 - \theta)^{-1} + (\Sigma_2 - \theta^T)^{-1} \right]^{-1} (\Sigma_2 - \theta^T)^{-1} \quad .$$

where Σ_1, Σ_2, and θ are as given by (7.8). This is easily shown by replacing N by I - M in (7.8), expanding in M and completing the square. The resulting covariance is

$$\Sigma_{MV} = (M_{MV}, N_{MV}) \begin{pmatrix} \Sigma_1 & \theta \\ \theta^T & \Sigma_2 \end{pmatrix} \begin{pmatrix} M_{MV}^T \\ N_{MV}^T \end{pmatrix}$$

$$= \Sigma_2 - M_{MV}(\Sigma_1 + \Sigma_2 - (\theta + \theta^T))^{-1} M_{MV}$$

Any other choice of M results in a covariance matrix Σ_3 where

$$\Sigma_3 - \Sigma_{MV} = (M - M_{MV})(\Sigma_1 + \Sigma_2 - (\theta + \theta^T))(M - M_{MV})^T.$$

Although the minimum variance estimator L_{MV} of Section 1 is difficult to calculate, the choice $M = M_{MV}$ is easily calculated if the matrices C_i, D_i, Σ_s and Γ_i are available. These in turn depend upon the estimators L_1 and L_2, which may well be taken as least squares estimators. In this case the C and D matrices may be readily computed.

7.4 THE INFLUENCE OF APRIORI INFORMATION

Suppose the estimate $\hat{x}_1$ is available along with a covariance matrix Σ_1, the "a priori" covariance, but that the dependence of Σ_1 on Σ_s and Γ_1 is not known. We may assume that

$$\hat{x}_1 = x + C_1 s + D_1 n_1 \quad ,$$

but the matrices C_1 and D_1 are not known. They may be regarded as variable but constrained so that $C_1 \Sigma_s C_1^T + D_1 \Gamma_1 D_1^T = \Sigma_1$.

It is often stated that the covariance of the combined estimate is given by $\tilde{\Sigma}_3$ which is computed from the augmented tracking matrix

$$[A_2^T W_2 A_2 + \Sigma_1^{-1} \;\vdots\; A_2^T W_2 B_2]$$

in the same manner employed for a weighted least squares estimator. This is a dependable covariance matrix corresponding to the combinative procedure $M = (A_2^T W_2 A_2 + \Sigma_1^{-1})^{-1} \Sigma_1^{-1}$ <u>if</u> $C_1 = 0$. If $C_1 \neq 0$, the matrix $\tilde{\Sigma}_3$ is not correctly calculated. It is the purpose here to investigate the magnitude of the error.

Consider the combinative method given by

$$M = (A_2^T W_2 A_2 + \Sigma_1^{-1})^{-1} \Sigma_1^{-1} \quad .$$

Let Σ_3 be the corresponding covariance matrix computed from (7.8).

The following formulas are easily established:

$$\Sigma_3 = (M,N)\begin{pmatrix}\Sigma_1 & \theta \\ \theta^T & \Sigma_2\end{pmatrix}\begin{pmatrix}M^T \\ N^T\end{pmatrix}$$

$$\tilde{\Sigma}_3 = (M,N)\begin{pmatrix}\Sigma_1 & 0 \\ 0 & \Sigma_2\end{pmatrix}\begin{pmatrix}M^T \\ N^T\end{pmatrix}$$

$$\Sigma_{MV} = (M_{MV},N_{MV})\begin{pmatrix}\Sigma_1 & \theta \\ \theta^T & \Sigma_2\end{pmatrix}\begin{pmatrix}M_{MV}^T \\ N_{MV}^T\end{pmatrix} \qquad (7.9)$$

hence

$$\Sigma_3 - \tilde{\Sigma}_3 = M\theta N^T + N\theta^T M^T.$$

It is seen that this error will be small compared to the size of θ if $M \approx 0$ or $N \approx I$, i.e. if the noise errors are much smaller than the system parametric errors. But θ is about the same size as the matrices Σ_{1C} and Σ_{2C}, hence, in this case $\tilde{\Sigma}_3$ is a reasonable approximation of Σ_3.

Since the pseudo covariance matrix $\tilde{\Sigma}_3$ could be unrealistically small, it would be useful to know under what circumstances it is at least greater than the minimum variance covariance matrix Σ_{MV}. Equation (7.9) shows that

$$\begin{aligned}\tilde{\Sigma}_3 - \Sigma_{MV} &= M\Sigma_1 M^T + N\Sigma_2 N^T - M_{MV}\Sigma_1 M_{MV}^T \\ &\quad - M_{MV}\theta N_{MV}^T - N_{MV}\theta^T M_{MV} + N_{MV}\Sigma_2 N_{MV}^T \\ &= N_{MV} Q N_{MV}^T + N(\Sigma_{2C} - (\Sigma_1 + \Sigma_{2n}))N^T,\end{aligned}$$

where $Q = (\Sigma_1 - \theta) + (\Sigma_2 - \theta^T)$.

It is clear that $\tilde{\Sigma}_3 > \Sigma_{MV}$ if $\Sigma_{2C} \geq \Sigma_1 + \Sigma_{2n}$. Otherwise the situation is not so clear.

A possible approach to establishing that $\tilde{\Sigma}_3 \geq \Sigma_{MV}$ would be to show the existence of matrices $\tilde{M}$ and $\tilde{N} = I - \tilde{M}$ such that $\tilde{\Sigma}_3$ may be computed from (7.8) using $\tilde{M}$ and $\tilde{N}$. However it can be shown that a necessary and sufficient condition for the existence of such matrices is $\tilde{\Sigma}_3 \geq \Sigma_{MV}$. This approach then is not helpful.

