MEASUREMENT COVARIANCE-CONSTRAINED ESTIMATION FOR POORLY MODELED DYNAMIC SYSTEMS

by

Daniel Joseph Mook

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in

ENGINEERING MECHANICS

Approved:

J. L. Junkins

E. M. Cliff

S. L. Hendricks

L. G. Kraige

D. P. Telionis

December 1985
Blacksburg, VA

Acknowledgements

The author wishes to express his sincere appreciation to Dr. John L. Junkins for his guidance and support throughout this work, both technical and otherwise, including the difficult last few months via "long distance". The education he has provided includes invaluable lessons on the inner workings of research institutions, in addition to the subject matter of the research itself. Both types of information should prove to be extremely useful in the years ahead.

It is a pleasure to acknowledge the excellent typing and administrative support provided by Ms. Vanessa McCoy.

This work is dedicated to my wife, , and son, , who made it seem worthwhile, and to my parents, , who made it all possible.

Predominant funding of this work was supplied by the Naval Surface Weapons Center, Dahlgren, Virginia, under contract N60921-83-G-9-A165. The contract monitor was Dr. J. N. Blanton, who contributed many useful discussions and an occasional needed prod. This support is gratefully acknowledged.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

CHAPTER 1

INTRODUCTION

1.1 Introduction to Optimal Estimation

The field of "optimal estimation" is concerned with utilizing any available information to estimate the state history of a system. Generally, this information is available in two forms: a mathematical model of the system, providing theoretical insights and information, and one or more sets of measurements, providing experimental insights and information.

An "optimal estimator" may be defined as a computational algorithm that combines measurement information and model information with the objective of determining a minimum error estimate of the state history of a system. In general, a properly weighted compromise estimate of the actual state history which utilizes both the model and the measurements is superior to an estimate based upon either the model or the measurements exclusively, since neither the model nor the measurements are perfect.

1.1.1 Statement of the Dynamic Estimation Problem

The typical statement of a dynamic estimation problem, and the one used throughout this work, is:

Given the system state (dynamic) model, in the form of a system of nonlinear differential equations

$$\dot{\underline{x}} = \underline{f}(\underline{x}(t),\underline{u}(t),t) \tag{1.1}$$

and the discrete measurements

$$\tilde{\underline{y}}_k = \underline{g}_k(\underline{x}(t_k),t_k) + \underline{v}_k \tag{1.2}$$

where

$\underline{x}$ = n x 1 state vector

$\underline{f}$ = n x 1 vector of state model functions

$\underline{u}$ = ℓ x 1 known forcing vector

$\underline{\tilde{y}}_k$ = m x 1 measurement vector at t_k

$\underline{g}_k$ = m x 1 vector of measurement model functions evaluated at t_k

$\underline{v}_k$ = m x 1 measurement error vector,

$$E\{\underline{v}_k\} = \underline{0},\ E\{\underline{v}_k \underline{v}_k^T\} = R_k = \text{known}$$

find the optimal estimate of $\underline{x}(t)$. The definition of "optimal" is clearly subjective; various criteria will be defined where appropriate in the developments which follow.

The present study does not address real-time estimation ("filtering"), wherein the measurements are processed sequentially as they become available, but post-experiment estimation - i.e., all of the measurements in the time domain of interest are assumed to be available at the time of execution. In many applications, the algorithms execute fast enough to essentially eliminate any practical distinction between real-time and post-experiment estimation.

1.1.2 Model Errors

The focus of the present work is on errors in the system model; specifically, for dynamic systems, unmodeled effects in the system of governing equations. In many dynamic systems, both the forces acting on the system and the system parameters themselves are difficult to determine accurately. In addition, and perhaps more importantly, errors in the state estimate $\underline{x}(t)$ will produce errors in the right hand side of

Eq. (1.1), which lead to incorrect integration of the model; this "compounding" of errors can quickly destroy the accuracy of the estimate.

Inaccurate state estimates can easily lead to other problems. For example, many systems are controlled by feedback controllers which depend upon the state estimates to determine the control forces. Poor state estimates may evoke inappropriate control forces, causing the system to behave erratically and possibly become unstable (e.g., Alspach[1] (1975) or Tse, Bar-Shalom, and Meier[2] (1973)). The entire September 1984 issue of Automatica[3] was devoted to the problem of "adaptive control" in the presence of state estimate uncertainties.

1.2 Process Noise in Kalman Filters

The overwhelming majority of estimation techniques in use today are essentially of the "Kalman filter" type, first proposed by Kalman[4] (1960) and Kalman and Bucy[5] (1961). This is especially true for the class of systems in which substantial model errors are known or suspected to be present. The extensions, modifications, clarifications, adaptations, applications, etc., of Kalman filters fills many shelves in the VPI&SU library. In Chapter 3 of the present study, a derivation of one form of the so-called "Extended Kalman Filter-Smoother" (EKFS) is given. ("Filtering" refers to real-time estimation; "smoothing" refers to post-experiment estimation. In the present study, which deals with post-experiment estimation, all references to "filter-type" algorithms are understood to include smoothing.) For a much broader theoretical

and practical development, dozens of books are available (c.f., e.g., 6-9).

The developments in Chapter 3, and the discussion in the current section, are included for comparison purposes. The method developed in this work is not an extension of any known estimation algorithm, but an alternative to existing methodology. The predominance of Kalman filter-type estimation algorithms in use today requires that any proposed alternative be compared directly with Kalman filters. The form of the Kalman filter developed in Chapter 3 follows a paper by Chang and Tabaczynski[10] (1984), in which they found this algorithm to be the most effective for target tracking applications; since the present work was motivated by the related problem of spacecraft orbit attitude estimation, this form is used in all of the numerical comparisons of the present study.

The handling of model errors in filter algorithms is via so-called "process noise". Equation (1.1) is modified to the form

$$\dot{\underline{x}} = \underline{f}(\underline{x}(t),\underline{u}(t),t) + \underline{w}(t) \qquad (1.3)$$

where $\underline{w}$ is usually assumed to be a zero mean white noise process. Literally hundreds of papers have appeared in the literature which treat model errors as in Eq. (1.3). The only known alternative for treating model errors is to parameterize them using assumed functions with unknown coefficients; the coefficients can then be estimated. Some survey papers ([10], [11], [7], [8] also Refs. 7, 8) are listed in the references; each of these overview the literature and cite many additional references.

The integration of Eq. (1.3) as written is impossible for unknown $\underline{w}$; consequently, the state vector estimate is actually integrated via Eq. (1.1). The process noise is incorporated by inflating the so-called "gain matrix"; details are in Chapter 3. For the present discussion, the gain matrix is used to determine the relative weighting between the measurements and the model in the estimate. Thus, including process noise does not actually improve the model, but simply produces an estimate which is more dependent on the recent measurements. This approach can work well when the measurements are relatively dense in time and are accurate. However, if the measurement accuracy is poor, and/or the measurements are sparse in time, the process noise approach often has difficulty obtaining an accurate estimate; these qualitative remarks are supported by example problem results presented in Chapters 4 and 5.

Furthermore, there is no mechanism for estimating the unmodeled effects; in fact, as previously stated, the model actually integrated is usually left unchanged by the inclusion of process noise. Thus, even though poor estimates obtained without the addition of process noise are generally improved when an appropriate amount of process noise is added, no information is obtained which might improve the model in future applications.

Finally, the amount of process noise to be added is an educated guess by the user. If the covariance of $\underline{w}$ is known, and if $\underline{w}$ is idealized (e.g., a white noise process), the modification of the gain matrix is deterministic (c.f., e.g., [7]). In practice, statistical information about $\underline{w}$ is rarely, if ever, known (e.g., Mehra[13] (1970)).

More often, the model errors are due to systematic non-white noise effects, and thus the algorithm must be applied artistically or "tuned". Thus, in practice, the choice of an appropriate level of process noise is based upon a subjective criterion determined by the user.

The existence of an extensive literature (and growing; dozens of new articles have appeared in 1985) on the subject of model errors in estimation algorithms indicates continuing interest in the subject. However, virtually since its inception, the "process noise" approach has been shown to have some undesirable properties. For example, Tukey[12] (1960) showed that deviations from Gaussian noise may seriously degrade the performance of filter algorithms based on Gaussian assumptions. The filter algorithms are dependent on the estimate error covariance matrix for the determination of the gains; however, process noise is added directly to this covariance matrix, casting serious doubt on its accuracy. In some cases, the filter algorithms may diverge (e.g., Fitzgerald[14] (1971); Huber[15] (1972); Breza and Bryson[16] (1974)). Of course, these criticisms are not meant to imply that useful results cannot be obtained from Kalman filter-type algorithms. The intent is to emphasize the substantial gap which exists between the idealized assumptions and the realities of most actual applications.

An extensive literature exists on various ad hoc solutions to some of the problems associated with process noise. The most extensive is in an area called "adaptive filtering", wherein the unknown process noise covariance matrix elements are calculated as part of the solution. Many approaches exist (e.g., Jazwinski[17] (1969); Mehra[18] (1972); Chin[19]

(1979); Tsai and Kurz[20] (1983); Dee, Cohn, Dalcher, and Ghil[21] (1985)). The principal advantage of these approaches is that the amount of process noise to be added is calculated as part of the solution, removing some of the ad hoc guesswork. However, the criteria for determining the process noise varies from method to method and remains a subjective choice of the user.

Adaptive filtering has been expanded to include time-varying process noise covariance, i.e., the amount of process noise to be added to the estimate error covariance matrix is allowed to vary with time (e.g., Bar-Shalom, Tse, and Dressler[22] (1973)); also, to non-gaussian, zero mean process noise (e.g., Alspach and Sorensen[23] (1971); Masreliez and Martin[24] (1977)). Some methods have been developed which essentially skip the addition of process noise to the estimate error covariance and instead calculate the gain directly (e.g., Carew and Belanger[25] (1973)).

To summarize, the Kalman filter-type algorithms use process noise to account for model errors. This approach has the following drawbacks:

(1) The model is not improved; rather, the state estimates are obtained by shifting the relative emphasis from the model to the recent measurements. Clearly, this can cause problems if the recent measurements are poor and/or sparse in time.

(2) Since the model is not improved, the between-measurement state estimates are still subject to accumulation of model errors even if the measurements are perfect.

(3) The amount of process noise added is a subjective choice of the user; consequently, two users utilizing identical models

and identical measurements will typically obtain different estimates. Clearly, this subjective tuning is undesirable.

(4) Upon completion of the estimation process, no information about the unmodeled effects is available, even though process noise may have significantly improved the state estimate.

1.3 Jump Discontinuities in Discrete Sequential Estimation Algorithms

Although the primary focus of the present work is on model errors, there is another important disadvantage of sequential estimation algorithms (such as Kalman filters) - jump discontinuities in the state vector estimates at each discrete measurement time. The sequential algorithms process each discrete measurement set by adjusting the state estimate at the measurement time (obtained by integrating the model forward from the previous measurement time) via a state estimate jump discontinuity toward the measurement values. For most physical systems, jump discontinuities in the state vector are not physically realizable; thus, the discontinuous Kalman filter estimate history does not strictly correspond to a physical motion of the system. This issue is examined in detail in Chapter 3. The method of the present work, developed in Chapter 2, does not produce jump discontinuities in the estimated state vector.

1.4 Outline of the Present Method

The essential feature of the present method is the application of control theory concepts to the development of an estimation algorithm which can account for non-random model errors in a more satisfactory

manner than the process noise approach of the Kalman filter-type algorithms which currently dominate estimation practice.

In control theory, the unknown functions (controls) which drive a system to produce an optimal state history are calculated. The state history may have known boundary conditions specified by the user. The control force history required to effect the optimal state history is determined by using variational calculus to minimize a functional which penalizes the deviation of the state history from some reference or target state history. In many practical problems, the functional to be minimized also penalizes the control forces themselves, so that an optimum balance between state departures (from the reference state) and the control history is found.

In the present method, model errors are represented in Eq. (1.1) as

$$\dot{\underline{x}} = \underline{f}(\underline{x}(t),\underline{u}(t),t) + \underline{d}(t) \tag{1.4}$$

where $\underline{d}(t)$ is the vector of unmodeled disturbances. No assumptions other than piecewise continuity are required concerning $\underline{d}(t)$; note that this is in sharp contrast to process noise, which is always zero mean and usually gaussian, and which must have a specified covariance. Equation (1.4) has the same form as many models used in the control literature, where $\underline{d}(t)$ would be interpreted as the to-be-determined control force.

The criterion used for determining $\underline{d}(t)$ in the present study is the minimization of the functional

$$J = \sum_{k=1}^{M} [\tilde{\underline{y}}_k - \underline{g}_k(\hat{\underline{x}}(t_k),t_k)]^T R_k^{-1} [\tilde{\underline{y}}_k - \underline{g}_k(\hat{\underline{x}}(t_k),t_k)]$$

$$+ \int_{t_1}^{t_M} \underline{d}^T(t) W \underline{d}(t) dt \qquad (1.5)$$

where $\hat{\underline{x}}$ is the state vector estimate, W is a positive definite weighting matrix, M is the number of measurement sets, and the other terms are defined in Eq. (1.2). The estimate $\hat{\underline{x}}(t)$ and the optimal $\underline{d}(t)$ are implicitly required to satisfy the differential equation, Eq. (1.4). The first term in Eq. (1.5) is the weighted sum square of the residuals between the actual measurements and the estimated measurements, and the second term is a weighted integral sum square of the unmodeled disturbance. By adjusting the value of W, thereby penalizing the unmodeled disturbance to a greater or lesser degree, the estimated measurements may be driven toward or away from the actual measurements. The criterion for determining W is to force the estimated measurements to match the actual measurements with the same statistical accuracy as the measurements match the truth; specifically, choose W so that

$$R_k \approx [\tilde{\underline{y}}_k - \underline{g}_k(\hat{\underline{x}}(t_k), t_k)][\tilde{\underline{y}}_k - \underline{g}_k(\hat{\underline{x}}(t_k), t_k)]^T \qquad (1.6)$$

Equation (1.6), which approximately equates the measurement-minus-estimated measurement covariance with the prescribed (R_k) measurement-minus-truth covariance, will hereafter be called the "covariance constraint".

To summarize the present method, the unmodeled disturbance which must be added to the model in order to produce the statistically correct estimated measurements will be determined using control theory concepts. Not only is a state estimate obtained which has a statistically correct accuracy, but the smallest unmodeled effects

required to produce this "statistically consistent" state estimate are also obtained. Thus, the model itself is improved; the estimate is improved; and the model error is estimated. Note that Eq. (1.5) is a more general functional than those typically encountered in most control formulations, since the first sum includes penalty terms at interior time points in addition to the initial and final times.

This approach is a significant departure from other estimation algorithms, since the state history is in a sense treated as a constraint. Of course, it is only constrained in the sense that its deviation from the measurements is required to be consistent with the measurement accuracy, i.e., the estimate is required to fit the measurements with the same accuracy as the measurements are assumed to be distributed about the truth.

CHAPTER 2

DEVELOPMENT OF THE MINIMUM MODEL ERROR ESTIMATION ALGORITHM

2.1 Introduction

In this chapter, a variational calculus development of the present method is given. The development is closely related to similar developments in optimal control theory; see, e.g., Bryson and Ho[26] (1969), Kirk[27] (1970), or Pun[28] (1970). However, the adaptation to development of an estimation algorithm is apparently unique.

The history of variational calculus includes many of the legendary figures in mathematics, physics, and classical mechanics - Galileo, Newton, Leibniz, John and James Bernoulli, Euler, Lagrange, Legendre, Jacobi, Hamilton, Mayer, Bolza, Bliss, Pontryagin, to name a few. A fascinating history is contained in Goldstine[29] (1980); unfortunately, it is far beyond the scope of the present work.

The classical calculus of variations was organized, formalized, and greatly extended by a group of researchers centered around the University of Chicago during the first five decades of this century. Led by Oskar Bolza and G. A. Bliss, this group, commonly referred to as the "Chicago School", essentially laid the groundwork for modern optimal control theory. Representative summaries of their work are contained in Bolza[30] (1904) and Bliss[31] (1946). Apparently, they did not pursue practical applications of their work, and it remained largely unknown to the engineering community.

In the 1950's, a group of Russian researchers began to explore variational calculus as a means of obtaining solutions to optimal control problems. The "Pontryagin's Maximum Principle" was first

hypothesized by Pontryagin[32] (1956) and Boltyanskii, Gamkrelidze, and Pontryagin[33] (1956) for some specific cases. The principle was proven for linear systems by Gamkrelidze[34] (1957) and Boltyanskii[35] (1958). Gamkrelidze[36] (1958) extended the maximum principle to the general case of minimizing (maximizing) an arbitrary functional which includes integrals of the system variables and/or control variables. Such a functional is defined in Section 2.2, and conditions for its minimization are derived. These are the so-called "Pontryagin's Necessary Conditions". Pontryagin's work extended the classical variational calculus by admitting discontinuous forcing functions and especially, by communicating the necessary conditions in a form easy to understand and apply.

There are several summary derivations of this basic work in the literature; see, e.g., Rozonoer[37,38,39] (1960) or Kopp[40] (1962). The variational principle which is presented in this chapter involves a modest departure from the literature, but is included for completeness since it leads directly to the intended applications in estimation theory.

In Section 2.3, the functional is expanded to include terms of the state variables at discrete times between the initial and final time. Although this problem was solved by Geering[41] (1976), the derivation in Section 2.3 is a new extension of the well-known results in Section 2.2 and is not similar to Geering. However, the results are identical. In Section 2.4, the results of Section 2.3 are adapted to the formulation of an estimation strategy. Finally, a simple example is presented in Section 2.5.

2.2 Pontryagin's Necessary Conditions

Given a system modeled by the (generally nonlinear) system of equations,

$$\dot{\underline{x}} = \underline{f}[\underline{x}(t), \underline{u}(t), t] \quad , \quad \underline{x}(t_o) \text{ specified} \tag{2.1}$$

where

$\underline{x} \equiv$ n x 1 state vector

$\underline{f} \equiv$ n x 1 vector of model equations

$\underline{u} \equiv$ n x 1 force vector

and a performance index defined as

$$J = \phi[\underline{x}(t_f), t_f] + \int_{t_o}^{t_f} L[\underline{x}(\tau), \underline{u}(\tau), \tau]d\tau \tag{2.2}$$

where

$\phi \equiv$ penalty on the final state $\underline{x}(t_f)$

$L \equiv$ penalty reflecting the deviation of $\underline{x}(t)$, $\underline{u}(t)$ from their desired trajectories

The problem may be stated as: find a smooth, differentiable, unbounded $\underline{u}(t)$, $t_o \leq t \leq t_f$, which minimizes J, subject to the constraint Eq. (2.1).

Utilizing the Lagrange multiplier technique, the augmented function is written as

$$J' \equiv \phi(\underline{x}(t_f), t_f) + \int_{t_o}^{t_f} \{L[\underline{x}(\tau), \underline{u}(\tau), \tau] + \underline{\lambda}^T(\tau)[\underline{f}[\underline{x}(\tau), \underline{u}(\tau), \tau] - \dot{\underline{x}}]\}d\tau \tag{2.3}$$

Here, the vector $\underline{\lambda}(t)$ may be interpreted as a vector of Lagrange

multipliers. In control theory literature, they are usually called the "co-states" or "adjoint variables". Note that if the constraint (Eq. (2.1)) is exactly satisfied, then J' = J regardless of the values for $\underline{\lambda}(t)$. The last term in the integrand of Eq. (2.3) may be integrated by parts to yield

$$\int_{t_o}^{t_f} \underline{\lambda}^T(\tau)\dot{\underline{x}}d\tau = (\underline{\lambda}^T\underline{x})_{t_f} - (\underline{\lambda}^T\underline{x})_{t_o} - \int_{t_o}^{t_f} \dot{\underline{\lambda}}^T(\tau)\underline{x}(\tau)d\tau \tag{2.4}$$

Primarily for convenience (although there may be physical interpretations in some problems), the remainder of the integrand in Eq. (2.3) is defined as the Hamiltonian:

$$H = L[\underline{x}(t), \underline{u}(t), t] + \underline{\lambda}^T(t)\underline{f}[\underline{x}(t), \underline{u}(t), t] \tag{2.5}$$

Equations (2.4) and (2.5) are substituted into (2.3) to obtain

$$J' = \phi(\underline{x}(t_f), t_f) + \int_{t_o}^{t_f} \{H[\underline{x}(\tau), \underline{u}(\tau), \tau)] + \dot{\underline{\lambda}}^T(\tau)\underline{x}(\tau)\}d\tau - (\underline{\lambda}^T\underline{x})_{t_f} + (\underline{\lambda}^T\underline{x})_{t_o} \tag{2.6}$$

Using the Lagrange multiplier rule (c.f., e.g., [32]) minimizing J' with respect to $\underline{u}(t)$ is equivalent to minimizing J with respect to $\underline{u}(t)$, provided the constraint is exactly satisfied. Let $\bar{\underline{u}}(t)$ denote the unknown $\underline{u}(t)$ which minimizes J'. Consider the variation in J' due to a variation $\delta\underline{u}$ about $\bar{\underline{u}}(t)$; for a stationary value of J', this variation should be zero:

$$\delta J' = 0 = [(\frac{\partial \phi}{\partial \underline{x}} - \underline{\lambda}^T)\delta \underline{x}]_{t_f} + (\underline{\lambda}^T \delta \underline{x})_{t_o}$$

$$+ \int_{t_o}^{t_f} [(\frac{\partial H}{\partial \underline{x}} + \dot{\underline{\lambda}}^T)\delta \underline{x} + \frac{\partial H}{\partial \underline{u}} \delta \underline{u}]d\tau \qquad (2.7)$$

Note that in Eq. (2.7), the initial and final times (t_o and t_f) are assumed to be fixed. Presumably, the initial conditions $\underline{x}(t_o)$ are independent of $\underline{u}(t)$, so

$$\delta \underline{x}(t_o) = 0 \qquad (2.8)$$

In general, it may be possible to determine the relationship between $\delta \underline{u}$ and $\delta \underline{x}$ in Eq. (2.7) in order to impose $\delta J' = 0$. However, because of the freedom introduced by the multipliers (co-states), which are as yet undefined, the problem of relating the variations $\delta \underline{x}$ and $\delta \underline{u}$ may be avoided altogether by requiring that the coefficients of $\delta \underline{x}$ and $\delta \underline{u}$ be zero independently; thus, for $\delta J' = 0$, the necessary conditions may be written as:

$$\underline{\lambda}^T(t_f) = \frac{\partial \phi}{\partial \underline{x}}\bigg|_{t_f} \qquad (2.9)$$

$$\frac{\partial H}{\partial \underline{x}} = -\dot{\underline{\lambda}}^T \qquad (2.10)$$

$$\frac{\partial H}{\partial \underline{u}} = 0 \qquad (2.11)$$

Equation (2.9) is the boundary condition on $\underline{\lambda}(t_f)$, commonly called (after Bernoulli) the "transversality condition"; Eqs. (2.10), (2.11), and (2.1) are commonly referred to as the "Pontryagin's Necessary Conditions".