7.5 A SIMPLIFIED ANALYSIS OF STEADY STATE ORBIT DETERMINATION

Suppose we have a steady state (constant geometry) orbit determination situation where the state of the system can be estimated quite well during a short tracking interval, but where the effects of state noise and unmodelled parameters add disturbance between tracking intervals. In order to optimally weight the data and to describe the tracking error, it is necessary to determine a realistic error covariance matrix which is characteristic of the operational tracking system.

Assuming one continuously estimates some limited number of system parameters, ignoring state noise and unmodelled parameters, the hypothetical error variance $(A^T W A)^{-1}$ would go to zero as time goes to infinity, (Figure 7.1 - curve 1). Consideration of these unmodelled effects in the calculation would yield an error variance (7.4) which becomes unacceptably large with time, (Figure 7.1 - curve 2). Neither of these calculations is meaningful, however, for in an operational system one would de-weight early data to reflect the information loss due to unmodelled effects, (Figure 7.1 - curve 3). For example, previous information might be completely ignored by using only the most recent tracking interval to estimate the state. Such an approach is a special case of the data weighting methods discussed above, where the covariance matrix describing prior information is degraded in some less arbitrary fashion.

Theoretically, de-weighting prior information corresponds to postulating systematic errors acting between tracking intervals. Thus the most rational way to treat the problem is to determine the statistics of this disturbance, and weight the data according to the true error covariance. A broken curve will result since the tracking is intermittent (Figure 7.1 - curve 4). The necessary calculations can be carried out with existing

computer programs, for it is only necessary to compute the normal matrices corresponding to the individual tracking intervals and add impulses to the error covariance matrices between intervals. In this subsection an approximate treatment of the steady state orbit determination problem will be presented. Such an analysis hopefully will be useful for illustrating the type of results to be expected, for approximate error analysis, and for suggesting a realistic operational data weighting scheme and tracking policy.

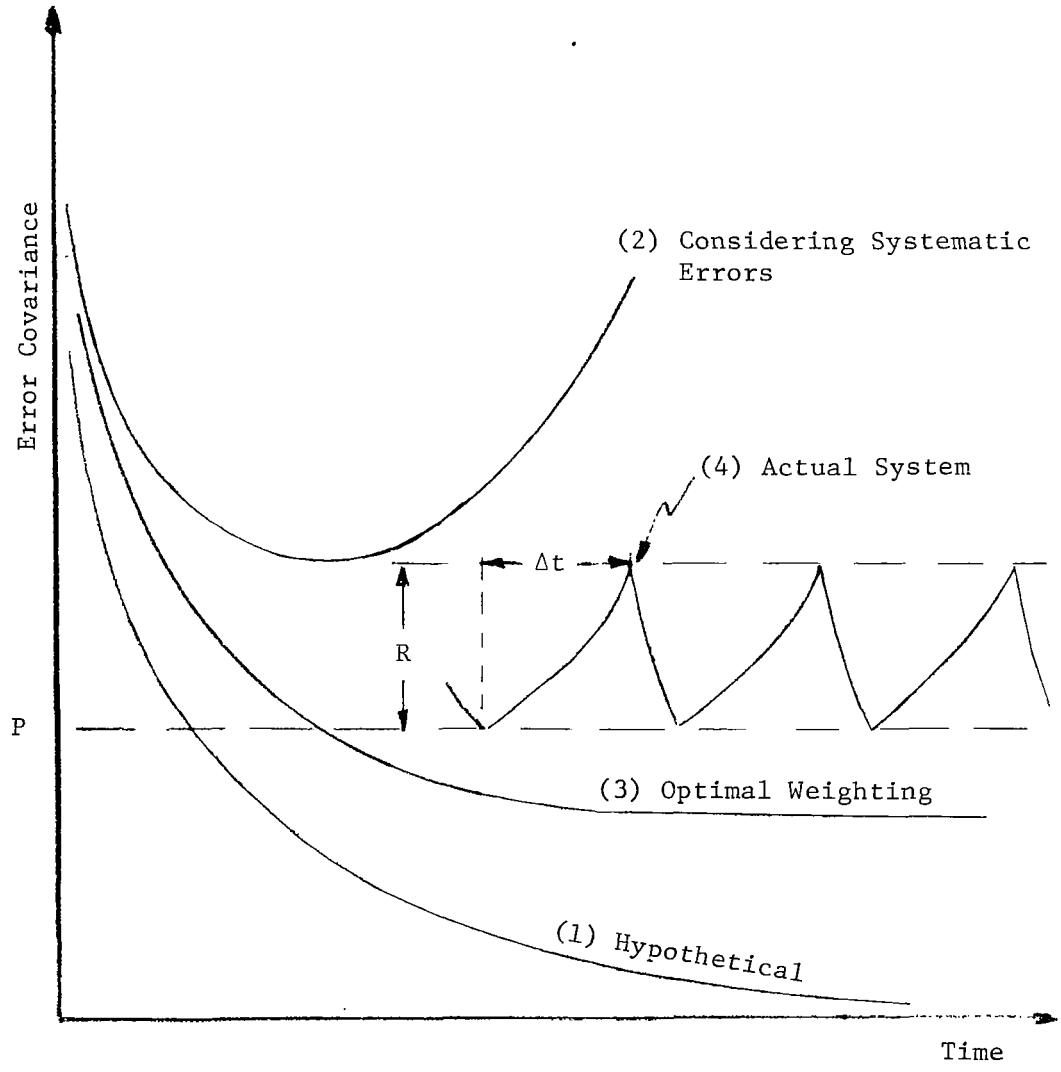

Figure 7.1 Tracking Error Covariance

Suppose the unmodelled parameters and state noise act only between tracking intervals to perturb the state by an amount Δx. Then let

$$x_i = \text{constant} \tag{7.10}$$

$$x_{i+1} = x_i + \Delta x \tag{7.11}$$

$$z_i = A\, x_i + n_i \tag{7.12}$$

where x is the state vector, i denotes the ith tracking interval, Δx is the state disturbance, n_i is white data noise, z_i is data. The A is a (constant) matrix, which is supposed to be of full rank. Let the variance of Δx be

$$E[\Delta x^2] = R(\Delta t) \tag{7.13}$$

where Δt is the time between tracking intervals. Thus,

$$R = \int_0^{\Delta t} \int_0^{\Delta t} \left[\frac{\partial x}{\partial q}\right] \left[Q(\tau,s)\right] \left[\frac{\partial x}{\partial q}\right]^T d\tau\, ds \tag{7.14}$$

where q is the perturbation acting in Δt and $Q(\tau,s)$ is its autocorrelation function. Let P be the "steady state" covariance matrix, obtained at the end of a tracking interval (Figure 7.1). The uncertainty at the beginning of a tracking interval is $P + \Delta R$, and the information added by a tracking interval is $N = (A^T W A)$, where W is the inverse data noise variance. The P matrix is obtained from

$$\left[\left(P + R\right)^{-1} + N\right]^{-1} = P \tag{7.15}$$

The solution of (7.15) is

$$P = \left[\left(I + 4\, N^{-1}\, R^{-1}\right)^{1/2} - I\right] \frac{R}{2} \tag{7.16}$$

If the norm of $\left(N^{-1}R^{-1}\right)$ is small, we have

$$P = \left[N^{-1} - N^{-1}\, R^{-1}\, N^{-1} + \ldots.\right] \cong N^{-1} - N^{-1}\, R^{-1}\, N^{-1} \tag{7.17}$$

The estimate at the end of a tracking interval is

$$\left(x_i^*\right)^+ = P\left[A^T W z_i + \left(P + R\right)^{-1}\left(x_i^*\right)^-\right] \tag{7.18}$$

where [+,-] refers, respectively, to the end and beginning of the interval. Thus (7.17) describes the desired steady state error covariance matrix, and (7.18) shows how to optimally combine the most recent data with prior information.† These results can be easily extended to the case where the normal matrix N depends upon i, or where the unmodelled parameters and state noise act during the tracking interval.

The analysis developed here could be used in the following way:

a. define the disturbance covariance R(Δt) from a (perhaps pessimistic) analysis of the state noise and unmodelled parameters.

b. define an acceptable value of R(Δt) and determine the time between tracking intervals to achieve that value.

c. given the data z in any tracking interval, find the estimate at the end of the tracking interval by (7.18).

Note that only the P matrix is needed for error analysis purposes and for definition of the tracking pattern. The validity of this approach could be checked by a Monte Carlo simulation with equation (7.18) introduced as the "fit world" estimator.

†From (7.17) and (7.18) it can be seen that prior information is ignored only if R = infinity, in which case $P = N^{-1}$ and $(P+R)^{-1} = 0$.

8. TREATMENT OF CORRELATED DATA

8.1 INTRODUCTION

The weighted least squares (WLS) form of the maximum likelihood estimator can treat correlated data by employing the weighting matrix $W = \Gamma^{-1}$, where Γ is the data noise covariance matrix $\Gamma = E[nn^T]$. This is a large matrix, with dimension equal to the total number of data points, and in general it is not practical to invert and store Γ.

In certain special cases the inverse is easily obtained. For example, if the data is uniformly spaced Δt seconds apart and is exponentially correlated, we have

$$E[n(t_j)n(t_i)] = \sigma^2 \exp(-\alpha|t_j - t_i|) = \sigma^2 \gamma^{|j-i|} \tag{8.1}$$

where σ^2 is the (stationary) variance, $\frac{1}{\alpha}$ is the correlation time constant, and,

$$\gamma = \exp(-\alpha \Delta t)$$

Then

$$\Gamma = \sigma^2 \begin{bmatrix} 1 & \gamma & \gamma^2 & \gamma^3 & . & . & . \\ \gamma & 1 & \gamma & \gamma^2 & . & . & . \\ \gamma^2 & \gamma & 1 & \gamma & . & . & . \\ \gamma^3 & \gamma^2 & \gamma & 1 & . & . & \gamma \\ . & . & . & . & . & \gamma & 1 \end{bmatrix} \tag{8.2}$$

and

$$\Gamma^{-1} = W = \frac{1}{\sigma^2(1-\gamma^2)} \begin{bmatrix} 1 & -\gamma & 0 & 0 & . & . & . \\ -\gamma & (\gamma^2+1) & -\gamma & 0 & . & . & . \\ 0 & -\gamma & (\gamma^2+1) & -\gamma & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & 0 & -\gamma & (\gamma^2+1) & -\gamma \\ . & . & . & . & 0 & -\gamma & 1 \end{bmatrix} \tag{8.3}$$

This weighting matrix causes adjacent data points to be differenced, as can be seen by performing the multiplication Wz, where z is the data vector. This approach is equivalent to adding a new state variable which is the correlated noise, as is sometimes done in Kalman filtering. Essentially, the inverse (8.3) can be easily found because (8.1) describes a first order Markoff process.