Collectively, Eqs. (2.1), (2.9), (2.10), and (2.11) constitute a two-point boundary-value problem (TPBVP), so-called because the boundary conditions are split between t_o and t_f. Specifically, the problem may be rewritten as:

Solve the 2n differential equations,

$$\dot{\underline{x}} = \underline{f}(\underline{x}(t), \underline{u}(t), t) \tag{2.1}$$

$$\dot{\underline{\lambda}} = -\left[\frac{\partial H}{\partial \underline{x}}\right]^T = -\left(\frac{\partial \underline{f}}{\partial \underline{x}}\right)^T \underline{\lambda} - \left(\frac{\partial L}{\partial \underline{x}}\right)^T \tag{2.12}$$

with boundary conditions

$$\underline{x}(t_o) = \text{given initial state} \tag{2.13}$$

$$\underline{\lambda}(t_f) = \left(\frac{\partial \phi}{\partial \underline{x}}\right)^T_{t_f} \tag{2.9}$$

and $\underline{u}$ in Eqs. (2.1) and (2.12) given by

$$\frac{\partial H}{\partial \underline{u}} = 0 = \left(\frac{\partial \underline{f}}{\partial \underline{u}}\right)^T \underline{\lambda} + \left(\frac{\partial L}{\partial \underline{u}}\right)^T \tag{2.14}$$

This problem can be solved iteratively by the shooting method: "guess" $\underline{\lambda}(t_o)$; integrate forward to t_f; check $\underline{\lambda}(t_f) \stackrel{?}{=} \left(\frac{\partial \phi}{\partial \underline{x}}\right)^T_{t_f}$; if not, update $\underline{\lambda}(t_o)$ and try again. There are numerous methods for numerically solving TPBVP's (e.g., Vadali[42] (1982) or Keller[43] (1976)); a general discussion is beyond the scope of the present work. The extensions to deal with discontinuous and/or bound constrained $\underline{u}(t)$ are available in the literature (e.g., [32]-[36]).

This section presents a basic, classical variational calculus derivation of the Pontryagin's Necessary Conditions for the smooth, unbounded functions $\underline{u}(t)$ which minimize the performance index J. There are many other possibilities for some of the assumptions in this

derivation. For example, if the final states are specified, then $\delta\underline{x}(t_f) = 0$, so the transversality condition (Eq. (2.9)) is replaced by the end condition

$$\underline{x}(t_f) = \text{known} \tag{2.15}$$

Another possibility is that the initial conditions are unknown; then $\delta\underline{x}(t_0) \neq 0$, so the initial condition is

$$\underline{\lambda}(t_0) = 0 \tag{2.16}$$

Another possibility, of course, is that $\underline{u}(t)$ may be bounded or discontinuous. Note that in all of these cases, a TPBVP still must be solved since the boundary conditions remain split. Other important cases which have not been covered here include unknown initial and/or final times. Many texts on variational calculus and optimal control contain exhaustive permutations of the various possibilities; for good practical summaries, see the texts by Kirk (27) or Bryson and Ho (26). In the next section, the important extension to internal penalty terms on the states at discrete times is derived.

2.3 Extension to Internal Penalty Terms

In this section, Pontryagin's Neccesary Conditions, derived in the previous section for a cost function including an integral penalty term and a final state penalty term, are extended to the more general case when penalty terms on the states at discrete times between t_0 and t_f are present. Accordingly, the cost functional is defined as

$$J = \sum_{i=0}^{N} K_i[\underline{x}(t_i), t_i] + \int_{t_o}^{t_N} L[\underline{x}(\tau), \underline{u}(\tau), \tau]dt \tag{2.17}$$

where for simplicity, $t_N \equiv t_f$.

The system state model is, as before,

$$\dot{\underline{x}} = \underline{f}[\underline{x}(t), \underline{u}(t), t] \tag{2.18}$$

so that the augmented cost function is

$$J' = \sum_{i=0}^{N} K_i[\underline{x}(t_i), t_i] + \int_{t_o}^{t_N} \{L[\underline{x}(\tau), \underline{u}(\tau), \tau] + \underline{\lambda}^T(\tau)[\underline{f}(\underline{x}(\tau), \underline{u}(\tau), \tau) - \dot{\underline{x}}]\}d\tau \tag{2.19}$$

At this point, there are no restrictions on $\underline{\lambda}(t)$. In order to accommodate the internal penalty terms, the co-states $\underline{\lambda}(t)$, and therefore $\underline{u}(t)$, will be allowed to have jump discontinuities at each discrete penalty time (c.f., e.g., Geering[41]). The last term in the integrand of Eq. (2.19) may be rewritten as

$$\int_{t_o}^{t_N} \underline{\lambda}^T(\tau)\dot{\underline{x}}(\tau)d\tau = \int_{t_o^+}^{t_1^-} \underline{\lambda}^T(\tau)\dot{\underline{x}}(\tau)d\tau + \int_{t_1^-}^{t_2^-} \underline{\lambda}^T(\tau)\dot{\underline{x}}(\tau)d\tau + \ldots + \int_{t_{N-1}^+}^{t_N^-} \underline{\lambda}^T(\tau)\dot{\underline{x}}(\tau)d\tau \tag{2.20}$$

where the superscripts (+) and (-) on the discrete times are defined to mean "approaching from the positive side" and "approaching from the negative side", respectively. Thus, $\underline{\lambda}(t_i^-)$ and $\underline{\lambda}(t_i^+)$ represent the two values of $\underline{\lambda}(t_i)$ due to the jump discontinuity - note, however,

that $t_i^- = t_i^+$. The right hand side of Eq. (2.20) is integrated by parts to yield

$$\int_{t_o}^{t_N} \underline{\lambda}^T(\tau)\dot{\underline{x}}(\tau)d\tau = (\underline{\lambda}^T\underline{x})_{t_1^-} - (\underline{\lambda}^T\underline{x})_{t_o^+} - \int_{t_o}^{t_1^-} [\dot{\underline{\lambda}}^T(\tau)\underline{x}(\tau)]d\tau$$

$$+ (\underline{\lambda}^T\underline{x})_{t_2^-} - (\underline{\lambda}^T\underline{x})_{t_1^+} - \int_{t_1^+}^{t_2^-} [\dot{\underline{\lambda}}^T(\tau)\underline{x}(\tau)]d\tau$$

$$+ \ldots$$

$$+ (\underline{\lambda}^T\underline{x})_{t_N} - (\underline{\lambda}^T\underline{x})_{t_{N-1}^+} - \int_{t_{N-1}^+}^{t_N} [\dot{\underline{\lambda}}^T(\tau)\underline{x}(\tau)]d\tau$$

$$= \sum_{i=0}^{N-1} [(\underline{\lambda}^T\underline{x})_{t_{i+1}^-} - (\underline{\lambda}^T\underline{x})_{t_i^+}] - \int_{t_o}^{t_N} \dot{\underline{\lambda}}^T(\tau)\underline{x}(\tau)d\tau \qquad (2.21)$$

By substituting Eq. (2.21) into Eq. (2.19), the augmented cost function J' may be written

$$J' = \sum_{i=1}^{N} K_i[\underline{x}(t_i), t_i] - \sum_{i=0}^{N-1} [(\underline{\lambda}^T\underline{x})_{t_{i+1}^-} - (\underline{\lambda}^T\underline{x})_{t_i^+}]$$

$$+ \int_{t_o}^{t_N} \{H[\underline{x}(\tau), \underline{u}(\tau), \tau] + \dot{\underline{\lambda}}^T(\tau)\underline{x}(\tau)\}d\tau \qquad (2.22)$$

where the Hamiltonian is given in Eq. (2.5) as

$$H \equiv L[\underline{x}(t), \underline{u}(t), t] + \underline{\lambda}^T(t)\underline{f}[\underline{x}(t), \underline{u}(t), t] \qquad (2.5)$$

Following the procedure used in Section 2.2, consider small variations in $\underline{u}(t)$ in the vicinity of a stationary value of J'. For

stationary J', $\delta J' = 0$:

$$\delta J' = 0 = \sum_{i=0}^{N} \left.\frac{\partial K_i}{\partial \underline{x}}\right|_{t_i} \delta\underline{x}(t_i) - \sum_{i=0}^{N-1} [(\underline{\lambda}^T \delta\underline{x})_{t_{i+1}^-} - (\underline{\lambda}^T \delta\underline{x})_{t_i^+}]$$
$$+ \int_{t_o}^{t_N} [(\frac{\partial H}{\partial \underline{x}} + \dot{\underline{\lambda}}^T)\delta\underline{x} + \frac{\partial H}{\partial \underline{u}} \delta\underline{u}]d\tau \quad . \tag{2.23}$$

As before, the coefficients of δx and δu can be set to zero independently. This yields the "modified Pontryagin's Necessary Conditions" (for the case when penalty terms are present on the states at discrete times within the time domain) as

$$\dot{\underline{x}} = \underline{f}(\underline{x}(t), \underline{u}(t), t) \tag{2.1}$$

$$\dot{\underline{\lambda}} = -[\frac{\partial H}{\partial \underline{x}}]^T = -[\frac{\partial \underline{f}}{\partial \underline{x}}]^T \underline{\lambda} - (\frac{\partial L}{\partial \underline{x}})^T \tag{2.24}$$

$$\frac{\partial H}{\partial \underline{u}} = 0 \tag{2.11}$$

and, at the internal penalty times,

$$\left.\frac{\partial K_i}{\partial \underline{x}}\right|_{t_i} + \underline{\lambda}^T(t_i^+) - \underline{\lambda}^T(t_i^-) = 0$$

or

$$\underline{\lambda}^T(t_i^+) = \underline{\lambda}^T(t_i^-) - \left.\frac{\partial K_i}{\partial \underline{x}}\right|_{t_i} \tag{2.25}$$

Equations (2.1), (2.24), (2.11), and (2.25) constitute a TPBVP as before, with the following boundary conditions:

At t_o: either $\underline{x}(t_o)$ = known,

or $\underline{\lambda}(t_o^-) = 0$ (2.26)

At t_f: either $\underline{x}(t_f)$ = known,

or $\underline{\lambda}(t_f^+) = 0$ (2.27)

2.4 Adaptation to Estimation

In this section, the results of Section 2.3 are applied to dynamic estimation problems. The discrete dynamic estimation problem is stated as: Given a system model

$$\dot{\underline{x}} = \underline{f}[\underline{x}(t), t] \tag{2.28}$$

with measurements $\tilde{\underline{y}}$ at discrete times modeled by

$$\tilde{\underline{y}}(t_j) = \underline{g}[\underline{x}(t_j), t_j] + \underline{V}(t_j) \quad , \quad j = 1,\ldots,m \tag{2.29}$$

where in general, $\tilde{\underline{y}}$ and $\underline{x}$ are of different length, and $\underline{V}(t)$ is a Gaussian white noise sequence with known covariance R_j, determine the "minimum error" estimate of $\underline{x}(t)$.

The approach taken in the current work is to minimize the functional

$$J = \sum_{j=1}^{m} \{[\tilde{\underline{y}}(t_j) - \underline{g}(\hat{\underline{x}}(t_j), t_j)]^T R_j^{-1} [\tilde{\underline{y}}(t_j) - \underline{g}(\hat{\underline{x}}(t_j), t_j)]\} + \int_{t_o}^{t_f} \underline{u}^T W \underline{u} d\tau \tag{2.30}$$

where

$\tilde{\underline{y}}(t_j) \equiv$ r x 1 measurement set at t_j

$\hat{\underline{x}} \equiv$ n x 1 estimated state vector

$\underline{u} \equiv$ k x 1 unmodeled disturbance vector

$R_j \equiv$ r x r assumed covariance of the measurement error at t_j

$W \equiv$ k x k weight matrix to be determined

The unmodeled disturbance vector, $\underline{u}(t)$, is simply added to the right hand side of the model equations as

$$\dot{\underline{x}} = \underline{f}[\underline{x}(t),t] + \underline{u}(t) \tag{2.31}$$

Minimization of J with respect to $\underline{u}(t)$ may now be performed utilizing the results of Section 2.3.

The functional J is clearly a weighted sum of discrete penalty terms on the estimated states at the discrete measurement times, and a weighted integral square penalty term on the unmodeled disturbances. The discrete terms penalize the deviation of the estimated measurements, obtained by substituting the estimated states into the measurement model, from the actual measurements. The weighting on each discrete term is the inverse of the associated measurement error covariance matrix; thus, deviations from "good" measurements (small covariance) are weighted (and hence penalized) more heavily than deviations from "bad" measurements (large covariance). The integral term reflects the assumption that the most likely values of the disturbance history are the smallest ones which yield the correct state history. The determination of the estimated state history and the weight matrix W are explained in the following paragraphs.

For a given value of W, the minimization of J in Eq. (2.30) proceeds exactly as in Section 2.3. Considering Eqs. (2.31), (2.24) and (2.11), the governing equations for the states and co-states may be written

$$\dot{\underline{x}} = \underline{f}[\underline{x}(t), \underline{u}(t), t] \tag{2.1}$$

$$\dot{\underline{\lambda}} = - [\frac{\partial \underline{f}}{\partial x}]^T \underline{\lambda} \tag{2.32}$$

$$\frac{\partial H}{\partial \underline{u}} = 0 = \frac{\partial L}{\partial \underline{u}} + [\frac{\partial \underline{f}}{\partial \underline{u}}]^T \underline{\lambda}$$

$$\rightarrow \underline{u} = - \frac{1}{2} W^{-1} [\frac{\partial \underline{f}}{\partial \underline{u}}]^T \underline{\lambda} \tag{2.33}$$

Equation (2.33) is substituted into Eqs. (2.31) and (2.32) to eliminate $\underline{u}$ in terms of λ. The discrete penalty term at time t_j is

$$K_j \equiv [\tilde{\underline{y}}(t_j) - \underline{g}(\hat{\underline{x}}(t_j),\ t_j)]^T R_j^{-1} [\tilde{\underline{y}}(t_j) - \underline{g}(\hat{\underline{x}}(t_j),\ t_j)] \qquad (2.34)$$

The jump discontinuities in the co-states at t_j are (from Eq. (2.25))

$$\underline{\lambda}(t_j^+) = \underline{\lambda}(t_j^-) - \left.\frac{\partial K_j}{\partial \hat{\underline{x}}}\right|_{t_j}$$

$$= \underline{\lambda}(t_j^-) + 2H_j^T R_j^{-1} [\tilde{\underline{y}}(t_j) - \underline{g}(\hat{\underline{x}}(t_j),\ t_j)] \qquad (2.35)$$

where

$$H \equiv \left.\frac{\partial \underline{g}}{\partial \hat{\underline{x}}}\right|_{\hat{\underline{x}}(t_j),\ t_j}$$

Thus, the TPBVP for a given W may be summarized as:

$$\dot{\underline{x}} = f[\underline{x}(t), t] + \underline{u}(t) \qquad (2.31)$$

$$\dot{\underline{\lambda}} = -\left[\frac{\partial f}{\partial x}\right]^T \underline{\lambda} \qquad (2.32)$$

$$\underline{u} = -\frac{1}{2} W^{-1} \left[\frac{\partial f}{\partial \underline{u}}\right]^T \underline{\lambda} \qquad (2.33)$$

$$\underline{\lambda}(t_o^-) = 0 \qquad (2.26)$$

$$\underline{\lambda}(t_j^+) = \underline{\lambda}(t_j^-) + 2H_j^T R_j^{-1} [\tilde{\underline{y}}(t_j) - \underline{g}(\hat{\underline{x}}(t_j),\ t_j)] \qquad (2.35)$$

$$\underline{\lambda}(t_f^+) = 0 \qquad (2.27)$$

An approximate set of initial conditions for $\hat{\underline{x}}(t_o)$ is required for the first trial integration. The trial can be taken as the solution of

$$\tilde{\underline{y}}(t_o) - \underline{g}(\hat{\underline{x}}(t_o),\ t_o) = 0 \qquad (2.36)$$

if $r \geq n$; otherwise, several sets of measurements near t_o could be used to construct a local least square approximation. As needed, $\hat{\underline{x}}(t_o)$

will be adjusted iteratively by the TPBVP solution technique, so it is only necessary to obtain a starting estimate sufficiently close to ensure convergence. $\underline{\lambda}(t_0^+)$ is determined using Eq. (2.35).

The criterion used for determining the optimal $\hat{\underline{x}}$ will be based on the covariance test. This covariance consistency test, to see if $\hat{\underline{x}}$ is acceptable, may take many forms. The most general is to check the covariance at each measurement time as follows: determine the actual estimate covariance,

$$\hat{R}_j = [\tilde{\underline{y}}(t_j) - \underline{g}(\hat{\underline{x}}(t_j), t_j)][\tilde{\underline{y}}(t_j) - \underline{g}(\hat{\underline{x}}(t_j), t_j)]^T \qquad (2.37)$$

and compare it element-by-element with the measurement covariance R_j. Due to "small sample" statistics idosyncracies, this check is not particularly realistic. In the typical case when all of the measurement sets are of the same quantities with the same error covariance, the <u>average</u> covariance can be checked:

$$\text{Is } R \approx \frac{1}{M} \sum_{j=1}^{M} \hat{R}_j? \qquad (2.38)$$

where $\hat{R}_j$ is given in Eq. (2.37), R is the actual measurement covariance, and M is the number of measurement sets. The interpretation of $\approx$ in Eq. (2.38) may itself take many forms. For example, the norms of the matrices could be compared; the diagonal elements (representing the individual variances of the $\hat{x}_i$'s) could be compared; or the comparison could be element-by-element. Specific schemes are investigated in the examples; the intent here is to indicate the qualitative concept.

By finding the minimum disturbance error estimate, $\hat{\underline{x}}$, which predicts the measurements with essentially the same variance as the prescribed measurement variance, the solution is "covariance

constrained". The observation is made that if the minimum disturbance estimate is tuned to match the measurements with the same statistics as the assumed moments of the measurement error distributions, then the estimate has been optimized in an intuitively appealing and physically reasonable way. Numerous test cases support the practicality of this principle; however, it has not been established analytically. The fact that arbitrary model errors are allowed probably precludes a more rigorous justification for this formulation.

The solution of the TPBVP for a given W yields a single minimum disturbance state history, $\hat{\underline{x}}(t)$. The issue of "tuning" is now addressed qualitatively. If the state history does not satisfy the covariance check, then the balance between small model errors and small measurement residuals has not been achieved. If the measurement residual covariance is higher than the apriori measurement covariance, then the estimate history is not close enough to the measurement history, and a larger disturbance history should be allowed to the model to correct the error. Thus, W should be decreased, so that $\underline{u}(t)$ is penalized less heavily. The minimization of J can then proceed with higher magnitudes of $\underline{u}(t)$, allowing more flexibility in $\hat{\underline{x}}(t)$ to better predict the measurements. On the other hand, if the estimate covariance is too low, then the estimate history is too close to the noisy measurements; i.e., the measurements have been matched better than necessary (or desirable) as indicated in the apriori measurement covariance. Thus, W should be increased, so that less unmodeled disturbance (correction) is added to the system model.

The solution of the TPBVP for a given W yields a continuous state history. The co-states, and consequently, the state derivatives, are only piecewise continuous with usually small discontinuities occurring at the measurement times. By comparison, discrete Kalman filter estimates are discontinuous in the states themselves. This is clearly undesirable, especially in precision post-experiment estimation of a smooth physical process.

When W has been found such that the covariance check is satisfied, the desired solution has been obtained. Using Eq. (2.33), the disturbance history can be calculated from the converged co-state history. Although the estimated disturbance history generally contains (usually small) jump discontinuities at the measurement times, it provides an estimate of the actual disturbance which is often very accurate, as is evident in the results presented below and in subsequent chapters.

2.5 Simple Scalar Example

To illustrate the application of the present method, consider the problem of estimating the state history of a scalar function of time for which measurements are the only information available. No knowledge of the system state governing equations is assumed. It is known that the measurements are direct measurements of the state itself, and the measurement noise is a zero mean gaussian process with an assumed known variance of σ^2.

Using the method presented in this chapter, the system state model is

$$\dot{x} = 0 + u(t) \tag{2.37}$$

where x is the state and u is the "ummodeled effect". The measurements are given as

$$\tilde{x}_k = x_k + v_k \tag{2.38}$$

where $\tilde{x}_k$ is the measurement at time t_k, x_k is the true state at time t_k, and v_k is a zero mean gaussian process. The cost functional to be minimized is

$$J = \sum_{i=0}^{N} (\tilde{x}_i - \hat{x}_i)^2 \sigma_2^{-1} + \int_{t_o}^{t_N} u^2 W dt \tag{2.39}$$

where $\hat{x}_i$ is the estimate at time t_i and W is the to-be-determined weight on the integral sum-square unmodeled effect history. The Hamiltonian is

$$H = u^2 W + \lambda^T u \tag{2.40}$$

so that the TPBVP may be summarized as

$$u = -\frac{1}{2} W^{-1} \frac{\partial f}{\partial u} \lambda = -\frac{\lambda}{2W} \tag{2.41}$$

$$\dot{x} = u = -\frac{\lambda}{2W}$$

$$\dot{\lambda} = -\frac{\partial f}{\partial x} \lambda = 0$$

$$\lambda(t_o^-) = \lambda(t_N^+) = 0$$

$$\lambda(t_k^+) = \lambda(t_k^-) + \frac{2}{\sigma^2} [\tilde{x}_k - \hat{x}_k]$$

The solution proceeds according to the following steps:

1) Choose W.