The treatment of more general forms of data correlation is not so straightforward. In this Section we will develop a practical method of processing exponential-cosine correlated data by a differencing method, and it will be shown how a general data differencing technique can be used to treat other types of correlation. It will also be shown how to transform the problem to that of estimating an unknown acceleration. As before, a linear system with no state noise will be assumed, which corresponds to a linear expansion about the modal trajectory.

8.2 STATEMENT OF THE PROBLEM

Consider the problem of estimating the initial state vector of an orbiting vehicle from a linear combination of noisy measurements. Let the equation relating the measurement vector z, to the initial state vector x_o, be

$$z = Ax_o + n \tag{8.4}$$

where $A = \partial z/\partial x_o$ and n is the vector of zero-mean random noise on the measurement. Then, the well known weighted least squares estimate of x_o is

$$\hat{x}_o = (A^T W A)^{-1} A^T W z \tag{8.5}$$

and the covariance of the estimate is

$$\Sigma_{WLS} = (A^T W A)^{-1} A^T W \Gamma W A (A^T W A)^{-1} \tag{8.6}$$

where W is a diagonal weighting matrix and $\Gamma = E[nn^T]$ is the covariance of the noise. If $W^{-1} = \Gamma$, then the estimate is a minimum variance one and (8.6) becomes

$$\Sigma_{MV} = (A^T W A)^{-1} = (A^T \Gamma^{-1} A)^{-1} \tag{8.7}$$

When Γ is not a diagonal matrix, i.e., the data are correlated, then the minimum variance problem can be transformed to the WLS problem if a matrix $S = [s_{ij}]$ can be found such that

$$S^T W S = \Gamma^{-1}$$

In this case, the estimate of x_o becomes

$$\hat{x}_o = (B^T W B)^{-1} B^T W q \tag{8.8}$$

where $B = SA$ and $q = Sz$. Because most orbit determination programs do not save more than one row of the A-matrix at a time in solving the WLS problem, transforming the minimum variance problem to the WLS problem as indicated above is only feasible if a small number of rows of the A matrix must be saved. That is, only a small number of elements of each row of S are non-zero. Section 8.1 has shown how to transform exponentially correlated data into uncorrelated data by this data differencing technique and the following sections will show how to treat data with an exponential cosine correlation by data differencing. The inverse problem of determining what type of correlation can be handled by differencing a given number of measurements will also be examined.

An alternative to the above approach is to augment the state vector with noise parameters which are perturbed by white noise, and solve for this new state vector in the presence of an "unknown acceleration," (see Section 4). This approach, which is analogous to the way correlated measurement noise is handled in the Kalman filter, is discussed in 8.5.

8.3 OPTIMAL PROCESSING OF DATA WITH EXPONENTIAL COSINE CORRELATION

8.3.1 The Noise Model

Let the autocorrelation function for the measurement noise have the form

$$R_{ij} = E[n_i n_j]$$

$$R_{ij} = (\cos \beta\tau_{ij} + \gamma \sin \beta\tau_{ij}) \exp(-\alpha\tau_{ij}) \tag{8.9}$$

where $\tau_{ij} = |t_i - t_j| = \tau_{ji}$, $n_i = n(t_i)$ = the ith component of the vector n, and α, β, and γ are constants. This is a realistic noise model since, for $\gamma = \alpha/\beta$, (8.9) represents the autocorrelation function of the output of a linear, second-order system forced by white noise. The equation of motion for such a system is

$$\ddot{n} + 2\alpha \dot{n} + (\alpha^2 + \beta^2)n = w(t) \tag{8.10}$$

where w(t) is zero-mean, gaussian white noise with variance σ_w^2. The autocorrelation function of the steady-state response of this system is given by Eqn. (8.9) if $\sigma_w^2 = 4\alpha(\alpha^2 + \beta^2)$. This will also be true for the transient response if the statistics on the initial conditions are

$$E(n_o) = E(\dot{n}_o) = 0$$

$$E(n_o^2) = 1$$

$$E(\dot{n}_o^2) = \alpha^2 + \beta^2$$

$$E(n_o \dot{n}_o) = 0$$

8.3.2 Solution For Equally Spaced Data

For equally spaced data, the solution of Eqn. (8.10) obeys the difference equation

$$n_{i+2} - c_1 n_{i+1} + c_o n_i = u_i \tag{8.11}$$

where

$$\tau = t_i - t_{i-1} ,$$

$$c_o = e^{-2\alpha\tau} ,$$

$$c_1 = 2e^{-\alpha\tau} \cos \beta\tau ,$$

$$u_i = \frac{e^{-2\alpha\tau}}{\beta} \left[\int_{t_{i+\tau}}^{t_{i+2\tau}} w(s) \sin \beta(t_i + 2\tau - s) e^{-\alpha(t_i - s)} ds \right.$$

$$\left. - \int_{t_i}^{t_{i+\tau}} w(s) \sin \beta(t_i - s) e^{-\alpha(t_i - s)} ds \right] \tag{8.12}$$