2) Set $\hat{x}_o = \tilde{x}_o$.

3) Integrate forward to t_N, adjusting λ at t_k.

4) Check: $\lambda_N^+ \approx 0$? If so, skip ahead to step 7.

5) Determine $\frac{\partial \lambda_N^+}{\partial \hat{x}_o}$, then $\Delta \hat{x}_o = - \left[\frac{\partial \lambda_N^+}{\partial \hat{x}_o} \right]^{-1} \lambda_N^+$.

6) Adjust $\hat{x}_o$ by $\Delta \hat{x}_o$; go to step 3.

7) Check "covariance constraint":

Is $\frac{1}{N} \sum_{i=0}^{N} (\tilde{x}_i - \hat{x}_i)^2 \approx \sigma^2$?

If so, stop.

8) Adjust W and go to step 3.

In Fig. (2.1-a), a set of 101 measurements with variance $\sigma^2 = 0.1136$ is shown, spanning the time 0 to 10. The measurements were obtained by adding gaussian noise to the unit cosine wave as

$$\tilde{x}_i = \cos t_i + V_i \qquad (2.42)$$

Thus, the true state history is x = cos t; however, this is not readily apparent in the scattered measurements of Fig. (2.1-a). In Fig. (2.1-b), the "covariance constrained" solution is shown, obtained by the solution procedure just outlined. The underlying true x is also shown for comparison. Note that the cosine wave has been reconstructed to an error variance of 0.0085, considerably better than the measurement variance.

In Fig. (2.1-c), the true state-minus-estimated state residual is shown, and in Fig. (2.1-d), the recovered disturbance is shown. The "true disturbance" - specifically, -sin t - is also plotted for comparison purposes. Although the recovered disturbance contains a considerable amount of noise, it is seen to follow the true

disturbance. A perceptive user, viewing the recovered unmodeled effect history in Fig. (2.1-d), might decide to attempt to model the state dynamics as

$$\dot{x} = - \sin t + u(t) \tag{2.43}$$

If this is done, the present method solution may be made to match the true state history to very high precision. Of course, it is not necessary to attempt fitting of the recovered estimate for u(t), and the validity of the basic approach does not depend upon this problem - dependent and intuitive step.

In Figs. (2.2-2.3), the solution is shown for different measurement sets. In Fig. (2.2), $\sigma^2 = 0.011$; in Fig. (2.3), $\sigma^2 = 0.00011$. (The nominal variances in Figs. (2.1-2.3) are 10^{-1}, 10^{-2}, 10^{-4} respectively; however, the random number generator used to generate the measurement errors is somewhat, well, random.) As the measurement noise decreases, the variance of the solution decreases to about 0.0020; apparently, this is the approximate limit of this approach in the absence of a model, with measurements spaced $\Delta t = 0.1$. Considering the TPBVP equations in Eq. (2.41), it is seen that

$$\dot{x} = - \frac{\lambda}{2W}$$

$$\dot{\lambda} = 0$$

Thus, the state estimate between measurements is a straight line with slope = $- \lambda/2W$. When fitting this sequence of straight lines to a cosine wave, the accuracy will be limited by the measurement frequency.

However, the disturbance recovery improves dramatically; see Fig. (2.3-d). In the event that the disturbance term is recognized and added

to the model (as in Eq. (2.43)), the estimate matches the truth to arbitrary accuracy.

In Chapters 4 and 5, the present method is compared with the Kalman filter algorithm developed in Chapter 3 for performance on some more interesting test problems.

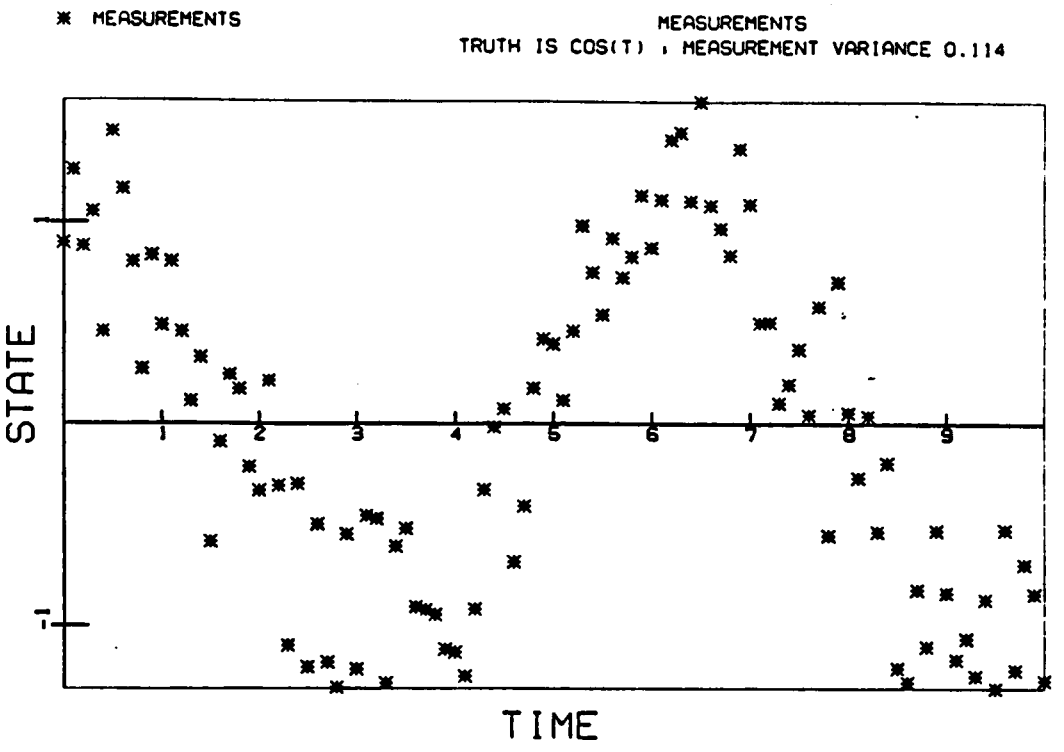


Figure 2.1-a

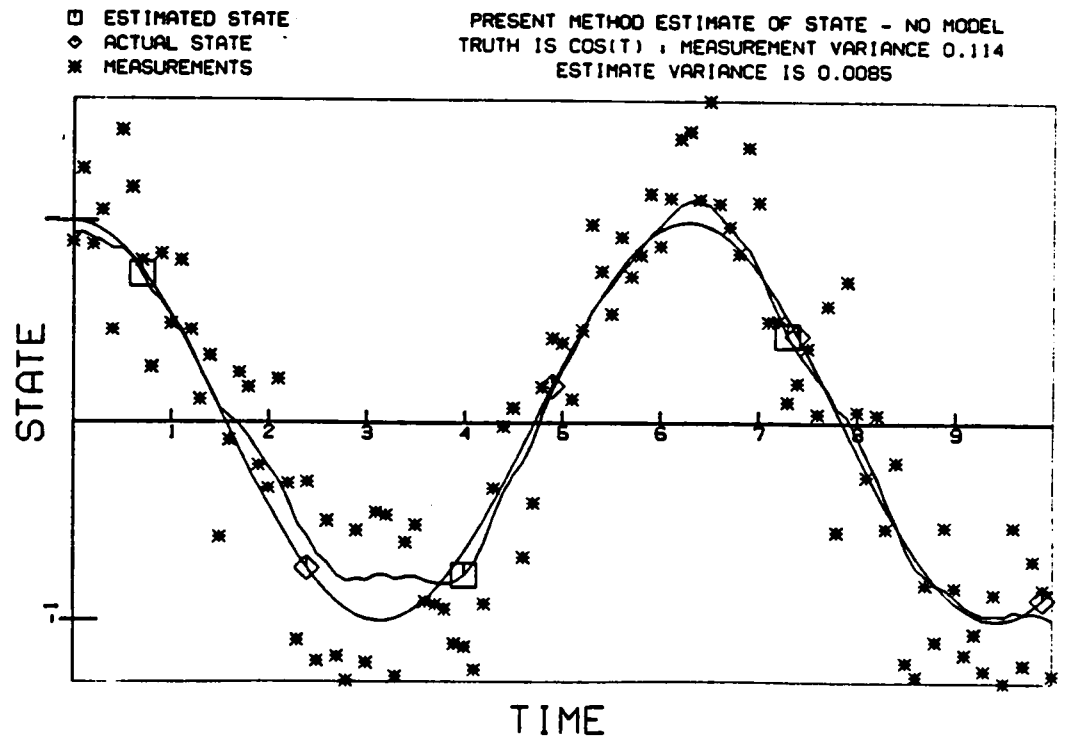


Figure 2.1-b

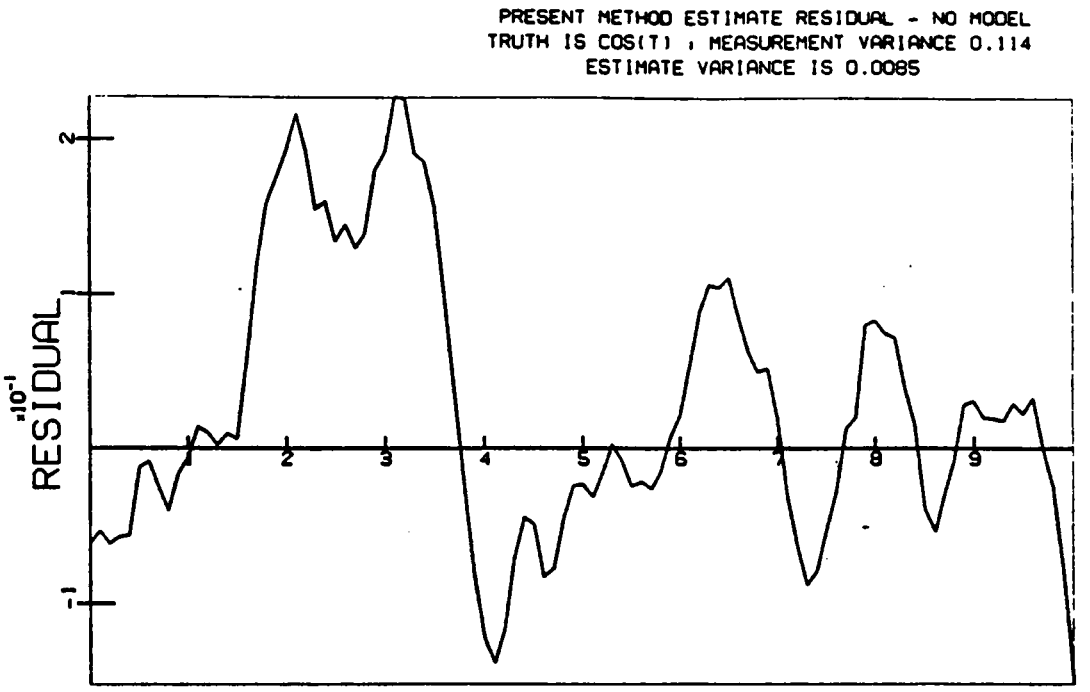


Figure 2.1-c

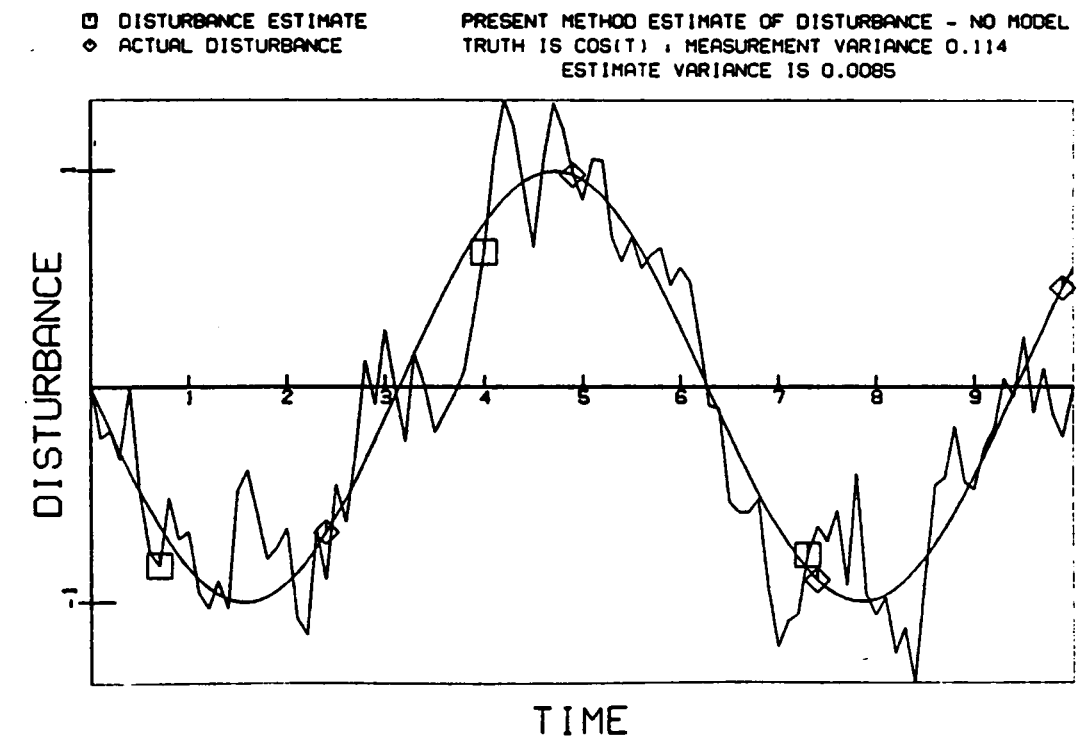


Figure 2.1-d

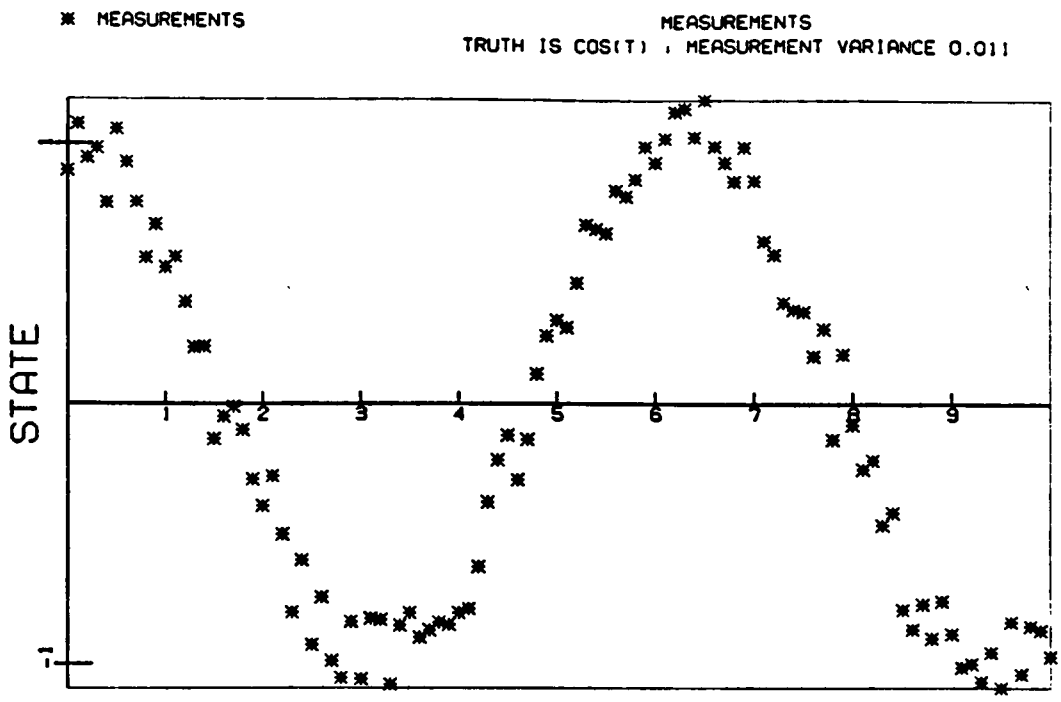


Figure 2.2-a

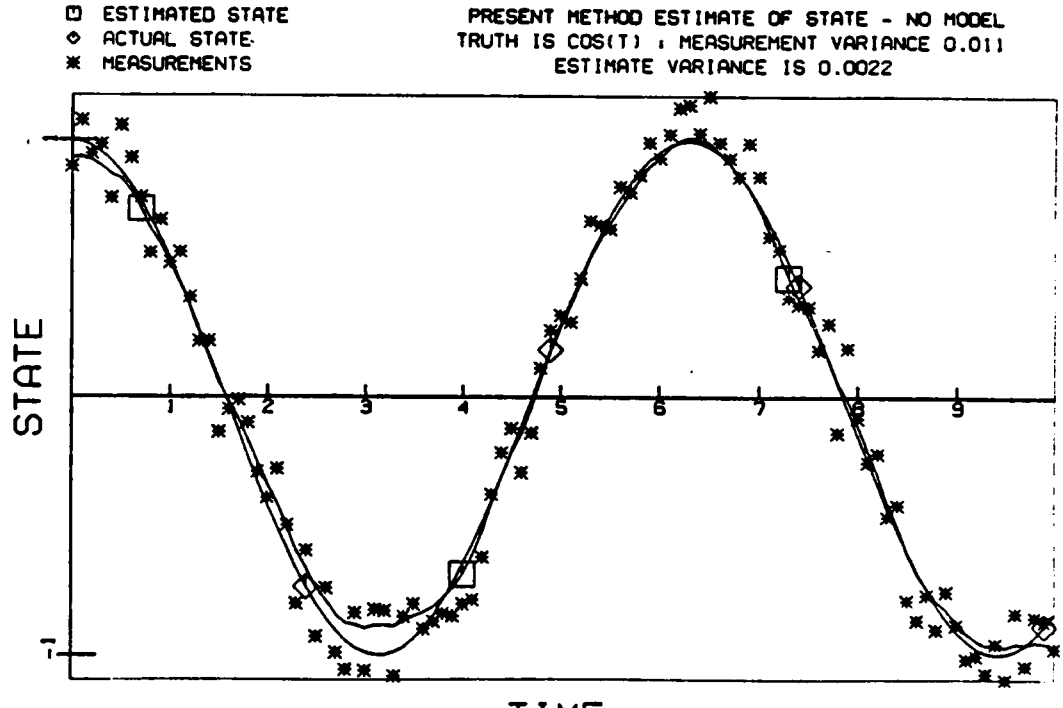


Figure 2.2-b

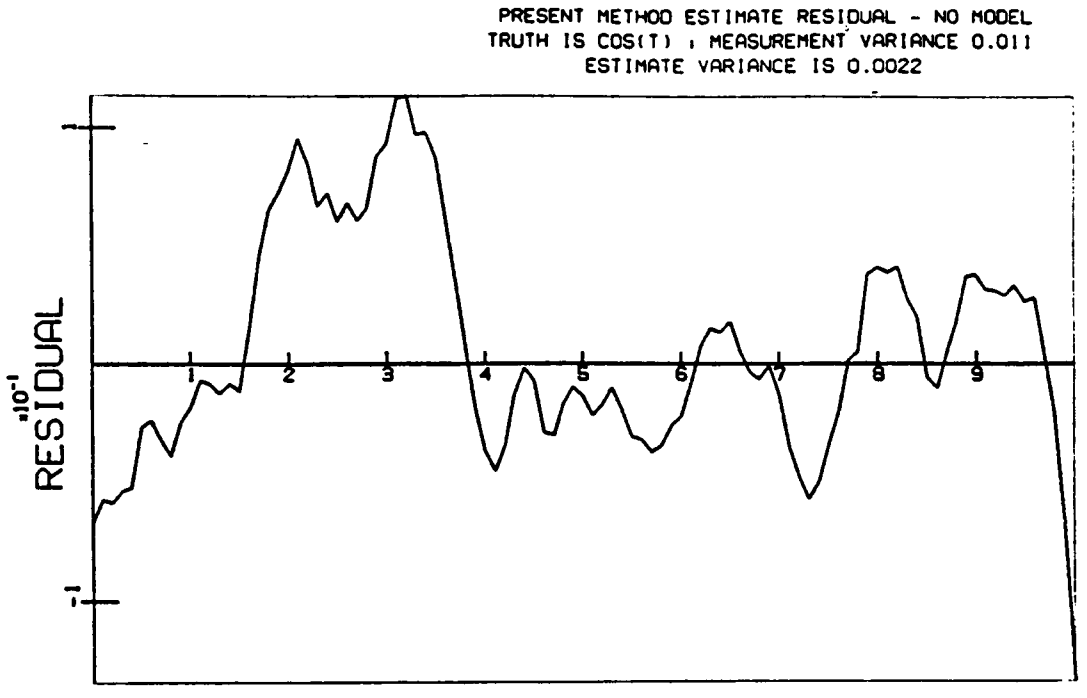


Figure 2.2-c

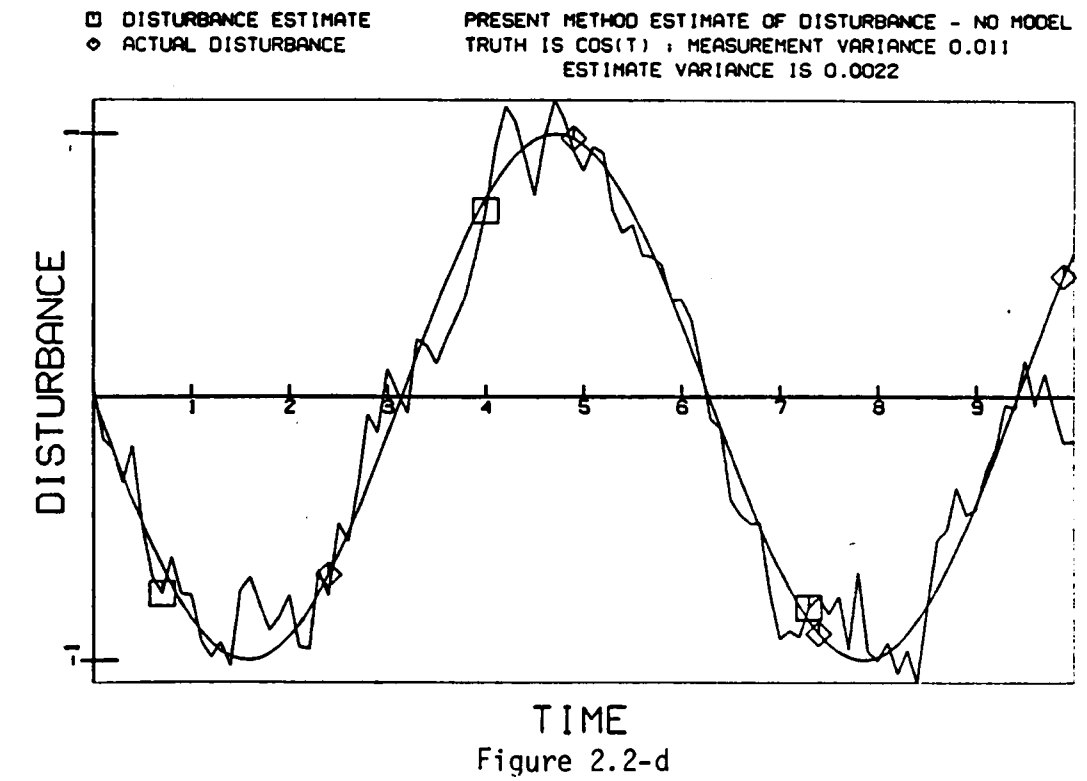


Figure 2.2-d

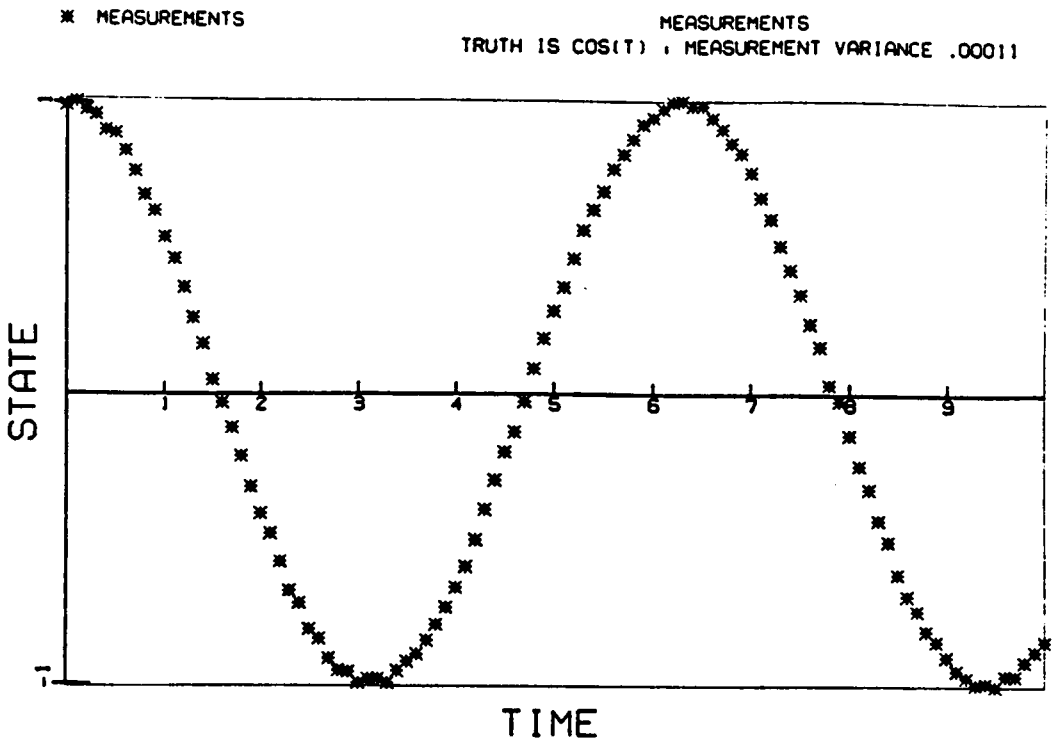


Figure 2.3-a

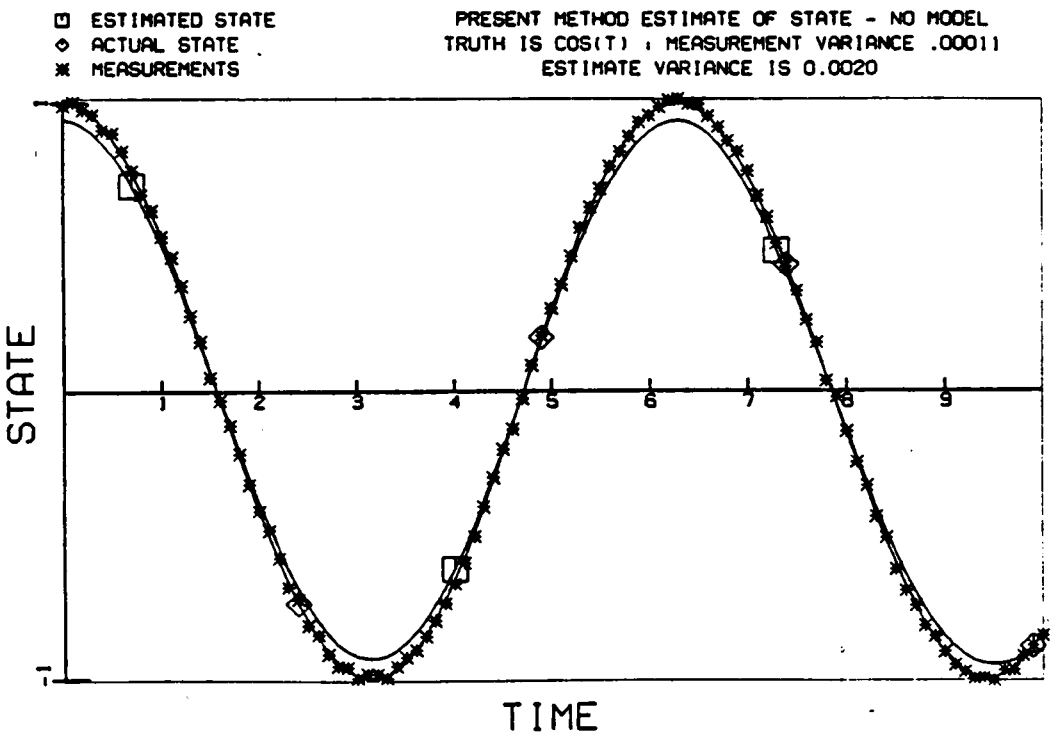


Figure 2.3-b

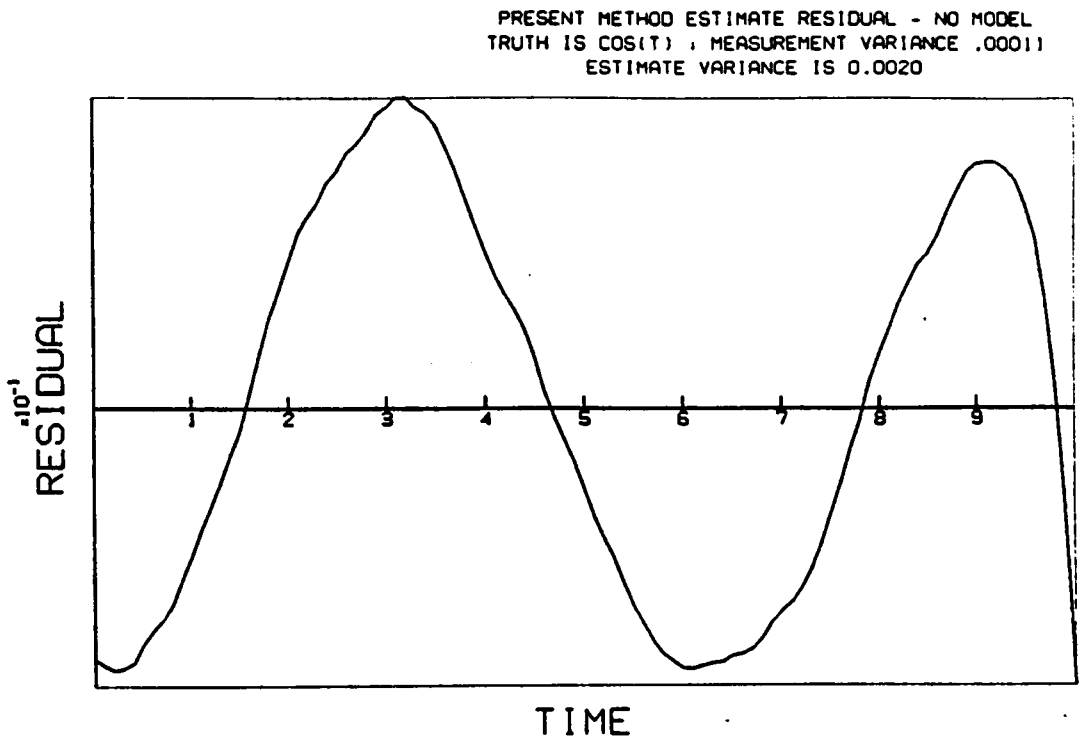


Figure 2.3-c

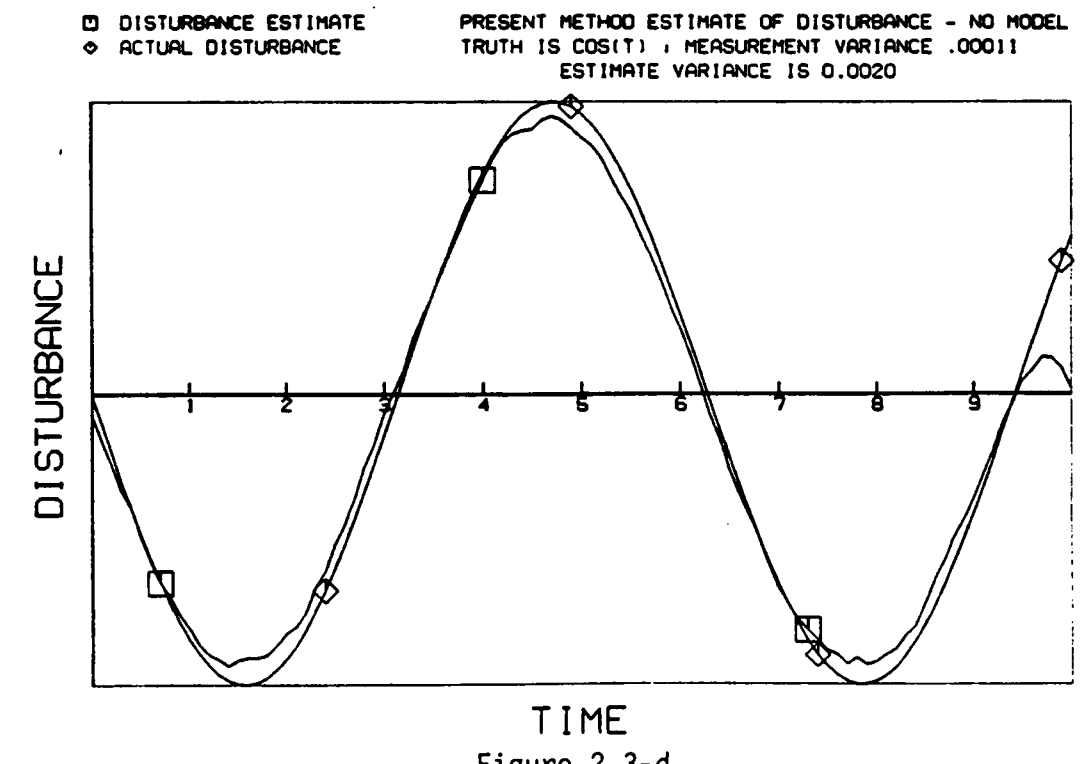


Figure 2.3-d

CHAPTER 3

KALMAN FILTER ESTIMATION

3.1 Introduction

In this chapter, several estimation strategies which are members of the "Kalman filter family" are derived and explored. The reason for including the Kalman filter algorithms in detail is the observation that these approaches are the most widely used estimation strategies in satellite applications (e.g., Chang and Tabaczynski[10] (1984)), and probably in all dynamic estimation applications. Also, since a variety of implementations exist, the specific algorithms used in the comparison studies of the present work are presented.

Kalman filters are "sequential estimators", which means that they process the measurement sets one at a time, sequentially through the time domain. Assuming that the calculations may be performed quickly enough (true for many applications), Kalman filters may be used in real time. The method proposed in this work, like other so-called "batch estimators", processes many measurement sets simultaneously, and so it is probably not appropriate for real time estimation. However, the length of time required to process a batch of measurements is shrinking as the boom in computational capability continues; a solution time of a few seconds or less may be adequate for many "real time" applications. Kalman estimation has become increasingly popular in non-real-time, post-experiment estimation problems. This trend is a direct consequence of the fact that classical batch estimators have no provisions for accommodating model errors.

A significant problem with Kalman filters is so-called "memory". Specifically, the Kalman gain matrix, which determines the corrections to the estimate based upon each new measurement set processed, tends to zero as the number of measurement sets processed increases. Thus, the classical Kalman filter tends to ignore any new measurements after it has processed some sufficient number of measurements. The estimator becomes "satisfied" that it has found the right answer.

However, if the model is imperfect and new measurement information is ignored, the estimate will diverge from the truth - perhaps very rapidly. Thus, the memory of the Kalman filter must be reduced, by increasing the covariance estimate associated with the present state estimate. This is accomplished using "process noise", as is shown in the developments of this chapter.

The net effect of adding process noise to the Kalman filter is to force the estimator to place more emphasis on the new measurements. Most often, this improves the estimate, as is shown in the test cases. However, the model is not improved, and if the measurements are sparse in time and/or of low accuracy, then shifting emphasis to them will be of limited value.

In Section 3.2, the classical Kalman filter is derived in detail. In Section 3.3, the so-called "extended Kalman filter" for nonlinear system models is developed. In Section 3.4, a "smoother" algorithm which reprocesses the estimate to reduce the size of the jump discontinuities at the measurement times is presented.

3.2 Classical Kalman Filter

Recall that the intent of any optimal estimator is to combine information from the model with the measurements in order to obtain a better estimate of the state history of a system than would be obtained by basing the estimate on the model or the measurements exclusively.

At the measurement time t_j, the following form of the estimate strategy is assumed:

$$\underline{\hat{x}}_j(+) = K'_j\underline{\hat{x}}_j(-) + K_j\underline{\tilde{y}}_j \tag{3.1}$$

where

$\underline{\hat{x}}_j(+) \equiv$ "New" estimate utilizing measurement set $\underline{\tilde{y}}_j$

$\underline{\hat{x}}_j(-) \equiv$ Model predicted value of $\underline{\hat{x}}(t_j)$, utilizing information up to but not including $\underline{\tilde{y}}_j$

$\underline{\tilde{y}}_j \equiv$ Measurement set at time t_j

$K'_j, K_j \equiv$ Gain matrices, used to specify how $\underline{\hat{x}}_j(-)$ and $\underline{\tilde{y}}_j$ will be combined or weighted to update the state estimate

Let the unadorned symbols represent the true value, and the $\hat{}$ symbols represent the estimated value. The error in the state estimate at t_j before and after processing the measurement set $\underline{\tilde{y}}_j$ is

$$\begin{aligned}\underline{e}_j(+) &= \underline{\hat{x}}_j(+) - \underline{x}_j = \text{error after processing } \underline{\tilde{y}}_j \\ \underline{e}_j(-) &= \underline{\hat{x}}_j(-) - \underline{x}_j = \text{error before processing } \underline{\tilde{y}}_j\end{aligned} \tag{3.2}$$

Equations (3.2) are substituted into Eq. (3.1) to obtain

$$\underline{e}_j(+) + \underline{x}_j = K'_j[\underline{e}_j(-) + \underline{x}_j] + K_j\underline{\tilde{y}}_j$$

or

$$\underline{e}_j(+) = [K'_j - I]\underline{x}_j + K'_j\underline{e}_j(-) + K_j\underline{\tilde{y}}_j \tag{3.3}$$

The measurements are assumed to be linear combinations of the states:

$$\underline{\tilde{y}}_j = H_j\underline{x}_j + \underline{V}_j \tag{3.4}$$

where $\underline{V}_j$ is a guassian white noise process of measurement errors. Substituting Eq. (3.4) into Eq. (3.3) and manipulating leads to

$$\underline{e}_j(+) = [K'_j + K_j H_j - I]\underline{x}_j + K'_j \underline{e}_j(-) + K_j \underline{V}_j \tag{3.5}$$

An unbiased estimate is desired; i.e., the expected value of $\underline{e}_j(+)$ should be zero. Taking the expected value of Eq. (3.5) yields

$$E\{\underline{e}_j(+)\} = 0 = [K'_j + K_j H_j - I]E\{\underline{x}_j\} + K'_j E\{\underline{e}_j(-)\} + K_j E\{\underline{V}_j\} \tag{3.6}$$

By definition, if $\underline{V}_j$ is white noise, $E\{\underline{V}_j\} = 0$. Generally, this requires that (1) H_k is a perfect model for the measurements (if the states are measured directly, this is trivial), and (2) the measurement errors themselves have zero mean. Now, if $E\{\underline{e}_j(-)\} = 0$, then, since in general $E\{\underline{x}_j\} \neq 0$, Eq. (3.6) leads to the conclusion that

$$K'_j = I - K_j H_j \tag{3.7}$$

The assumption that $E\{\underline{e}_j(-)\} = 0$ is not trivial. Assuming that this strategy is applied sequentially, $E\{\underline{e}_j(-)\} = 0$ will generally require the following: (1) $E\{\underline{e}_{j-1}^{(+)}\} = 0$, i.e., the expected value of the error after processing the previous measurement is zero, and (2) the model is unbiased, i.e., the model error has zero mean, so that the estimate propagates with zero expected error. For this reason, the Kalman filter process noise term is always assumed to have zero mean. However, since this term is commonly used to represent the unmodeled dynamic effects, there is no reason in general to suspect that the process noise (i.e., the model error) does indeed have zero mean. Moreover, there are many examples wherein model errors bear very little resemblance to a white noise process.

Substituting Eq. (3.7) into Eq. (3.1) and rearranging, the standard form of the linear discrete Kalman filter is obtained as

$$\hat{\underline{x}}_j(+) = \hat{\underline{x}}_j(-) + K_j[\tilde{\underline{y}}_j - H_j\hat{\underline{x}}_j(-)] \tag{3.8}$$

In Eq. (3.8), the new estimate (after processing $\tilde{\underline{y}}_j$), $\hat{\underline{x}}_j(+)$, is obtained by combining the old estimate (before processing $\tilde{\underline{y}}_j$), $\hat{\underline{x}}_j(-)$, with a correction term which is the product of K_j (the "Kalman Gain" matrix) and the residual of the actual measurement ($\tilde{\underline{y}}_j$) minus the estimated measurement ($H_j\hat{\underline{x}}_j(-)$).

The Kalman gain matrix is determined by choosing an optimality criterion based upon K_j, then optimizing it with respect to K_j. One classical criterion is to minimize the trace of the estimate error covariance matrix. This is called "minimum variance" estimation; other criteria are sometimes used (e.g., the principle of least squares, maximum likelihood estimation, and Bayesian estimation). (Remarkably, upon introducing a judicious set of auxilliary constraints, all lead to the same algorithm.)

The estimate error covariance matrix is

$$P_j(+) \equiv E\{\underline{e}_j(+)\underline{e}_j^T(+)\} = \text{covariance after processing } \tilde{\underline{y}}_j$$

Combining Eqs. (3.2) and (3.7), an expression for $\underline{e}_j(+)$ in terms of K_j is found:

$$\underline{e}_j(+) = K_j\underline{V}_j + (I - K_jH_j)\underline{e}_j(-) \tag{3.9}$$

Thus,

$$P_j(+) = E\{\underline{e}_j(+)\underline{e}_j^T(+)\}$$

$$= E\{[(I - K_jH_j)\underline{e}_j(-) + K_j\underline{V}_j][(I - K_jH_j)\underline{e}_j(-) + K_j\underline{V}_j]^T\}$$

$$= (I - K_j H_j) E\{\underline{e}_j(-)\underline{V}_j^T\} K_j^T + K_j E\{\underline{V}_j \underline{e}_j^T(-)\}(I - K_j H_j)^T$$

$$+ (I - K_j H_j) E\{\underline{e}_j(-)\underline{e}_j^T(-)\}(I - K_j H_j)^T + K_j E\{\underline{V}_j \underline{V}_j^T\} K_j^T \quad (3.10)$$

The estimate error prior to processing the t_j measurement set, $\underline{e}_j(-)$, is assumed to be uncorrelated with the t_j measurement error, $\underline{V}_j$, so that

$$E\{\underline{e}_j(-)\underline{V}_j^T\} = 0 = E\{\underline{V}_j \underline{e}_j^T(-)\} \quad (3.11)$$

The measurement error covariance at t_j and the estimate error covariance at t_j prior to processing the measurement are assumed to be known as

$$P_j(-) \equiv E\{\underline{e}_j(-)\underline{e}_j^T(-)\} = \text{known}$$
$$R_j \equiv E\{\underline{V}_j \underline{V}_j^T\} = \text{known} \quad (3.12)$$

Substituting Eqs. (3.11) and (3.12) into Eq. (3.10) yields

$$P_j(+) = (I - K_j H_j) P_j(-)(I - K_j H_j)^T + K_j R_j K_j^T \quad (3.13)$$

The optimality criterion is to minimize the trace of $P_j(+)$. Using the linear algebra theorem,

$$\frac{\partial}{\partial A} [\text{trace } (ABA^T)] = 2AB \quad (B \text{ symmetric})$$

the gradient of trace $P_j(+)$ with respect to K_j is

$$\frac{\partial}{\partial K_j} [\text{trace } P_j(+)] = -2(I - K_j H_j) P_j(-) H_j^T + 2K_j R_j \quad (3.14)$$

The necessary condition for an extremum of trace $P_j(+)$ is that the gradient in Eq. (3.14) be zero; imposing this condition, Eq. (3.14) may be rewritten

$$K_j R_j = (I - K_j H_j) P_j(-) H_j^T$$

from which

$$K_j = P_j(-)H_j^T(R_j + H_jP_j(-)H_j^T)^{-1} \tag{3.15}$$

Equation (3.15) is the classical form of the Kalman gain matrix.