The autocorrelation function of u is then

$$E[u_i u_j] = \begin{cases} 0 & \text{if } |i-j| > 1 \\ \rho\sigma_o^2 & \text{if } |i-j| = 1 \\ \sigma_o^2 & \text{if } i-j = 0 \end{cases} \tag{8.13}$$

where $\sigma_o^2 = 1 - e^{-4\alpha\tau} - 4 \frac{\alpha}{\beta} e^{-2\alpha\tau} \sin \beta\tau \cos \beta\tau$

$$= 2e^{-2\alpha\tau} \left[\sinh 2\alpha\tau - \frac{\alpha}{\beta} \sin 2\beta\tau \right] , \tag{8.14}$$

$$\rho = 2\left[\frac{\alpha}{\beta}\sin\beta\tau\cosh\alpha\tau - \cos\beta\tau\sinh\alpha\tau\right]\Big/ e^{2\alpha\tau}\sigma_o^2 \tag{8.15}$$

The new variable u_i is only correlated with u_{i-1} and u_{i+1}, its nearest neighbors. An uncorrelated set of variables with variance σ_o^2 can now be formed from the $\{u_i\}$ by letting

$$v_i = (u_i - a_i v_{i-1})\Big/\sqrt{1 - a_i^2} \tag{8.16}$$

$$a_i = \rho\Big/\sqrt{1 - a_{i-1}^2} \quad , \quad i > 1 \tag{8.17}$$

where $v_1 = u_1$,

$a_1 = 0$.

For N large, then the sequence of a_i defined in (8.17) converges to the value

$$a^\infty = \left[\frac{1}{2} - \frac{1}{2}\left(1 - 4\rho^2\right)^{1/2}\right]^{1/2} \cdot \text{sign}(\rho) \quad , \; N >> 1 \tag{8.18}$$

Notice that a real solution of (8.18) only exists for $|\rho| \leq \frac{1}{2}$. However, from (8.15) this can be shown to hold for all positive values of α, β, and τ.

Combining the above results with Eqn. (8.4), denoting the ith row of the A matrix as A_i, yields

$$q_i = B_i\, x_o + v_i \quad , \quad i = 1, 2, \ldots, N - 2$$

where

$$q_i = (Lz_i - a_i q_{i-1})\Big/\sqrt{1 - a_i^2} \quad ,$$

$$B_i = (LA_i - a_i B_{i-1})\Big/\sqrt{1 - a_i^2} \quad , \quad i > 1 \; ,$$

$$B_1 = LA_1 \quad .$$

L is a linear difference operator defined by

$$Lz_i = z_{i+2} - c_1 z_{i+1} + c_o z_i$$

Thus, letting $B^T = [B_1^T, B_2^T, \ldots, B_{N-2}^T]$, $q^T = [q_1^T, q_2^T, \ldots, q_{N-2}^T]^T$ and $W^{-1} = \sigma_o^2 I$, Eqn. (8.8) can be used to obtain a WLS estimate from the new data vector q. Note that the calculation of $B^T WB$ and $B^T Wq$ by summing the individual elements $B_i^T W_i B_i$ and $B_i^T W_i q_i$ requires the temporary storage of A_i, A_{i+1}, A_{i+2}, z_i, z_{i+1}, z_{i+2}, A_1, B_{i-1}, and q_{i-1}, in addition to the constants c_1, and c_2. Also, note that there are only (N-2) of the new uncorrelated observations as compared to the N original observations; no practical way of forming N uncorrelated observations was found for data with exponential cosine correlation. In this case, the covariance of the estimate is

$$\Sigma = (B^T WB)^{-1}$$

$$\Sigma = (A^T S^T (S\Gamma S^T)^{-1} SA)^{-1} \qquad (8.19)$$

where S is an (N-2) x N transformation matrix. If S were invertable, then (8.19) would reduce to Eqn. (8.7) and the estimate of x becomes a minimum variance one. However, further study is required to determine the conditions under which Σ closely approximates Σ_{MV} .

The above method assumes the data are equally spaced. If they are not, then, in general the computation of q_i and B_i requires the storage of all data prior to z_{i+2} and their associated A_i's. This would require an extraordinary amount of storage and is impractical for use in a computer program. Since most real data will not be equally-spaced, due to data editing or missing points, practical implementation of the method described in this section requires forming "dummy" observations at the missing data points. These "dummy" observations could be computed by interpolating between adjacent observations.

8.3.3 A Simple Example

To illustrate the method described in Section 3.2 and compare it to the WLS technique, consider the problem of estimating a constant scalar parameter. Then the measurement equation (Eqn. (8.4)) is

$$z = x_o + n \tag{8.20}$$

and the variance of the WLS estimate is

$$\Sigma_{WLS} = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{ij} \tag{8.21}$$

If R_{ij} is given by Eqn. (8.9) and the measurements are equally spaced, then (8.21) becomes

$$\Sigma_{WLS} = \frac{\sigma_o^2 (1+2\rho)}{NB_1^2}$$

$$+ 2e^{-2\alpha\tau} \left[2(1-e^{\alpha\tau} \frac{\alpha}{\beta} \sin \beta\tau - e^{-\alpha\tau} \cos \beta\tau) \right.$$

$$\left. + \rho\sigma_o^2 e^{-2\alpha\tau} + e^{-N\alpha\tau} C_N \right] \Big/ N^2 B_1^2 \tag{8.22}$$

where $$B_1 = 1 - 2e^{-\alpha\tau} \cos \beta\tau + e^{-2\alpha\tau}, \tag{8.23}$$

$$C_N = e^{\alpha\tau} \left[\cos(N+1)\beta\tau + \frac{\alpha}{\beta} \sin (N+1)\beta\tau \right] - 2 \left[\cos N\beta\tau + \frac{\alpha}{\beta} \sin N\beta\tau \right]$$

$$+ e^{-\alpha\tau} \left[\cos(N-1)\beta\tau + \frac{\alpha}{\beta} \sin (N-1)\beta\tau \right] , \tag{8.24}$$

and σ_o^2 and ρ are defined in Eqns. (8.14) and (8.15), respectively.