Substituting Eq. (3.15) into Eq. (3.13) and manipulating leads to the covariance update equation at measurement time t_j,

$$P_j(+) = [I - K_jH_j]P_j(-) \tag{3.16}$$

Thus, the linear discrete Kalman filter for processing the measurement set $\underline{\tilde{y}}(t_j)$ (modeled by Eq. (3.4)) consists of Eqs. (3.8), (3.15), and (3.16).

Note that $P_j(+)$ in Eq. (3.16) is less than $P_j(-)$. If the system is stable, then $P_j(-)$ is of approximately the same magnitude as $P_{j-1}(+)$. Thus, as j increases, $P_j(+)$ decreases, and thus K_j decreases, so that the new measurements are multiplied by smaller and smaller gain matrices. Eventually, the new measurement information is effectively ignored.

The determination of $\underline{\hat{x}}_j(-)$ and $P_j(-)$ from $\underline{\hat{x}}_{j-1}(+)$ and $P_{j-1}(+)$ proceeds as follows. For the propagation of the state vector, the system model is integrated from t_{j-1} to t_j using $\underline{\hat{x}}_{j-1}(+)$ as the initial condition. For the linear system modeled by

$$\dot{\underline{x}} = A(t)\underline{x}(t) + B(t)\underline{u}(t) \tag{3.17}$$

the estimate $\underline{\hat{x}}_j(-)$ is calculated as

$$\underline{\hat{x}}_j(-) = \Phi(t_j,t_{j-1})\underline{\hat{x}}_{j-1}(+) + \int_{t_{j-1}}^{t_j} \Phi(t,\tau)B(\tau)\underline{u}(\tau)d\tau \tag{3.18}$$

where $\Phi(t_j,t_{j-1})$ is the state transition matrix calculated by integrating

$$\dot{\Phi}(t,t_{j-1}) = A(t)\Phi(t,t_{j-1}) \tag{3.19}$$

from the initial condition

$$\Phi(t_{j-1},t_{j-1}) = I$$

In Eq. (3.17), the forcing term $B(t)\underline{u}(t)$ is assumed to model the known forces. The unknown effects are called process noise. Although it is possible to include process noise in the propagation of the state vector if the process noise can be described statistically (i.e., the mean and covariance of the process noise is known, as well as its probability distribution), in practice this is rarely (if ever) done (see, e.g., Gelb[7]). Instead, process noise terms are included in the propagation of the estimate error covariance matrix, as will be shown next.

The estimate covariance $P_{j-1}(+)$ is propagated forward to time t_j using linear error theory. For a general linear system modeled by

$$\underline{y} = M\underline{z} \tag{3.20}$$

where

$$E\{\underline{z}\} = \bar{\underline{z}} \quad , \quad E\{[\underline{z} - \bar{\underline{z}}][\underline{z} - \bar{\underline{z}}]^T\} = P_{zz}$$

it can be shown (see, e.g., Junkins[6]) that the covariance of $\underline{y}$ is

$$P\{[\underline{y} - \bar{\underline{y}}][\underline{y} - \bar{\underline{y}}]^T\} = MP_{zz}M^T \tag{3.21}$$

In Eq. (3.17), the term $B(t)\underline{u}(t)$ accounts for <u>known</u> forcing terms and thus contributes nothing to the uncertainty in $\hat{\underline{x}}$. Comparing Eq. (3.18) with Eq. (3.20), and utilizing (3.21), the estimate covariance propagates as

$$P_j(-) = \Phi(t_j,t_{j-1})P_{j-1}(+)\Phi^T(t_j,t_{j-1}) \tag{3.22}$$

For a stable system, i.e., a system with approximately constant or decreasing magnitude of the states, the state transition matrix will have a magnitude (e.g., maximum eigenvalue) of less than unity. Thus,

$P_j(-)$ will be of approximately the same magnitude or less than $P_{j-1}(+)$. Considering the measurement processing update equation, Eq. (3.16), clearly as j increases, $P_j(+)$ tends to zero and thus any new measurement is ignored. For this reason, an extra term is often added to the covariance propagation, Eq. (3.22), in order to increase the magnitude of $P_j(-)$ so that the next measurement is not ignored:

$$P_j(-) = \Phi(t_j,t_{j-1})P_{j-1}(+)\Phi^T(t_j,t_{j-1}) + Q_{j-1} \tag{3.23}$$

The term Q_{j-1} is the integrated effect of process noise upon the state covariance matrix (note the units are covariance) and is usually used as an artistic device to keep the Kalman filter from ignoring new measurements. An analytic expression can be derived for Q_{j-1} in the unlikely event that the model errors are strictly additive random effects in the form

$$\dot{\underline{x}} = A(t)\underline{x}(t) + B(t)\underline{u}(t) + L(t)\underline{w}(t) \tag{3.24}$$

where L(t) is a known matrix and $\underline{w}(t)$ is a Gaussian white noise process of known mean and covariance. It is trivial to note that if L(t) and the statistics of $\underline{w}(t)$ are known, then the resulting algorithm cannot rigorously be used for more general effects (e.g., non-random, non-zero mean). In practice, L(t) and $\underline{w}(t)$ are unknown and process noise is added as in Eq. (3.23). Thus, the Kalman filter algorithms are routinely and artistically applied to problems with quite general unmodeled effects, for which the underlying theory is not rigorously applicable.

In summary, the discrete Kalman filter consists of the two "propagation between measurement" equations, (3.18) and (3.23), and the

four "measurement processing" equations, (3.4), (3.8), (3.15), and (3.16).

3.3 Extended Kalman Filter

The extended Kalman filter is an adaptation of the discrete Kalman filter of Section 3.2 to nonlinear system and measurement models (cf., e.g., Jazwinski[44] (1970) or Sage and Melsa[45] (1970)). In this section, the linear system model Eq. (3.17) and the linear measurement model Eq. (3.4) are replaced by the nonlinear models

$$\dot{\underline{x}} = \underline{f}(\underline{x}(t),\underline{u}(t),t) \tag{3.25}$$

$$\tilde{\underline{y}}(t_j) = \underline{g}(\underline{x}(t_j),t_j) + \underline{v}_j \tag{3.26}$$

where the unadorned $\underline{x}(t)$ represents the true state vector. The extended Kalman filter state propagation is obtained by linearizing Eq. (3.25) in a Taylor Series expansion about the estimated state vector, $\hat{\underline{x}}(t)$, as

$$\dot{\hat{\underline{x}}} = \underline{f}(\hat{\underline{x}}(t),\underline{u}(t),t) \tag{3.27}$$

so that the state vector is propagated between measurements as

$$\hat{\underline{x}}_j(-) = \hat{\underline{x}}_{j-1}(+) + \int_{t_{j-1}}^{t_j} \underline{f}[\hat{\underline{x}}(\tau),\underline{u}(\tau),\tau]d\tau \tag{3.28}$$

The estimate covariance matrix is propagated using Eq. (3.23) as in the linear Kalman filter, except the state transition matrix is obtained for the "linear departure motion" (see, e.g., Junkins[6]) as

$$\dot{\Phi}(t,t_{j-1}) = \left.\frac{\partial \underline{f}}{\partial \underline{x}}\right|_{\hat{\underline{x}}(t),\underline{u}(t),t} \Phi(t,t_{j-1}) \tag{3.29}$$

Following a similar development as in Section 3.2, the measurement processing equations are obtained as

$$\hat{\underline{x}}_j(+) = \hat{\underline{x}}_j(-) + K_j[\tilde{\underline{y}}_j - \underline{g}(\hat{\underline{x}}_j(-),t_j)] \tag{3.30}$$

where

$$K_j = P_j(-)H_j^T[H_jP_j(-)H_j^T + R_j]^{-1} \tag{3.31}$$

$$P_j(+) = [I - K_jH_j]P_j(-) \tag{3.32}$$

$$H_j \equiv \left.\frac{\partial \underline{g}}{\partial \underline{x}}\right|_{\hat{\underline{x}}_j(-),t_j} \tag{3.33}$$

Note that the estimated measurement value, $\underline{g}(\hat{\underline{x}}_j(-),t_j)$, is a linear Taylor Series approximation about $\hat{\underline{x}}_j(-)$ for the actual value, $\underline{g}(\underline{x}(t_j),t_j)$.

Equations (3.25)-(3.33) constitute the extended Kalman filter.

3.4 The Rauch-Tung-Striebel[46] (RTS) Filter-Smoother

The state vector estimates at any time t obtained using a Kalman filter algorithm are based only upon the measurements available up through the time t. This is certainly a requirement for real-time estimation. However, if real-time estimates are not required, the Kalman filter estimates may generally be improved by including measurements taken after the time at which the estimate is desired. This process is commonly referred to as "smoothing".

There are many versions of smoothing currently in use. Some reprocess the actual original measurements; some reprocess the filter estimates but do not actually reconsider the original measurements. The simplest version of smoothing is simply to integrate backwards from the final estimate using the system model. Although this "smoothed" estimate is a smooth (continuous) estimate, if the model is imperfect the smoothed estimate may not be an improvement. By reprocessing the

original measurements or estimates, the smoothed estimate is usually an improvement over the original estimate. However, note that at every measurement or filter estimate processing time, a jump discontinuity is introduced as in the Kalman filter. Thus, although the "smoothed" estimate generally has smaller jumps than the original estimate, it is still not a "smooth" (continuous) estimate.

The filter-smoother algorithm which is presented in the remainder of this section was developed by Rauch, Tung, and Striebel[44] (1961) and is generally referred to as the "RTS filter-smoother". It is presented because of its popularity in attitude estimation applications. The version originally reported dealt with linear system models and linear measurement models. The filter portion is identical to a discrete extended Kalman filter as outlined in the previous section. The smoother equation is

$$\hat{\underline{x}}_{K/N} = \hat{\underline{x}}_K(+) + C_k[\hat{\underline{x}}_{K+1/N} - \hat{\underline{x}}_{K+1}(-)] \tag{3.33}$$

where

$\hat{\underline{x}}_{K/N}$ = "smoothed" estimate at t_K

$\hat{\underline{x}}_K(+)$ = filter estimate at t_K after processing $\tilde{\underline{y}}_K$

C_K = "Gain" = $P_K(+)\Phi^T(t_{K+1},t_K)P^T_{K+1}(-)$

Equation (3.33) is applied sequentially backwards from the final measurement time, t_N. The filter estimates, $\hat{\underline{x}}_K(+)$, should be stored from the filter solution. The covariance estimates, $P_K(+)$ and $P_{K+1}(-)$, may be stored from the filter estimates or recalculated using

$$P_K(-) = P_K(+) - P_K(+)H_K^T[H_KP_K(+)H_K^T - R_K]^{-1}H_KP_K(+) \tag{3.34}$$

$$P_{K-1}(+) = \Phi(t_K,t_{K-1})[P_K(-) - Q_{K-1}]\Phi^T(t_K,t_{K-1}) \tag{3.35}$$

where Q_{K-1} is the "process noise" from the filter estimate, and H_K and $\phi(t_K,t_{K-1})$ are given by Eqs. (3.4) and (3.19) for the linear case and by Eqs. (3.32) and (3.28) for the nonlinear case. The smoother covariance estimate is obtained as

$$P_{K/N} = P_K(+) + C_K[P_{K+1/N} - P_K(+)]C_K^T \tag{3.36}$$

where

$P_{K/N}$ ≡ smoother estimate of the covariance matrix at time t_K

Equation (3.36) is applied sequentially backwards through the measurement times along with Eq. (3.33).

3.5 Simple Scalar Example

The extended Kalman filter-smoother (EKFS) algorithm developed in this chapter is now applied to a scalar example. The example has been chosen to highlight the effects of using process noise to account for model errors. The true state is given by

$$x(t) = \frac{1}{2}\{[t^2 - 4(\frac{t^3}{3} - x_o^2)]^{1/2} - t\} \tag{3.37}$$

so that the exact dynamic state model is

$$\dot{x} = -\frac{x + t^2}{2x + t} \tag{3.38}$$

The states are measured directly as

$$\tilde{x}_k = x_k + V_k \tag{3.39}$$

where V_k is a zero-mean gaussian process with known variance σ^2. The assumed state model is

$$\dot{x} = -\frac{t^2}{2x + t} \tag{3.40}$$

so that the "unmodeled effect" is

$$d(t) = - \frac{x}{2x + t} \tag{3.41}$$

For demonstration purposes, the value of σ^2 is 0.59 (nominally, 1). The state history estimates obtained for various values of process noise (Q) are shown plotted with the true state history and the measurements in Fig. (3.1). In Fig. (3.1-a), no process noise is added. The process noise in Fig. (3.1-b), (3.1-c), (3.1-d), (3.1-e), and (3.1-f) is $Q = 10^{-2}$, 10^{-1}, 10^{0}, 10^{1}, and 10^{2}, respectively.

When no process noise is added, the smoothed EKFS estimate is actually smooth (i.e., continuous) as shown in Fig. (3.1-a). However, when process noise is added, the estimated state history contains jump discontinuities at the measurement times. These jumps are small in Fig. (3.1-b), where $Q = 10^{-2}$; however, for larger Q, the jumps are larger. This is due to the fact that process noise forces the estimate to be based more heavily on the measurements, which in this case are imperfect. For very large Q, the measurements are matched exactly; however, the resulting estimate is not particularly accurate (Fig. (3.1-f)).

For this sample of measurements, the minimum error variance estimate is obtained for $Q \approx 1$, Fig. (3.1-d). This is also the estimate for which the measurement-minus-estimate variance is approximately equal to the measurement-minus-truth variance; thus, in the usual case where the truth is unknown, this is the EKFS "optimal" estimate. In Fig. (3.2), the estimate obtained by the present method is shown. Comparing Figs. (3.1-d) and (3.2), the following observations are noted:

(1) The EKFS estimate contains jump discontinuities, some of which are quite large; the present method estimate does not.

(2) The EKFS estimate shape between measurements is fixed by Eq. (3.40); thus, in regions where the unmodeled effect term is relatively large (low t; see Eq. (3.41)), the shape between measurements is incorrect. The addition of process noise does not improve the model.

(3) Since the EKFS estimate is obtained by shifting emphasis from the model to the measurements, it is more sensitive to measurement anomalies. Between t = 5.5 and 6.5, three positive measurement errors in a row are present. Comparing the EKFS estimate error with the present method estimate error in this region, clearly the EKFS error is considerably larger.

In Chapters 4 and 5, a number of test cases are utilized to compare the EKFS performance with the present method.