On the other hand, forming the new uncorrelated data type using the technique of Section 8.3.2 transforms the measurement equation to

$$q = Bx_o + v \tag{8.25}$$

and the variance of the estimate is

$$\Sigma = (B^T W B)^{-1}$$

$$= \sigma_o^2 \Big/ \sum_{i=1}^{N-2} B_i^2 \tag{8.26}$$

where

$$B_i = (B_1 - a_i B_{i-1}) \Big/ \sqrt{1 - a_i^2} \ , \tag{8.27}$$

$$a_i = \rho \Big/ \sqrt{1 - a_{i-1}^2} \ , \ i = 2, 3, \ldots, N-2, \tag{8.28}$$

$$a_1 = 0 \ .$$

For NB_1 large Σ_{WLS} becomes

$$\Sigma_{WLS} \simeq \sigma_o^2 \ (1 + 2\rho) / \ NB_1^2 \ , \quad NB_1 >> 1 \tag{8.29}$$

Approximating a_i in Eqn. (8.27) by $a = a_\infty$ and substituting in Eqn. (8.26) yields

$$\Sigma \approx \frac{\sigma_o^2 \ (1 + 2\rho)}{NB_1^2} \ , \quad N >> 1 \tag{8.30}$$

Thus $\Sigma \approx \Sigma_{WLS}$ for $NB_1 >> 1$. However, since B_1 can be quite small for small values of $\alpha\tau$ and $\beta\tau$, N can be fairly large while NB_1, is still small. In these cases both terms of (8.22) must be considered and it is difficult to see how Σ and Σ_{WLS} compare. Table 1 shows how Σ_{MV}, Σ_{WLS}, and Σ behave for values of $\alpha\tau$ and $\beta\tau$ ranging from .01 to 10 for intermediate values of N.

It is interesting to note that the data differencing technique compares very closely with MV when WLS performs poorly, and the data differencing technique does poorly when WLS compares closely with MV. Further analysis is required, however, to compare the MV, WLS, and data differencing techniques for more realistic problems.

$\alpha\tau$	$\beta\tau$	N	Σ_{MV}	Σ_{WLS}	Σ	Σ/Σ_{WLS}
.01	.001	12	.98	1.0	38	38
.01	.01	12	.95	1.0	19	19
.01	.1	12	.28	.89	.38	.43
.01	.1	24	.15	.63	.18	.28
.1	.01	12	.79	.92	3.8	4.1
.1	.1	12	.66	.85	1.9	2.2
.1	.1	18	.55	.74	1.2	1.6
.1	1	12	.036	.044	.038	.85
1	.1	12	.27	.29	.38	1.3
1	1	6	.30	.31	.46	1.5
1	1	12	.16	.16	.19	1.2
1	1	24	.081	.082	.090	1.1
1	10	12	.041	.042	.045	1.1

Table 1. Variances of MV, WLS, and Data Differencing Techniques for Different Values of $\alpha\tau$, $\beta\tau$, and N.

8.4 A GENERAL FORM OF THE DIFFERENCING METHOD

The general technique of forming a normalized, uncorrelated set of measurement from k data points can be applied to other types of correlated data. Difference equations which the autocorrelation function of the data must satisfy can be determined and, if an autocorrelation function which satisfies these equations approximates that of the actual data, then this method may be used to satisfactorily uncorrelate the data. The following paragraphs show how to form an uncorrelated measurement set from 3 data points and the extension to k data points is easily obtained.

Form the new variable, q_i, from the original data z_i by

$$q_i = (z_i - a_i z_{i-1} - b_i z_{i-2}) \Big/ K_i \quad , \; i = 3, 4, \ldots, N \tag{8.31}$$

where

$$q_1 = z_1 \; ,$$

$$q_2 = (z_2 - R_{21}\, z_1) \Big/ \sqrt{1-R_{21}^2} \; ,$$

$$E(z_i^2) = 1 \quad .$$

The requirements that

$$E(q_i q_j) = \begin{cases} 0 & \text{if} \quad i \neq j \\ 1 & \text{if} \quad i = j \end{cases} \tag{8.32}$$

allow a_i, b_i and K_i to be determined from

$$E(q_i q_{i-1}) = 0 \quad , \tag{8.33}$$

$$E(q_i q_{i-2}) = 0 \quad , \tag{8.34}$$

$$E(q_i^2) = 1 \quad , \quad i = 3, 4, \ldots, N \tag{8.35}$$

and constrain the autocorrelation function to satisfy

$$E(q_i q_j) = 0 \quad , \quad j < i-2 \ , \ i = 4, 5, \ldots, N \tag{8.36}$$

Substituting Eqn. (8.31) into (8.33) - (8.35) and noting that $E(q_i z_j) = 0$ for $j < i$ yields