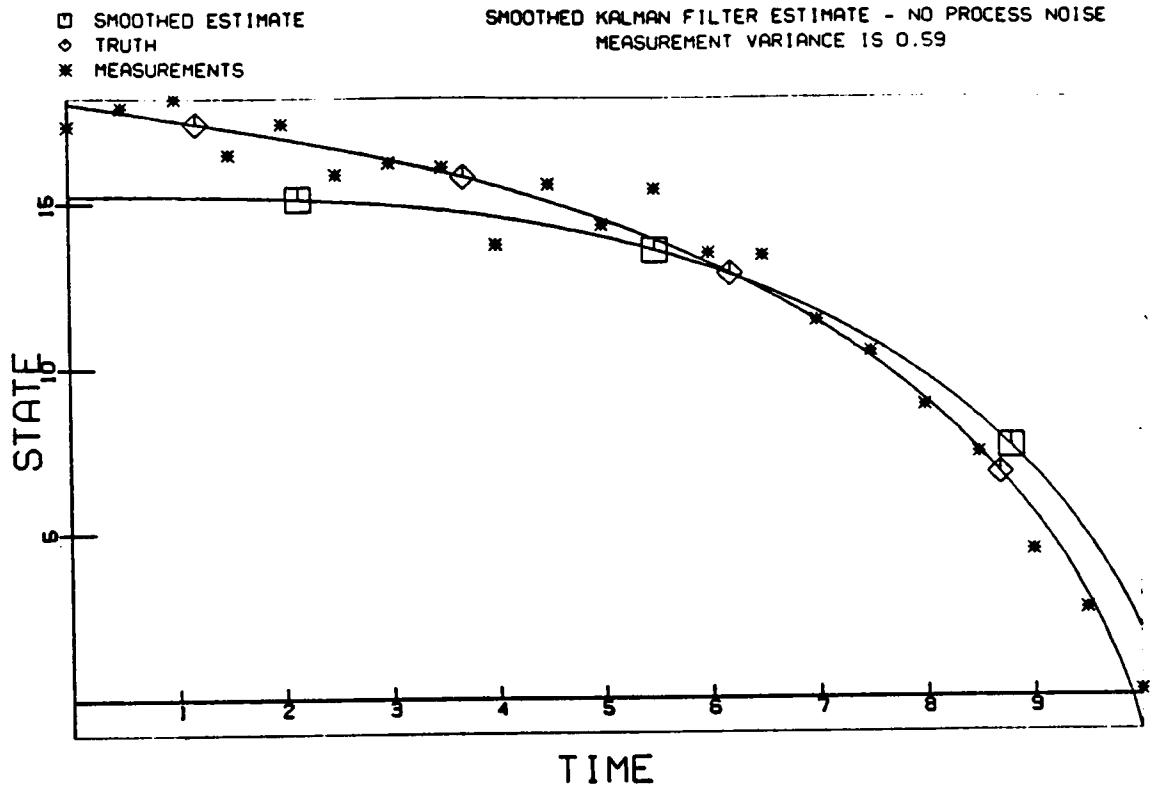


Figure 3.1-a

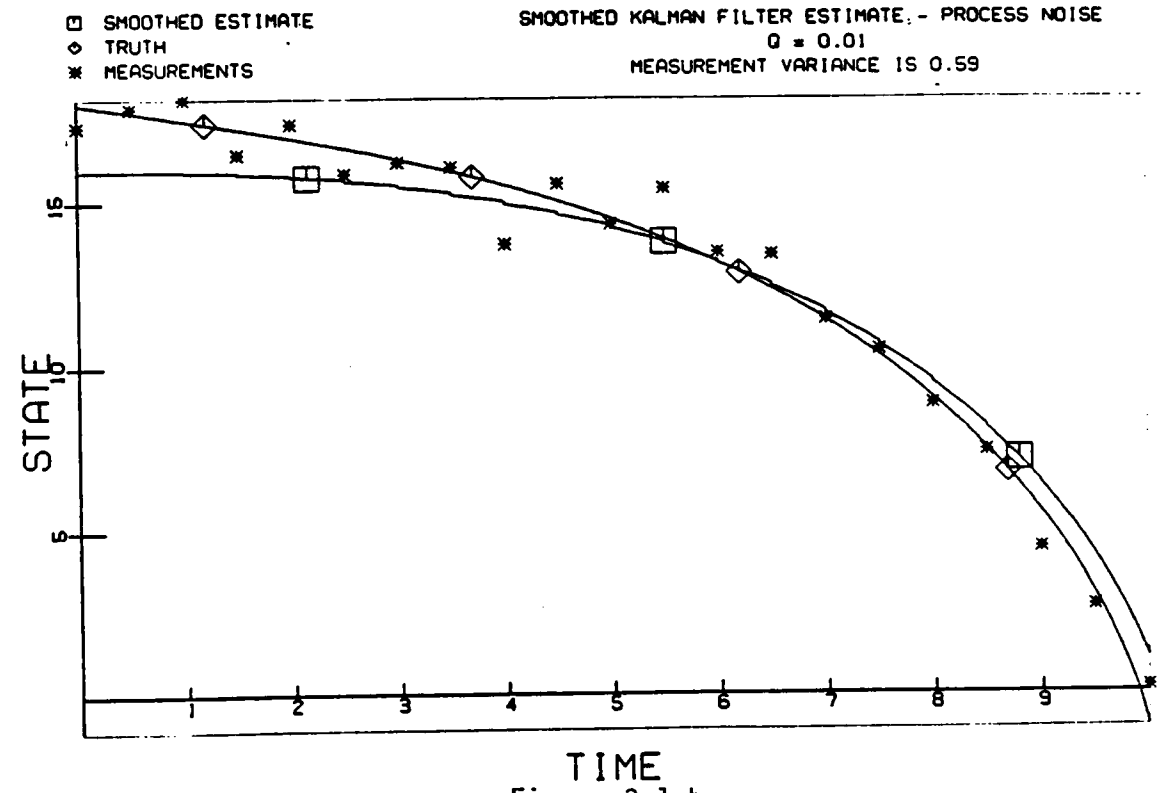


Figure 3.1-b

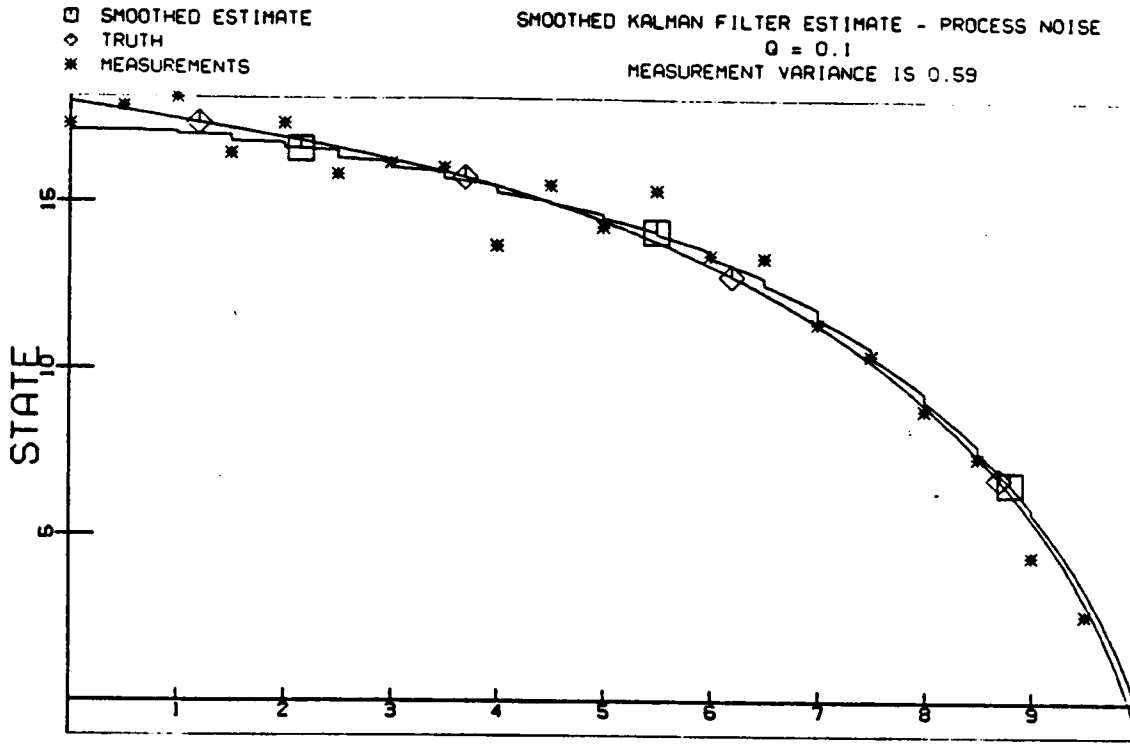


Figure 3.1-c

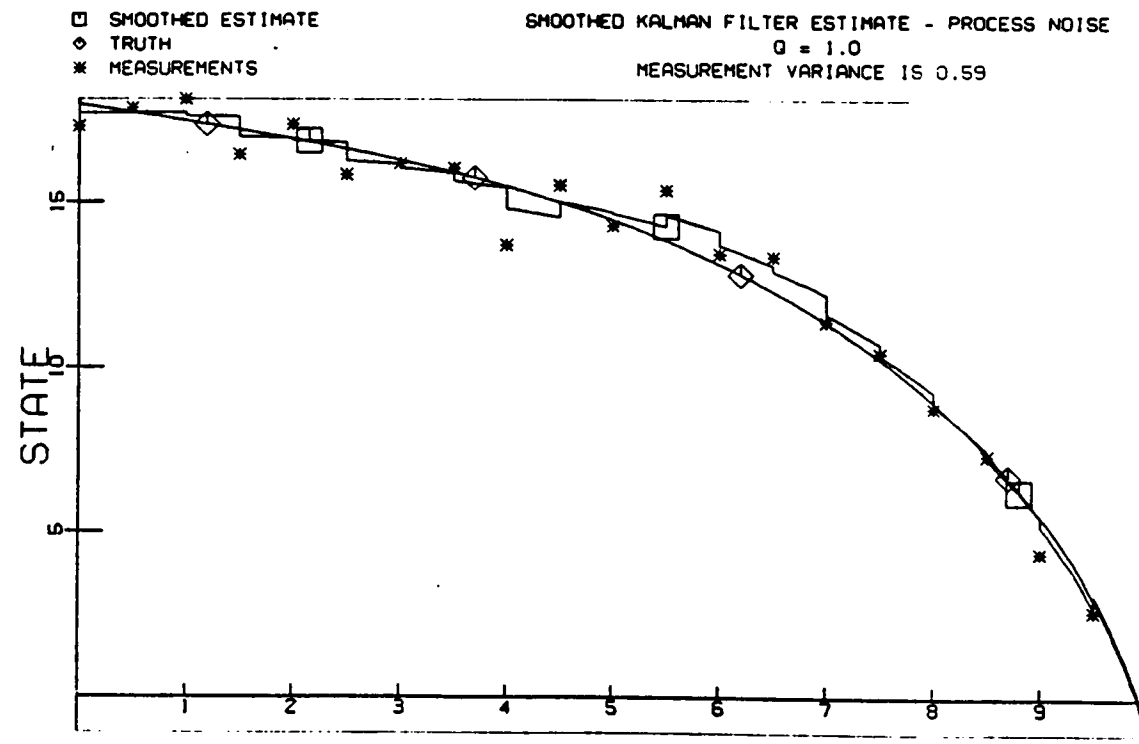


Figure 3.1-d

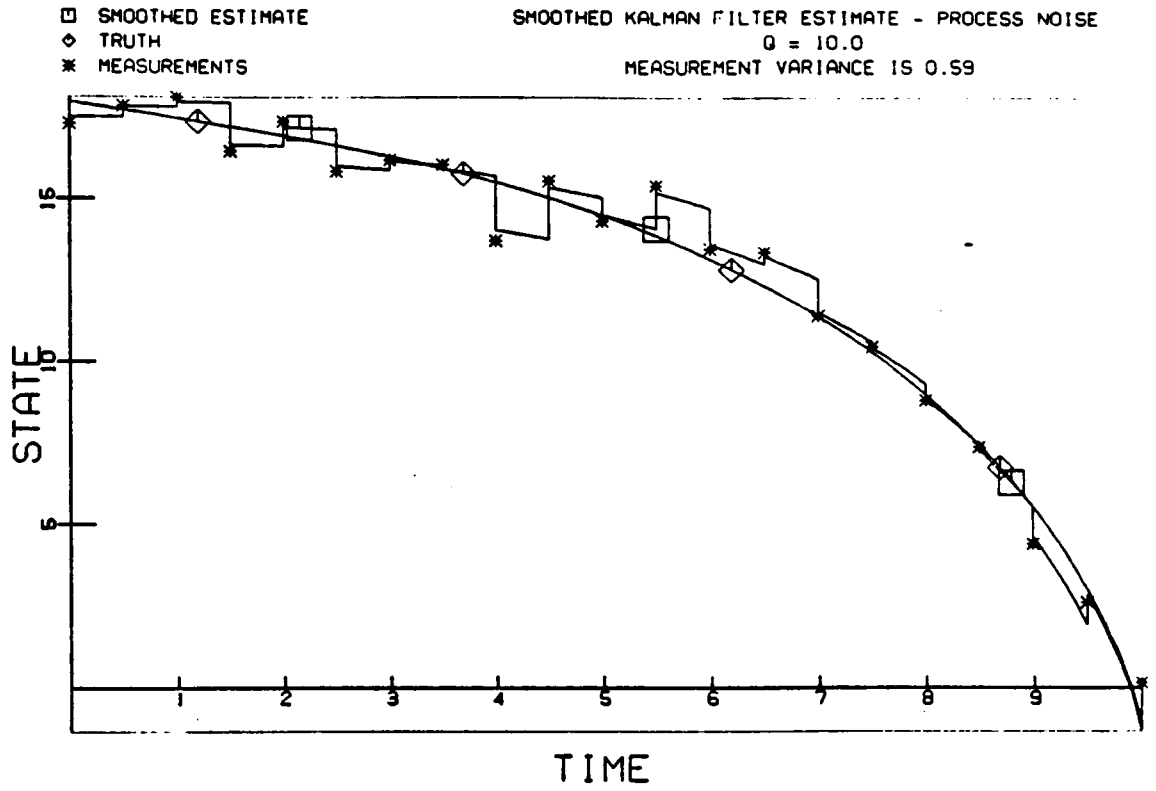


Figure 3.1-e

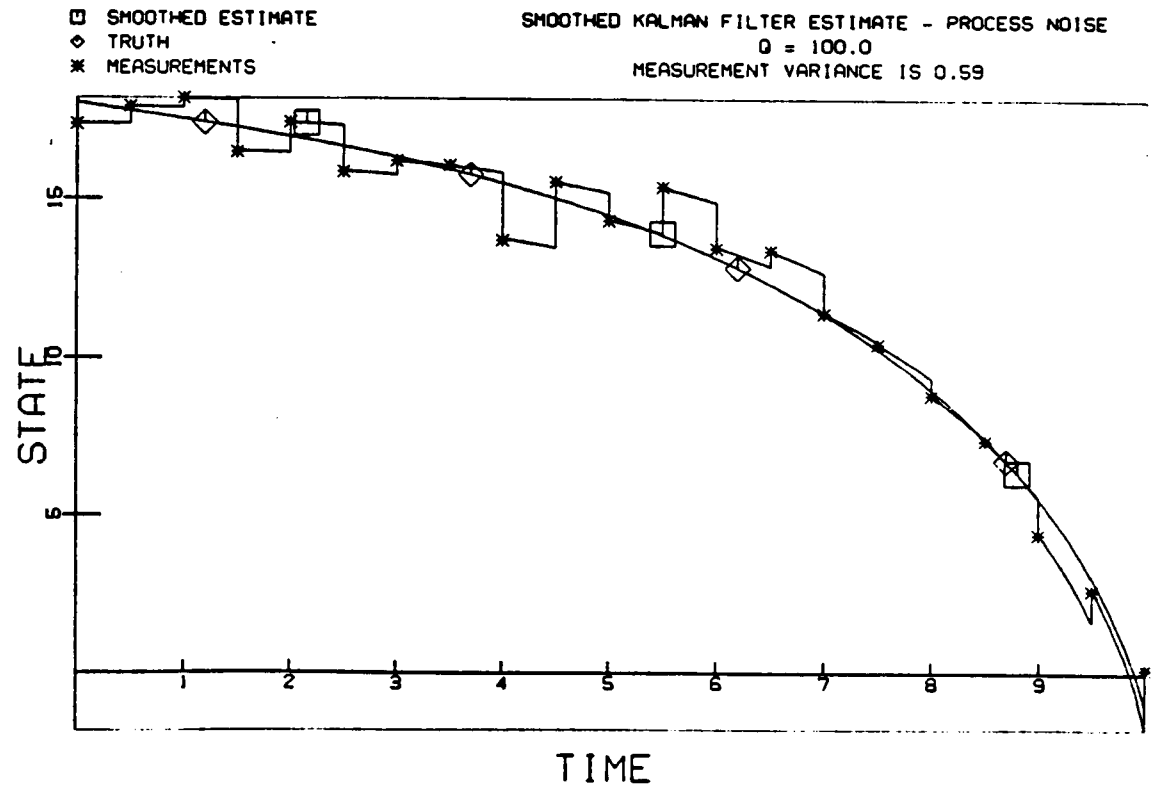


Figure 3.1-f

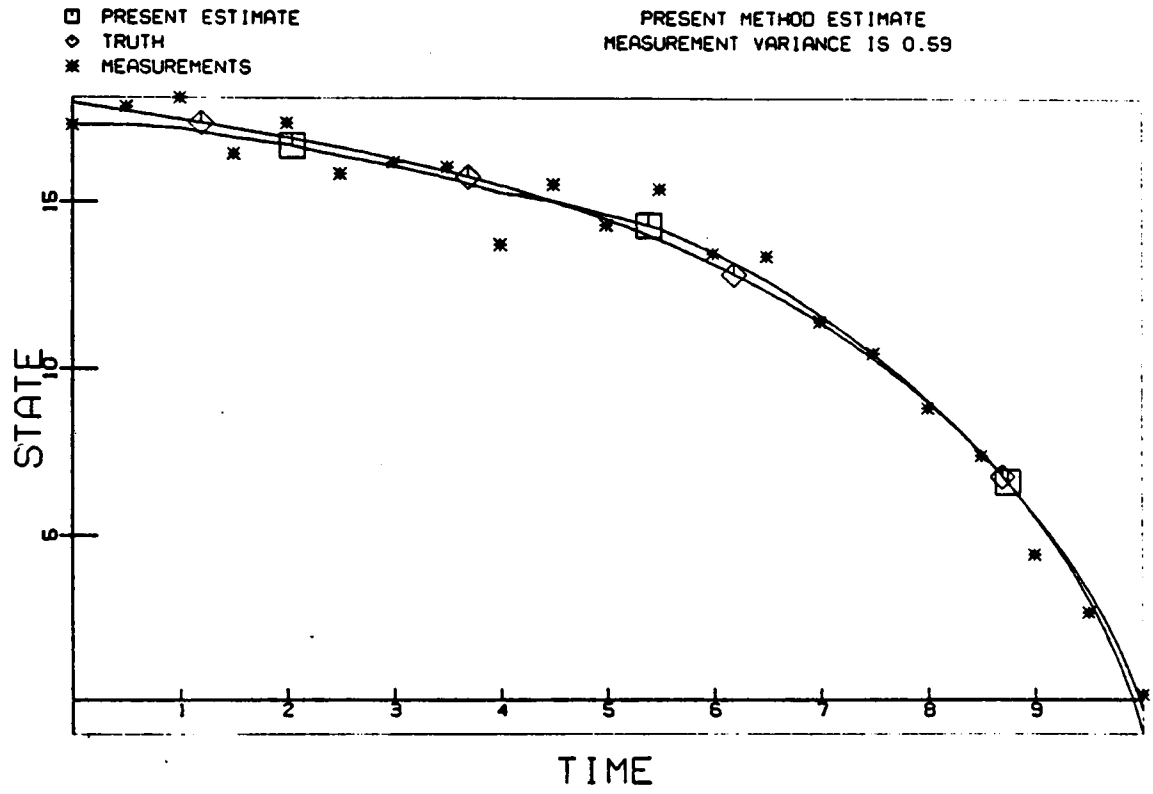

PRESENT ESTIMATE
TRUTH
MEASUREMENTS
PRESENT METHOD ESTIMATE
MEASUREMENT VARIANCE IS 0.59
STATE
15
10
5
1
2
3
4
5
6
7
8
9
TIME

Figure 3.2

CHAPTER 4

SCALAR EXAMPLES

4.1 Introduction

In this chapter, the estimates obtained using the present method are compared with the estimates obtained using the EKFS for some scalar examples. The sample problems will include cases where no model is assumed as well as cases where a near-perfect model is assumed; relatively high measurement frequency as well as relatively low measurement frequency; and perfect measurements as well as relatively poor measurements. To aid in the comparison, all problems were solved for $0 \leq t \leq 10$, with measurements at $t = 0$ and $t = 10$ in addition to the interior measurement times.

In all cases, the underlying truth is known so that the actual variances can be calculated for comparison purposes. However, the criterion used for determining the appropriate level of noise (EKFS) or unmodeled effect weight (present method) is not based on knowledge of the truth. The criterion was taken as the "covariance constraint", i.e., the measurement-minus-estimate variance must approximately equal the measurement-minus-truth variance. When the appropriate level of process noise (weight) was determined, the solution was considered optimal.

There are two underlying test problems; the test cases consist of solving these problems for various combinations of measurement frequency and measurement accuracy. These two problems are presented individually in Sections 4.2 and 4.3.

4.2 Test Problem 1

Here, the truth is the same as the one used in Section 3.5, specifically,

$$x(t) = \frac{1}{2}\{[t^2 - 4(\frac{t^3}{3} - x_o^2)]^{1/2} - t\} \tag{4.2.1}$$

Thus, the true state model is

$$\dot{x} = -\frac{x + t^2}{2x + t} \tag{4.2.2}$$

In this section, no model is assumed. The EKFS state model is

$$\dot{x} = 0 \tag{4.2.3}$$

while the present method state model and co-state governing equation is

$$\begin{aligned} \dot{x} &= u = -\frac{\lambda}{2w} \\ \dot{\lambda} &= 0 \end{aligned} \tag{4.2.4}$$

Clearly, the EKFS solution will consist of a sequence of constant values for x between measurements, and jump discontinuities at the measurements; by comparison, the present method solution will consist of a sequence of connected straight lines, with only the slopes discontinuous at the measurements.

4.2.1 Six Measurements

The problem was solved with two different measurement sets of six measurements each, spaced two time units apart. Results are shown in Figs. (4.1)-(4.3), with part (a) of each figure showing results for a measurement variance of 0.383, and part (b) showing results for a measurement variance of ~0. Figure (4.1) shows the measurements, true state history, EKFS estimate, and present method estimate. The present

method variance is at least an order of magnitude lower than the EKFS variance for both measurement sets. Note that the EKFS solution for the perfect measurement case is actually worse than the EKFS solution for the imperfect measurement case. This is due to the fact that the process noise approach assumes zero mean model error; however, the model error in this case is not zero mean, but always negative (see Eq. (4.2.2) and (4.2.3)). When measurement errors are present, the EKFS estimate is occasionally "pulled" below the true state by a negative measurement error (e.g., t = 6 in Fig. (4.1-a)); thus, the error passes back and forth through zero, as shown in the residuals in Fig. (4.2-a). However, when the measurements are "perfect", the EKFS solution never has negative error, so the positive error between measurements grows to larger magnitudes, as shown by comparing the residuals in Figs. (4.2-a) and (4.2-b).

By comparison, the present method variance is reduced by approximately 1/3 (0.464 vs. 0.312) when perfect measurements are used.

The demonstrated result that the EKFS solution may actually be degraded when measurements of better accuracy are available is disconcerting; certainly, a user would like to believe that regardless of the model used for obtaining the estimate, more accurate measurements will lead to more accurate estimates. However, if the model error is not zero mean, this is not guaranteed for the EKFS.

Figure (4.3) shows the estimated unmodeled effect (disturbance) obtained by the present method. Since $\dot{\lambda} = 0$ (Eq. (4.2.4)), the disturbance estimate is a sequence of constant value line segments. Although these line segments follow the actual disturbance, especially

for perfect measurements (Fig. (4.3-b)), the measurements are spaced too far apart for an accurate recovery.

4.2.2 Eleven Measurements

The identical problem was solved twice more, with two eleven measurement sets of variances 0.532 and 0, respectively. Thus, the measurement frequency has been doubled to once per time unit. Results are given in Figs. (4.4)-(4.6); part (a) contains the imperfect measurement results, and part (b) contains the perfect measurement results. The estimate state histories are shown in Fig. (4.4); the residuals are shown in Fig. (4.5); and the present method estimate of the disturbance is shown in Fig. (4.6).

Again, the present method performs considerably better than the EKFS, although the difference has decreased somewhat due to the fact that the EKFS solution is much more dependent on measurement frequency. However, the perfect measurement EKFS solution is still worse than the imperfect measurement EKFS solution, whereas the perfect measurement present method solution is considerably better than the imperfect measurement present method solution. This is clearly desirable.

The disturbance recovery in Fig. (4.6) is better than the recovery for six measurements in Fig. (4.3); however, it is still hampered by relatively low measurement frequency, and the fact that it is simply a sequence of horizontal straight line segments.

4.2.3 Twenty-one Measurements

The identical problem was solved twice more, with the measurement frequency doubled again to two measurements per unit of time. These two twenty-one measurement sets have variances of 0.590 and 0, respectively. Results are shown in Figs.(4.7)-(4.9); part (a) contains imperfect measurement results, and part (b) contains the perfect measurement results. The estimated state histories are shown in Fig. (4.7), the residuals in Fig. (4.8), and the present method estimate of the disturbance in Fig. (4.9).

Again, the present method estimate is considerably better than the EKFS estimate, especially in the perfect measurement case, as shown in Fig. (4.7). The variance of the present method in the perfect measurement case is 0.014, compared with 0.562 for the EKFS. Note, however, that the perfect measurement EKFS has lower variance than the imperfect measurement EKFS, unlike the results from Sections 4.2.1 and 4.2.2.

The estimated disturbance is considerably more accurate for the increased measurement frequency. However, comparing Fig. (4.9-a) with Fig.(4.9-b), the disturbance estimate is seen to suffer considerably from the presence of the measurement errors.

4.2.4 101 Measurements

As a final variation of this test problem, the measurement frequency was increased to ten measurements per unit of time, resulting in 101 measurements. Again, two sets were used; one has variance 0.872, the other has variance 0. Results are shown in Figs. (4.10)-(4.12).

Part (a) shows the results for imperfect measurements, and part (b) shows the results for perfect measurements. Figure (4.10) shows the estimated state histories, Fig. (4.11) shows the residuals, and Fig. (4.12) shows the estimated disturbance.

Again, the present method estimate has a lower variance than the EKFS solution in both cases. The difference, however, is much smaller than is for fewer measurements, reflecting the dependence of the EKFS on measurement frequency when the model is poor.

The disturbance estimate from the present method is now very good for the perfect measurement case; however, it is noisy for the imperfect measurement case (see Fig.(4-12)).

4.2.5 Summary of Test Problem 1

The model error for this problem was not zero mean, and thus served to illustrate some of the drawbacks of the process noise approach of the Kalman filter-type algorithms.

The results are summarized in Table 4.1, which shows the estimate error variances for both the EKFS and the present method for all of the results presented in this section. Clearly, the EKFS solution is much more dependent on the measurement frequency. In addition, for low measurement frequency, the EKFS gave a better solution for imperfect measurements than it did for perfect measurements. This is clearly undesirable.

It is interesting to note that the EKFS solution accuracy for twenty-one measurements is not as good as the present method accuracy for six measurements, corresponding to a measurement frequency ratio of

four to one. This is explained by the fact that the EKFS model is not improved by the addition of process noise; it remains a constant between measurements. The approach of the present method leads to straight line estimates between measurements.

It should be noted that in the limit of continuous, perfect measurements, the EKFS gives a perfect solution, because process noise can be used to force the estimate to match the measurements exactly. This generally cannot be done with the present method, as explained in the next paragraph. The frequency of perfect measurements for which the accuracy of the EKFS exceeds that of the present method is not known; however, it is obviously higher than ten per time unit.

Finally, the perfect measurement solutions of the present method for twenty-one and 101 measurements have approximately identical variances. This reflects a limit on the accuracy of the present method. The present method fit is a series of line segments whose slopes are proportional to the co-state. Considering the curvature of the true state trajectory, this slope must change value. However, the co-state only changes value at a measurement, if the estimate does not match the measurement (see Eq. (2.35)). Thus, if one measurement is matched perfectly, the next one cannot be matched perfectly since the slope (co-state) hasn't changed. The estimate must contain nonzero error in order to produce a change in the co-state, which must change in order to correct the estimate, etc. Consequently, the present method cannot obtain a perfect estimate.

4.3 Test Problem 2

In this section, a more "realistic" test problem is considered. The "true state" is given by

$$x(t) = 10\, e^{-\omega t}(\sin t + \cos 5t) \tag{4.3.1}$$

Thus, the true state model is

$$\dot{x} = -\omega x + 10\, e^{-\omega t}(\cos t - 5 \sin 5t) \tag{4.3.2}$$

The assumed model is

$$\dot{x} = 10\, e^{-\omega t}(\cos t - 5 \sin 5t) \tag{4.3.3}$$

Thus, the unmodeled effect is

$$d = -\omega x \tag{4.3.4}$$

In all cases, a value of $\omega = 1/3$ is used.

This problem is more realistic since the true state tends to zero with zero slope, reflecting many physical systems. Moreover, the assumed model contains all of the qualitative features of the truth: exponential decay, plus low frequency and high frequency components with the correct frequencies. The presence of model error is not at all obvious, unlike the previous example. Also, the model error is approximately zero mean, averaged over time.