$$a_i = (R_{i,i-1} - R_{i-1,i-2}\, R_{i,i-2})/(1 - R^2_{i-1,i-2})$$

$$b_i = (R_{i,i-2} - R_{i-1,i-2}\, R_{i,i-1})/(1 - R^2_{i-1,i-2})$$

$$K_i = \left[1 - a_i R_{i,i-1} - b_i R_{i,i-2} \right]^{1/2}$$

For equally spaced data, Eqns. (8.33) - (8.35) become

$$a_i = a = R_1(1 - R_2)/(1 - R_1^2)$$

$$b_i = b = (R_2 - R_1^2)/(1 - R_1^2)$$

$$K_i = K = \sqrt{(1 - R_2)(1 - 2R_1^2 + R_2)/(1 - R_1^2)}$$

where $R_k = R_{i,i-k}$. Eqn. (8.36) yields the second order difference equation

$$R_{k+2} - aR_{k+1} - bR_k = 0 \quad , \ 3 \le k \le N-2 \tag{8.37}$$

where $k = i - j$. The solution of (8.37) is of the form

$$R_k = A_1(r_1)^k + A_2(r_2)^k \quad , \quad 3 \le k \le N-2 \tag{8.38}$$

where

$$r_1 = \frac{a}{2} + \sqrt{d} \quad ,$$

$$r_2 = \frac{a}{2} - \sqrt{d} \quad ,$$

$$d = \left(\frac{a}{2}\right)^2 + b \quad .$$

A_1 and A_2 can be chosen to make R_3 and R_4 approximate the correlation of the real data.

Thus, if the data is such that Eqn. (8.38) can be used to approximate the autocorrelation function, then the method described in this section can be used to uncorrelate the data by differencing three terms.

8.5 TRANSFORMATION TO THE PROBLEM OF ESTIMATING AN UNKNOWN ACCELERATION

By interpreting the correlated data noise as the output of a second-order linear system (Section 8.3), the problem posed in Section 8.2 can be transformed to that of estimating an unknown acceleration as described in Section 4. To do this, augment the state vector to include n_o and $\dot{n}_o$, the correlated noise and its first derivative at epoch $t_o = 0$. Then, the linearized equations of motion of the new system become

$$\dot{X} = F\,X + Hw \tag{8.39}$$

where

$$X = (x^T,\ n,\ \dot{n})^T\ ,$$

$$F = \begin{pmatrix} F_1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -(\alpha^2+\beta^2) & -2\alpha \end{pmatrix},$$

$$F_1 = \frac{\partial \dot{x}}{\partial x}\ ,$$

$$H = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T\ .$$

The measurement equation then becomes

$$z = \begin{bmatrix} A & B & C \end{bmatrix}\ X_o + v \tag{8.40}$$

where $B = \partial z/\partial n_o$, $C = \partial z/\partial \dot{n}_o$, and v is zero-mean white noise uncorrelated with n. Although not included in the problem stated in Section 8.2, the white noise v allows the problem to be formulated as in Section 4.

8.6 CONCLUSIONS

The preceding sections presented three methods of processing correlated data. Of these, the two data differencing techniques are most easily implemented in a WLS orbit determination program. Before they are actually used, however, a comparison with the conventional WLS method should be performed to determine whether their implementation is warranted. As Magness and McGuire (Reference [34]) and the results of Table 1 have shown, the WLS estimate is often quite close to the MV estimate, even in the presence of highly correlated data.

9. CONCLUSIONS AND RECOMMENDATIONS

Orbit determination may be thought of as a hypothesis testing problem, in the sense that a statistical model of the data is postulated, an appropriate estimator is derived, real data is processed, a "best-estimate" trajectory is calculated, and the data residuals are examined to see if they are consistent with the statistical hypothesis. If not, the model must be changed or a rationale for accepting the result must be devised. From the theoretical point of view, it is necessary to develop a firm mathematical basis for any estimation algorithm so that the fundamental validity of the estimate is not in question. From the practical point of view, the estimator must be feasible to implement. With these considerations in mind, this report has discussed some of the theoretical and practical questions arising in sequential processing of tracking data.

In general it can be concluded that the maximum likelihood (weighted least squares) estimation technique, which has been much used in orbit determination work, is still the most practical algorithm for nonlinear orbit determination, sequential or otherwise. Several questions still remain to be investigated, however:

(1) the separation of estimation and control - the discussion of Section 3 offers a rationale for the presently used approach, but counterexamples can be constructed to show that such a separation is not correct in general. This subject needs to be pursued further, in particular, non-hamiltonian systems should be studied.

(2) the modal trajectory vs. the marginal mode - equations (3.24) and (3.28) show that two different forms of the maximum likelihood estimate can be constructed for nonlinear systems, whether control is present or not. This subject needs to be pursued further.

(3) nonlinear error analysis - throughout this report various treatments of nonlinear error analysis have been presented. For example, in Section 3.8 it was suggested that the bias in the estimate can be found by solving a certain differential equation. Such methods should be studied further and tested.

(4) estimation of unknown acceleration - this problem, which arises primarily in the treatment of model errors, has never been properly treated

in practical applications. It needs to be.

There are, of course, many other aspects of estimation theory which need to be examined for particular applications, such as the treatment of numerical errors, systematic errors, and correlated data. In general, it is recommended that some of the approaches described in this report be tested by numerical simulation, so that experience can be gained to suggest efficient design of future orbit determination computer programs, and, if necessary, further research.

REFERENCES

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," J. Basic Engrg., pp. 35-45, March 1960.

[2] R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," J. Basic Engrg., pp. 95-108, March 1961.

[3] C. G. Pfeiffer, "On the Sequential Estimation Technique," IEEE Transactions on Automatic Control, Vol. AC-11, pp. 132-133, Jan. 1966.

[4] K. G. Brammer and C. G. Pfeiffer, "Comment on the Sequential Estimation Techniques," IEEE Transactions on Automatic Control, Vol. AC-11, pp. 767, October 1966.