4.3.1 Six Measurements

Two sets of six measurements each were used, one with a variance of 0.383, the other with a variance of 0. The results are shown in Figs. (4.13)-(4.15); part (a) in each figure is from the imperfect measurement solution, and part (b) is from the perfect measurement solution. Figure (4.13) shows the estimated state histories; Fig. (4.14) shows the residuals; and Fig. (4.15) shows the estimated disturbance.

Again, the EKFS solution for imperfect measurements is actually better than for perfect measurements, as shown in Fig. (4.13). The present method solution is better than the EKFS solution in both cases, and the present method solution is better for perfect measurements than it is for imperfect measurements. This is desirable.

Figure (4.14) highlights the accumulation of error in the EKFS solution between measurements. The residuals grow from one measurement to the next due to the model error; at the next measurement, the EKFS gain causes a jump discontinuity, usually reducing the error. Including process noise forces the state closer to the measurement at the measurement times; however, for imperfect measurements, this may not improve the estimate. At t = 6 in Fig. (4.14-a), for instance, the jump discontinuity is toward a larger error. In any event, process noise does not improve the model; when the measurements are infrequent, the model error may accumulate to relatively large values.

Note that the present method residuals are continuous, and generally smaller than the EKFS residuals. This is due to the improvement in the model. However, Fig. (4.15) demonstrates that the estimate of the disturbance is very poor for this example. Since the assumed model (Eq. 4.3.3) does not include the state specifically on the right hand side, the governing equation for the co-state is

$$\dot{\lambda} = - \frac{\partial f}{\partial x} \lambda = 0 \qquad (4.3.5)$$

Clearly, a series of constant value line segments of length two time units will not reconstruct the busy disturbance shown in Fig. (4.15). Nevertheless, the addition of a constant to the right hand side of Eq.

(4.3.3) (due to the disturbance term in the present method) has significantly improved the estimate.

4.3.2 Eleven Measurements

Two sets of eleven measurements each, one set perfect, the other with variance .701, were used for the problem of this section. Results are shown in Figs. (4.16)-(4.18); part (a) in each figure shows results using the imperfect measurement set, and part (b) shows results using the perfect measurement set. Figure (4.16) shows the estimated state histories, Fig. (4.17) shows the residuals, and Fig. (4.18) shows the present method estimate of the unmodeled disturbance.

The EKFS solution is improved when the perfect measurements are used, unlike the previous section. However, the present method solution is still better than the EKFS solution for both measurement sets, and shows an order of magnitude improvement when the perfect measurement set is used. In Fig. (4.17), the accumulation of error between measurements is once again readily apparent in the EKFS residuals.

The disturbance estimate of the present method is not good; however, the presence of the disturbance term has significantly improved the present method estimate over the EKFS estimate.

4.3.3 Twenty-one Measurements

Figures (4.19)-(4.21) show results for two twenty-one measurement sets; part (a) shows the results for a measurement error variance of 0.590, and part (b) shows the results for a measurement error variance of 0. Figure (4.19) gives the estimated state histories, Fig. (4.20)

gives the estimate residuals, and Fig. (4.21) gives the present method estimate of the disturbance.

Once more, the present method is able to obtain more accurate estimates than the EKFS for both perfect and imperfect measurement sets. Although the present method estimate of the disturbance is poor (Fig. (4.21)), the presence of the disturbance term significantly improves the estimate. Again, the accumulation of errors between measurements in the EKFS solution is apparent in Fig. (4.20); note, however, that with the increased measurement frequency, the error does not accumulate to as high a value before being corrected at a measurement. Thus, the EKFS solution for twenty-one measurements is significantly better than for six or eleven measurements.

4.3.4 101 Measurements

Finally, two cases are considered with very high measurement frequency, ten measurements per unit time. Again, perfect and imperfect measurement sets are used. Results are shown in Figs. (4.22-4.24). In part (a), the measurement variance is 0.876; in part (b), it is zero. Figure (4.22) shows the estimated state histories, Fig. (4.23) shows the residuals, and Fig. (4.24) shows the present method estimate of the disturbance.

The present method estimate is better when the imperfect measurements are used; however, the EKFS solution is better when the perfect measurements are used. This is a reflection of the fact that the EKFS solution may be forced to match the measurements perfectly; if the measurements themselves are perfect and "dense enough", the EKFS

solution is superior to that of the present model. As explained in Section 4.2.5, the present method generally cannot be forced to match every measurement perfectly, since the disturbance term (co-state) must change value, and this can only happen when the estimate does not match the measurement.

In Fig. (4.23), the between-measurement error accumulation in the EKFS is minimal for this dense measurement case. When the measurements are perfect and the error does not have enough time to grow large, the EKFS method obtains a better solution than the present method.

Finally, in Fig. (4.24), the present method estimate of the unmodeled disturbance is seen to be poor. Apparently, the disturbance is simply too busy, and the measurement frequency still too low, to obtain a good estimate as a sequence of straight lines.

4.3.5 Summary of Test Problem 2

Table (4.2) summarizes the variances of the two solution methods for the test cases of Section 4.3. The problem and associated model are qualitatively similar to some actual physical systems. The present method obtains a more accurate estimate for every situation except dense, perfect measurements. This is explained by the ability of the EKFS method to match every measurement perfectly if enough process noise is added. Although the estimate is not improved between the measurements, if the measurements are "dense enough" and "accurate enough", and the model is "good enough", the EKFS method may obtain a very accurate estimate. However, if the measurements are imperfect or spaced too far apart, the model error may lead to large estimate errors

between measurements. The process noise approach, which does not improve the model, cannot obtain as accurate a solution as the present method.

The addition of the "unmodeled disturbance" term is seen to lead the present method to more accurate solutions than the EKFS method, even though the disturbance estimate is poor, for every case except dense, perfect measurements.

TABLE 4.1

Estimate Error Variances of Test Problem 1

Number of Measurements	EKFS		Present	
	Perfect	Imperfect	Perfect	Imperfect
6	6.17	5.36	.312	.464
11	1.93	1.33	.052	.36
21	.562	.648	.014	.316
101	.030	.204	.015	.177

TABLE 4.2

Estimate Error Variances of Test Problem 2

Number of Measurements	EKFS		Present Method	
	Perfect	Imperfect	Perfect	Imperfect
6	.968	.690	.085	.592
11	.459	.512	.041	.452
21	.119	.130	.017	.086
101	.006	.077	.015	.066