[5] R. E. Mortensen, "Mathematical Problems of Modeling Stochastic Nonlinear Dynamic Systems," NASA CR-1168, September 1968.

[6] D. D. Morrison, "Methods for Nonlinear Least Squares Problems and Convergence Proofs," Seminar Proceedings, Tracking and Orbit Determination Seminar, Jet Propulsion Laboratory, Pasadena, California February 23-26, 1960.

[7] C. G. Pfeiffer, "On the Identification of Observable Orbit Parameters, with Application to Lunar Orbiter Tracking," to appear in Journal of the Astronautical Sciences," Jan-Feb. 1969.

[8] V. O. Mowery, "Least Squares Recursive Differential-Correction Estimation in Nonlinear Problems," IEEE Trans. Automatic Control, Vol. Ac-10, pp. 399-407, October 1965.

[9] R. L. Stratonovich, "On the Theory of Optimal Nonlinear Filtering of Random Functions," Theo. Prob. Appl., Vol. 4, pp. 223-225.

[10] R. L. Kashyap, "On the Partial Differential Equation for the Conditional Probability Distribution for Nonlinear Dynamic Systems with Noisy Measurements," Tech. Report No. 432, Division of Engineering and Applied Physics, Howard University, 1964.

[11] H. J. Kushner, "On the Dynamical Equation of Conditional Probability Functions, with Applications to Optimum Stochastic Control Theory," J. Math. Anal. Appl., Vol. 8, pp. 332-355, 1964.

[12] H. J. Kushner, "On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes, with Applications," SIAM J. Control, Vol. 2, pp. 106-119, 1964.

[13] R. S. Bucy, "Nonlinear Filtering Theory," IEEE Trans. Automatic Control, Vol. AC-10, p. 198, 1965.

[14] R. E. Mortensen, "Optimal Control of Continuous-time Stochastic Systems," Report No. ERL-66-1, Electronics Research Laboratory, College of Engineering, University of California, Berkeley, 19 August 1966.

[15] S. R. McReynolds, "A Note on Stratonovich's Equation for the Conditional Probability Distribution of the State of a Nonlinear System, with Application to Nonlinear Filtering," Space Programs Summary 37-42, Vol. IV, pp. 22-27, Jet Propulsion Laboratory, Pasadena, 31 Dec. 1966.

[16] T. E. Duncan, "Probability Densities for Diffusion Processes with Applications to Nonlinear Filtering Theory and Detection Theory," Report No. SU-SEL-67-035, Stanford University, Center for Systems Research, May, 1967.

[17] M. Zakai, "On the Optimal Filtering of Diffusion Processes," E. E. Publication No. 80, Israel Institute of Technology, Haifa, Israel, October 1967.

[18] F. P. Romeo, "On Stochastic Differential Equations Arising in State Estimation Problems," Memorandum No. ERL-M246, Electronics Research Laboratory, College of Engineering, University of California, Berkeley.

[19] R. E. Mortensen, "Continuous-time Stochastic Problems," lectures notes for Engineering Extension course 867.10, University of California, Los Angeles.

[20] R. L. Stratonovich, "Conditional Markov Processes," Theory of Probability and its Applications, Vol. V, pp. 156-178, 1960.

[21] W. M. Wonham, "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering," J. SIAM Control, Series A, Vol. 2, pp. 347-369, 1965.

[22] H. J. Kushner, "Dynamical Equations for Optimal Nonlinear Filtering," J. of Differential Equations, Vol. 3, April 1967.

[23] H. J. Kushner, "Approximations to Optimal Nonlinear Filters," Proceedings of 1967 Joint Automatic Control Conference.

[24] J. R. Fisher, "Optimal Nonlinear Filtering," Report No. 66-5, Dept. of Engineering, University of California, Los Angeles, January 1966.

[25] R. W. Bass, V. D. Norum, and L. Schwartz, "Optimal Multichannel Nonlinear Filtering," J. Math. Anal. Appl., Vol. 16, No. 1, pp. 152-164, October 1966.

[26] L. Schwartz, "Approximate Continuous Nonlinear Minimal-Variance Filtering," Report No. SSD-60472 R, Hughes Aircraft Co., Space Systems Division, December 1966.

[27] R. S. Bucy, and P. D. Joseph, "Filtering for Stochastic Processes with Applications to Guidance," Wiley - Interscience, New York, 1968.

[28] A. V. Balakrishnan, editor, "Communication Theory," McGraw-Hill, 1968. See Chapter 3, "Estimation Theory."

[29] R. E. Mortensen, "Maximum Likelihood Recursive Nonlinear Filtering," Journal of Optimization Theory and Applications, Vol. 2, No. 6, pp. 386-394, 1968.

[30] E. A. Coddington, and N. Levinson, "Theory of Ordinary Differential Equations," McGraw-Hill, New York, 1955.

[31] R. H. Battin, "Astronautical Guidance," McGraw-Hill, New York, 1964.

[32] H. E. Rauch, "Optimum Estimation of Satellite Trajectories Including Random Fluctuations in Drag," AIAA Journal, Vol. 3, No. 4, April 1965.

[33] A. E. Bryson, and M. Frazier, "Smoothing for Linear and Nonlinear Dynamic Systems," Proceedings of the Optimum System Conference, Wright-Patterson Air Force Base, Ohio, Aeronautical Systems Division, TDR-63-119, pp. 353-364, September 1962.

[34] T. A. Magness, and J. B. McGuire, "Statistics of Orbit Determination-Correlated Observations," STL Doc. No. 8976-6001-RU-000, 15 Dec 1961.