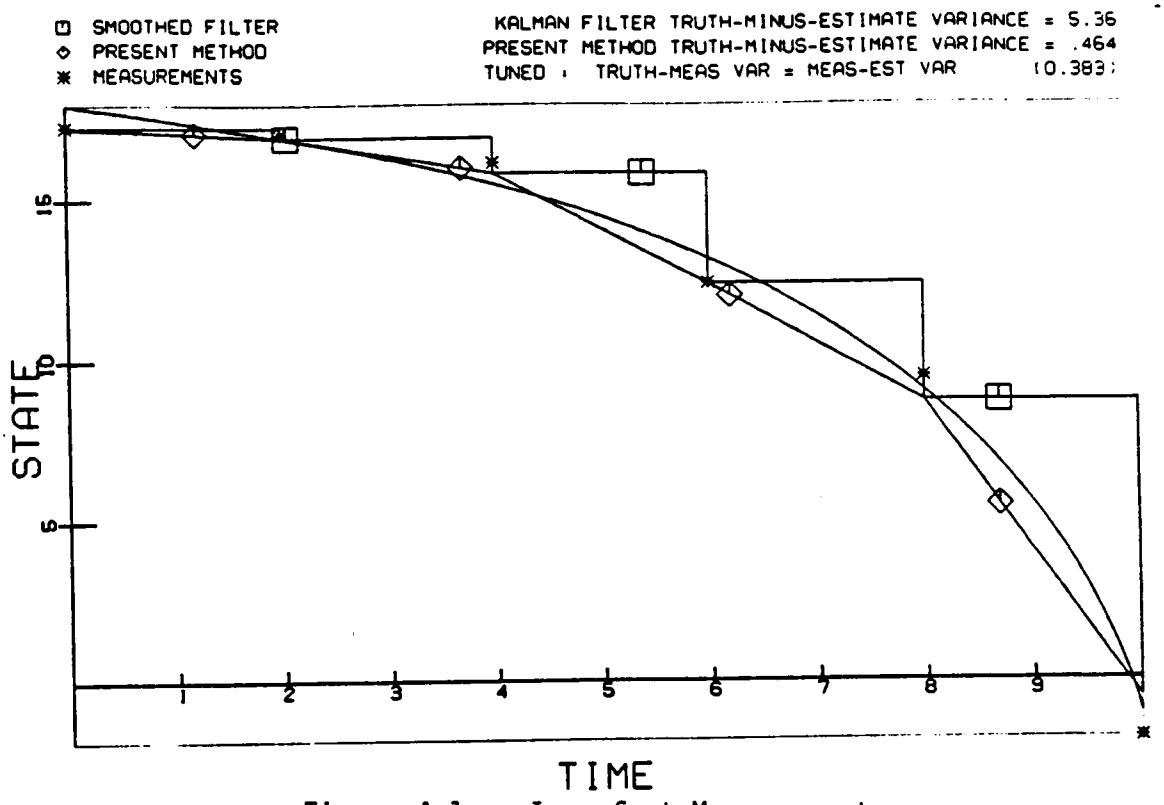


Figure 4.1-a Imperfect Measurements

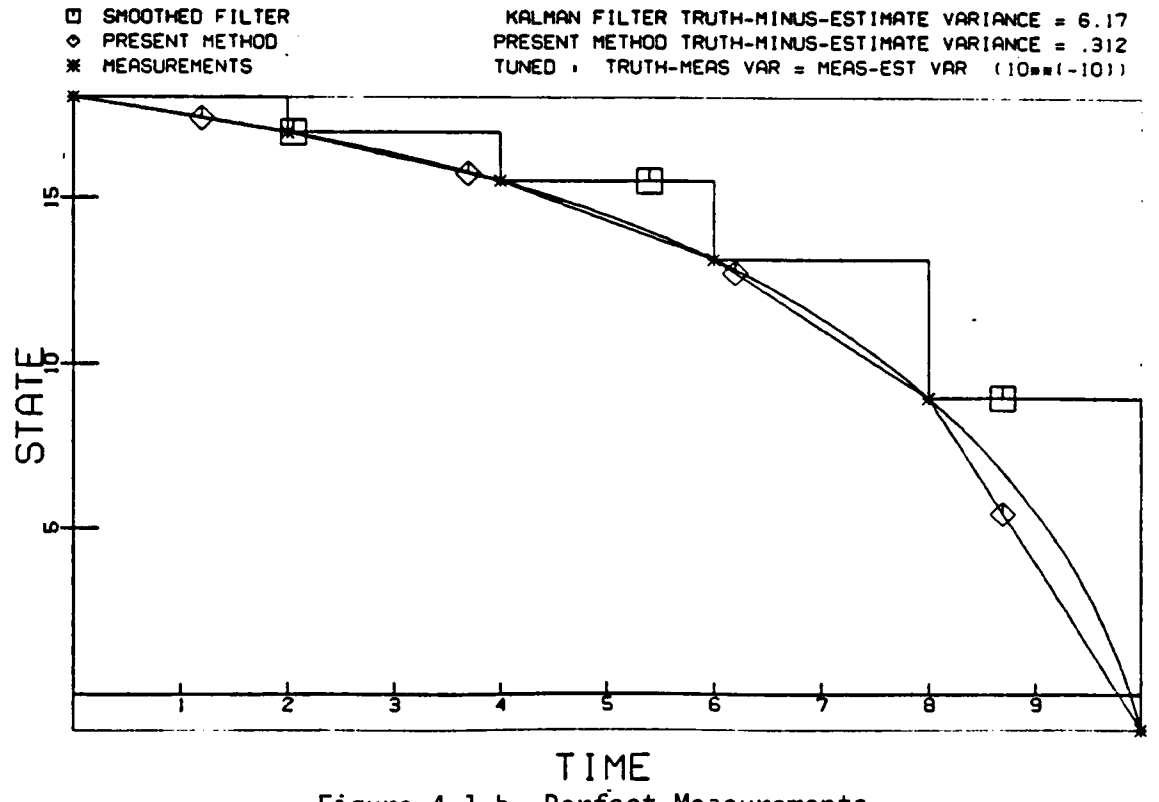


Figure 4.1-b Perfect Measurements

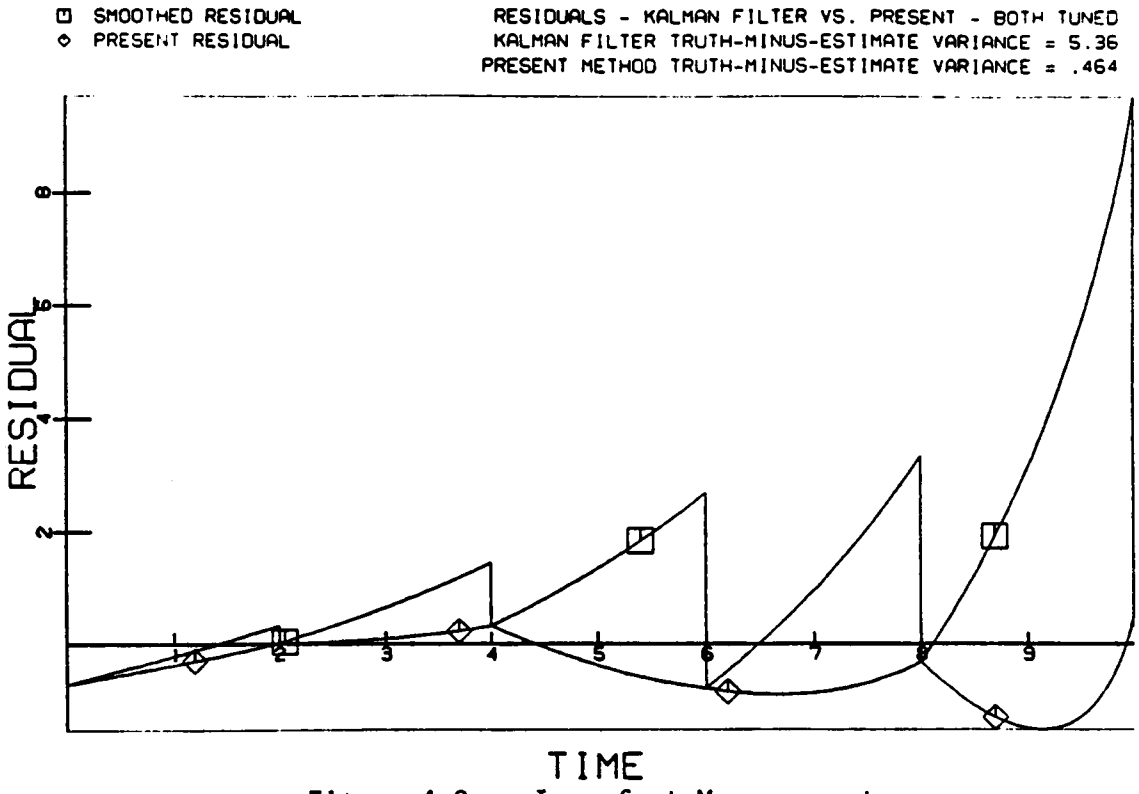


Figure 4.2-a Imperfect Measurements

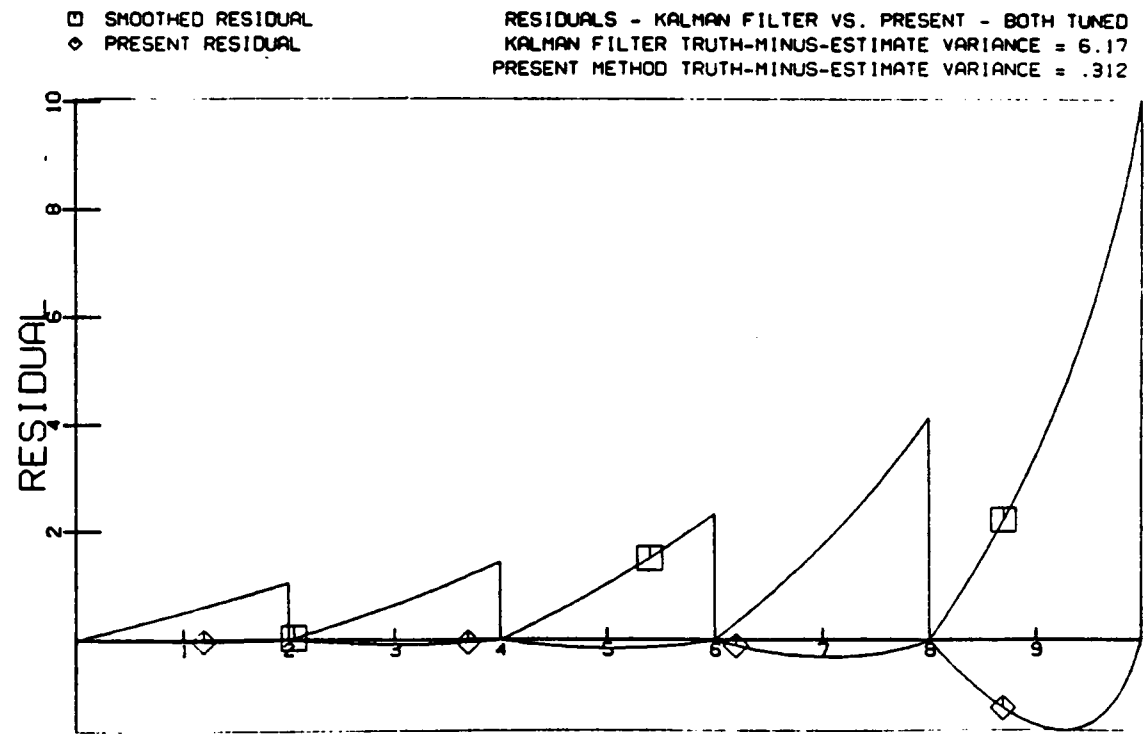


Figure 4.2-b Perfect Measurements

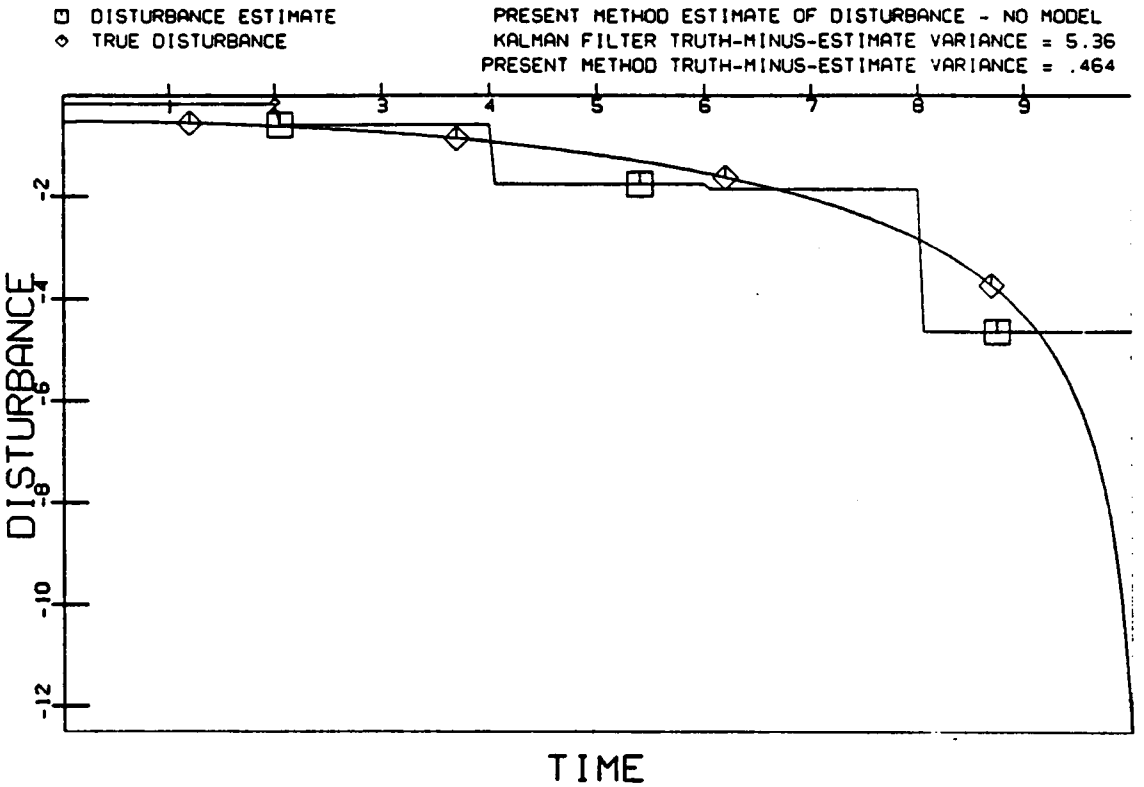


Figure 4.3-a Imperfect Measurements

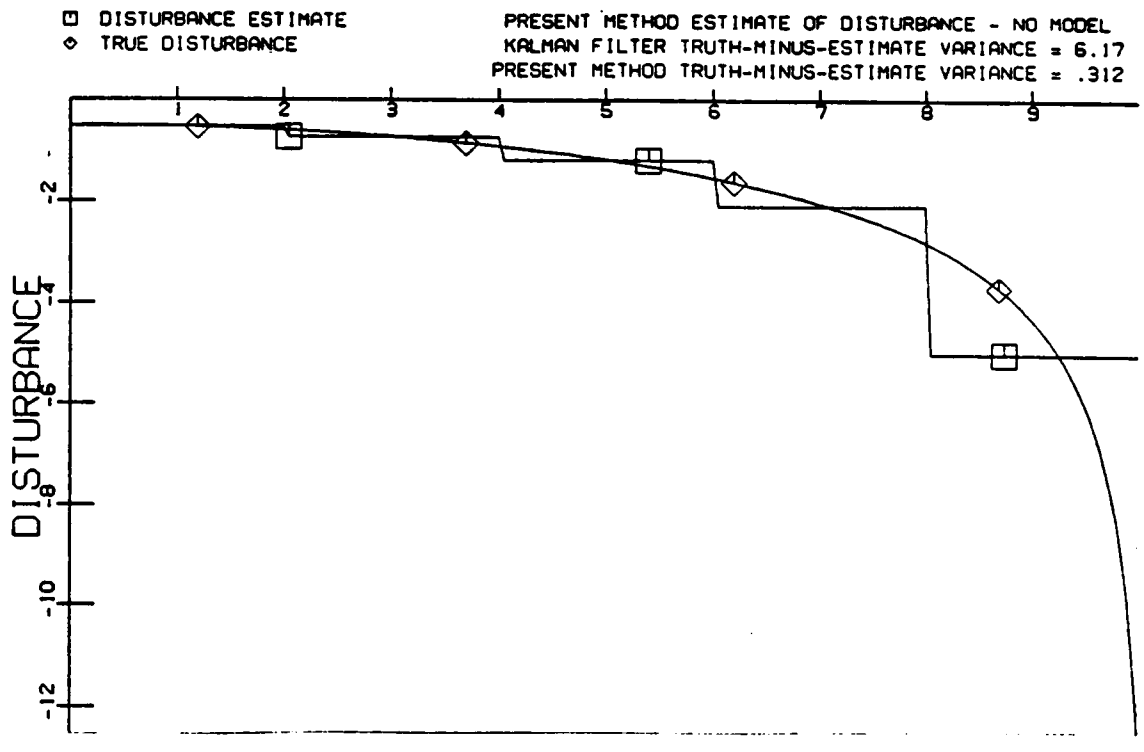


Figure 4.3-b Perfect Measurements

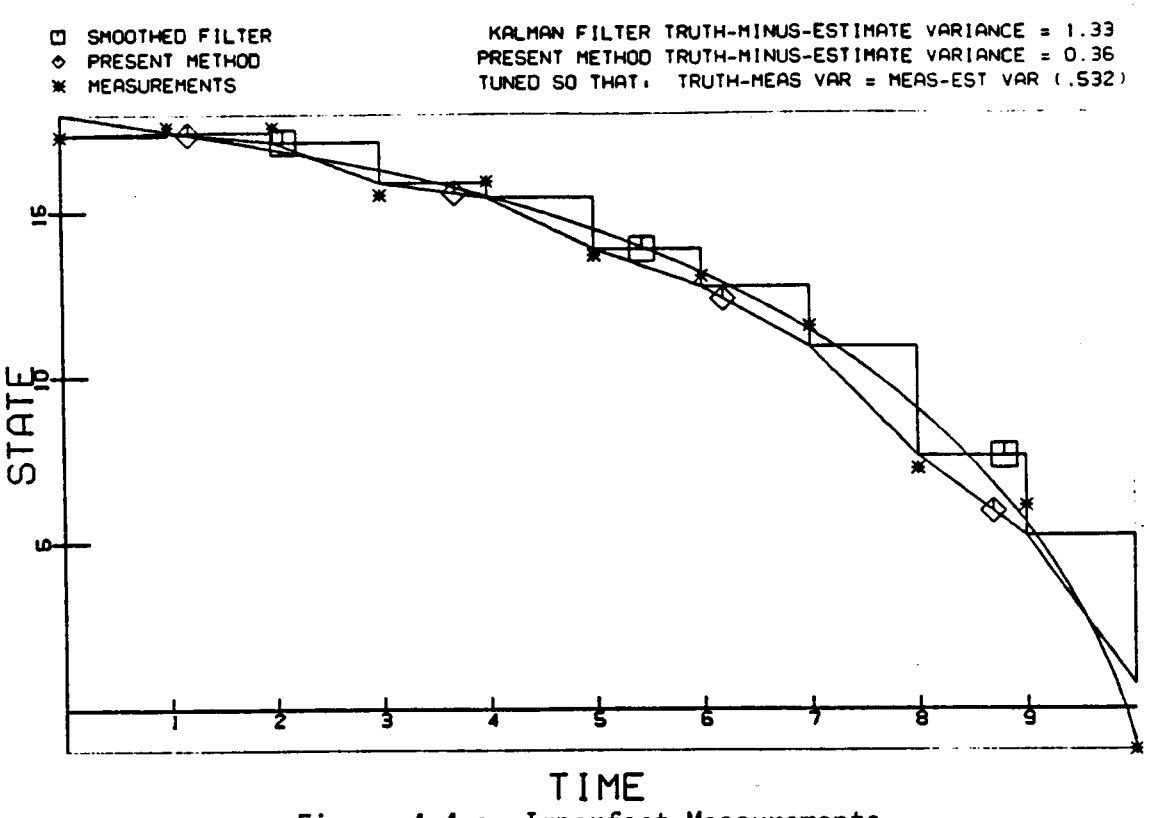


Figure 4.4-a Imperfect Measurements

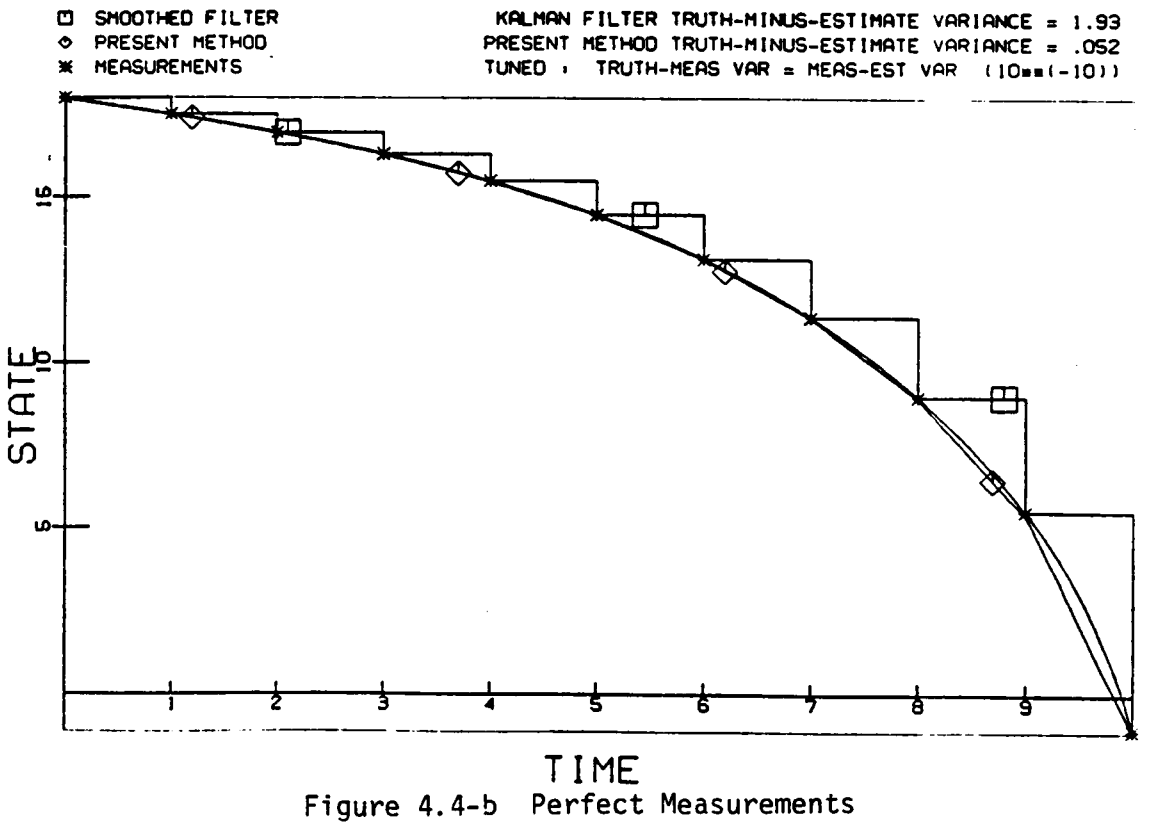


Figure 4.4-b Perfect Measurements

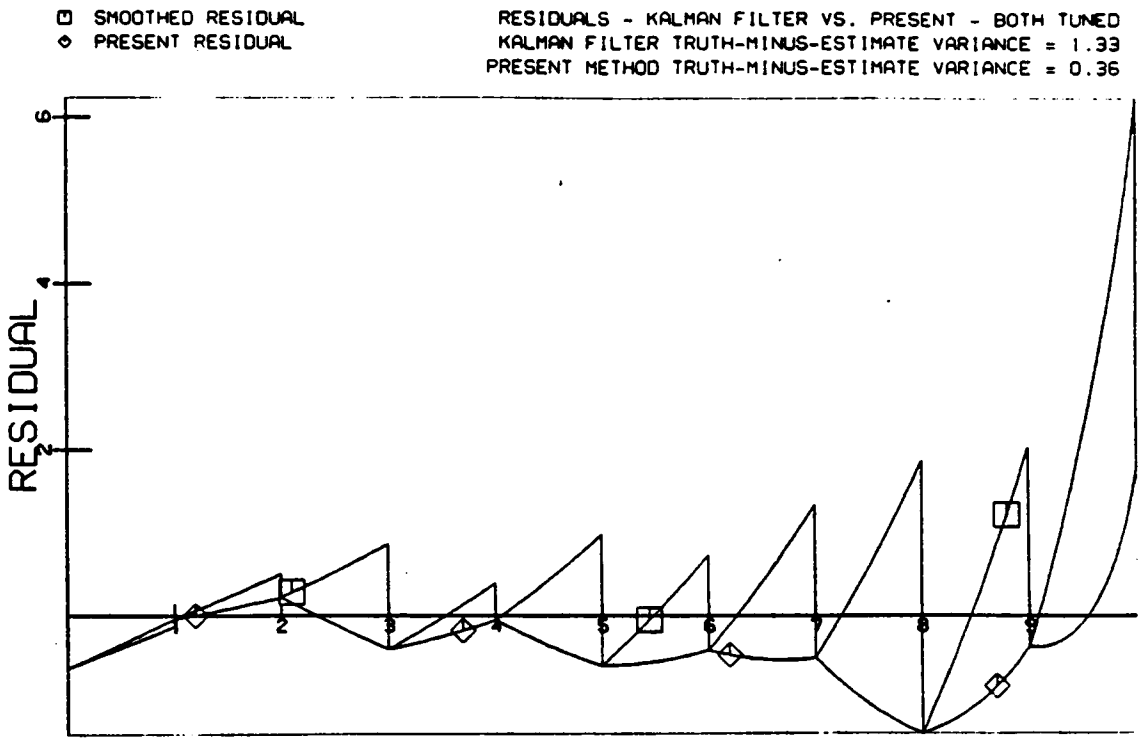


Figure 4.5-a Imperfect Measurements

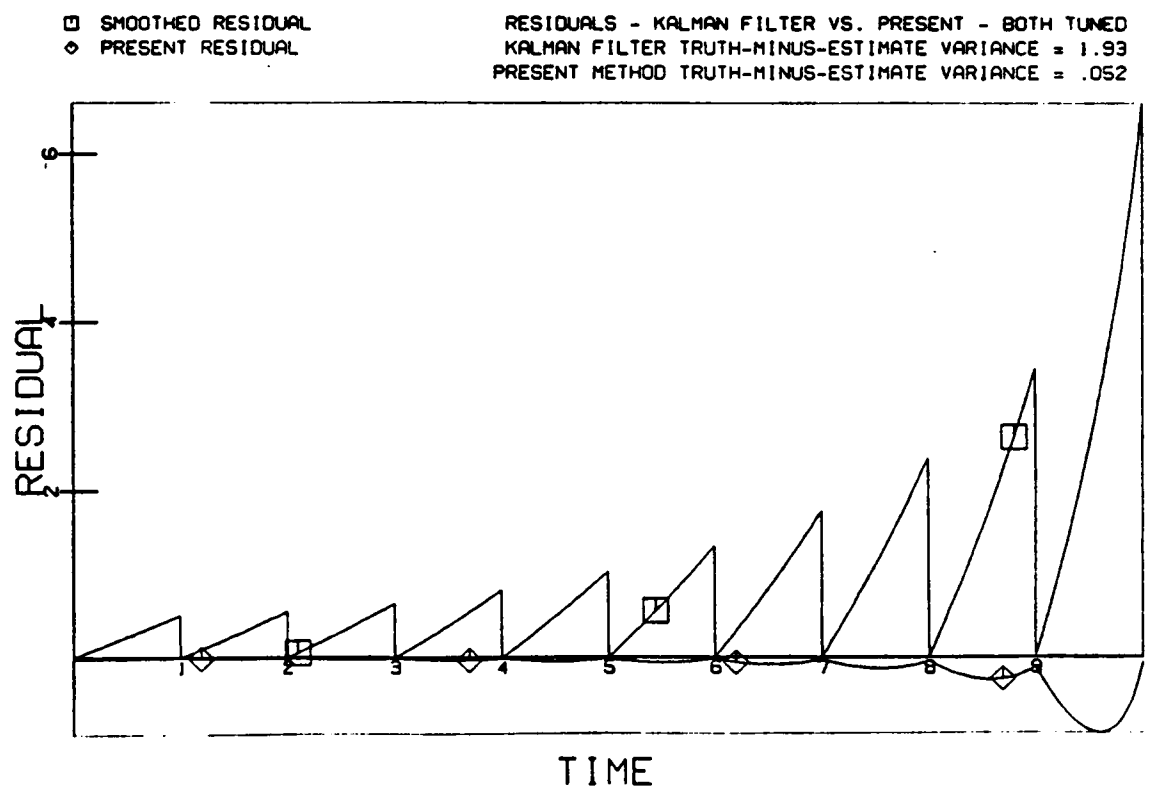


Figure 4.5-b Perfect Measurements

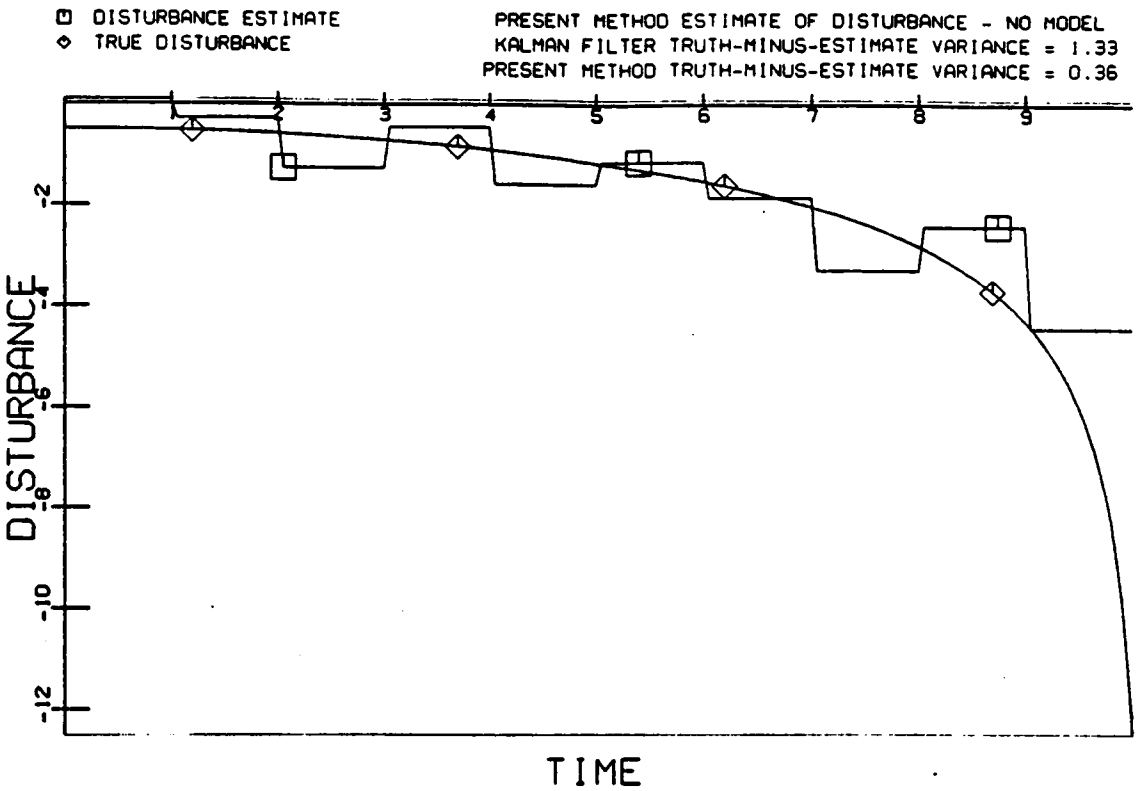


Figure 4.6-a Imperfect Measurements

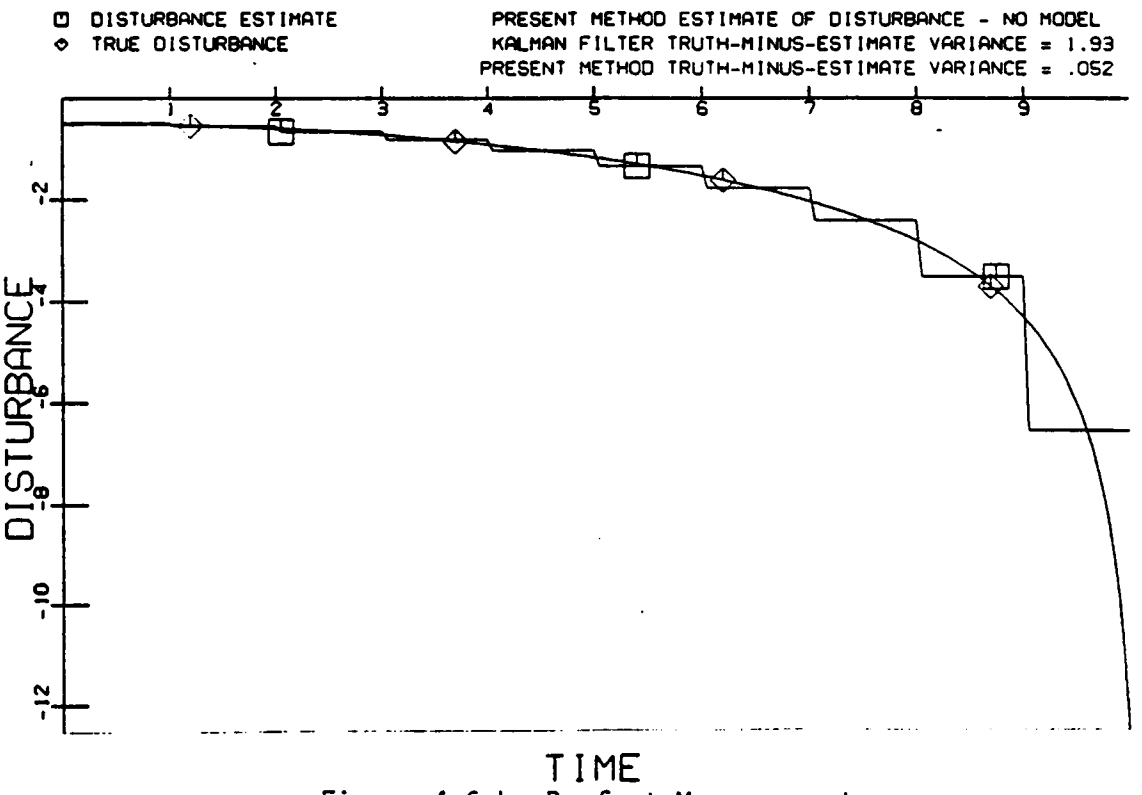


Figure 4.6-b Perfect Measurements

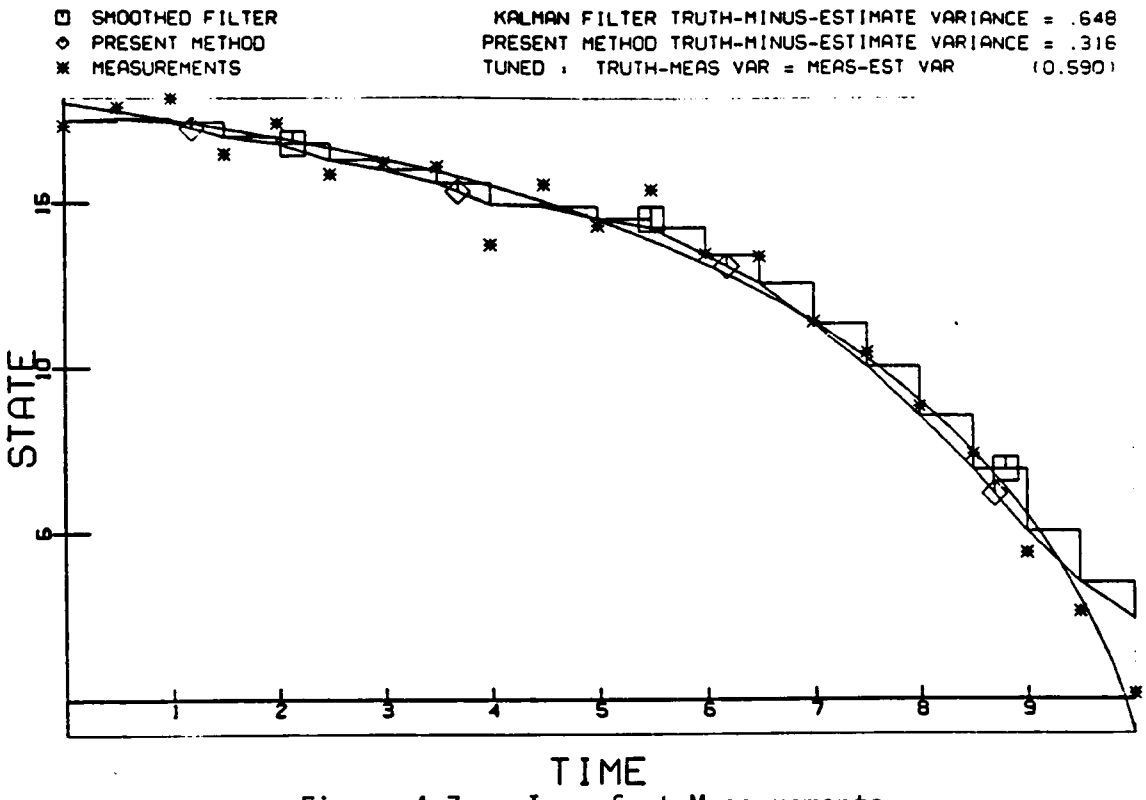


Figure 4.7-a Imperfect Measurements

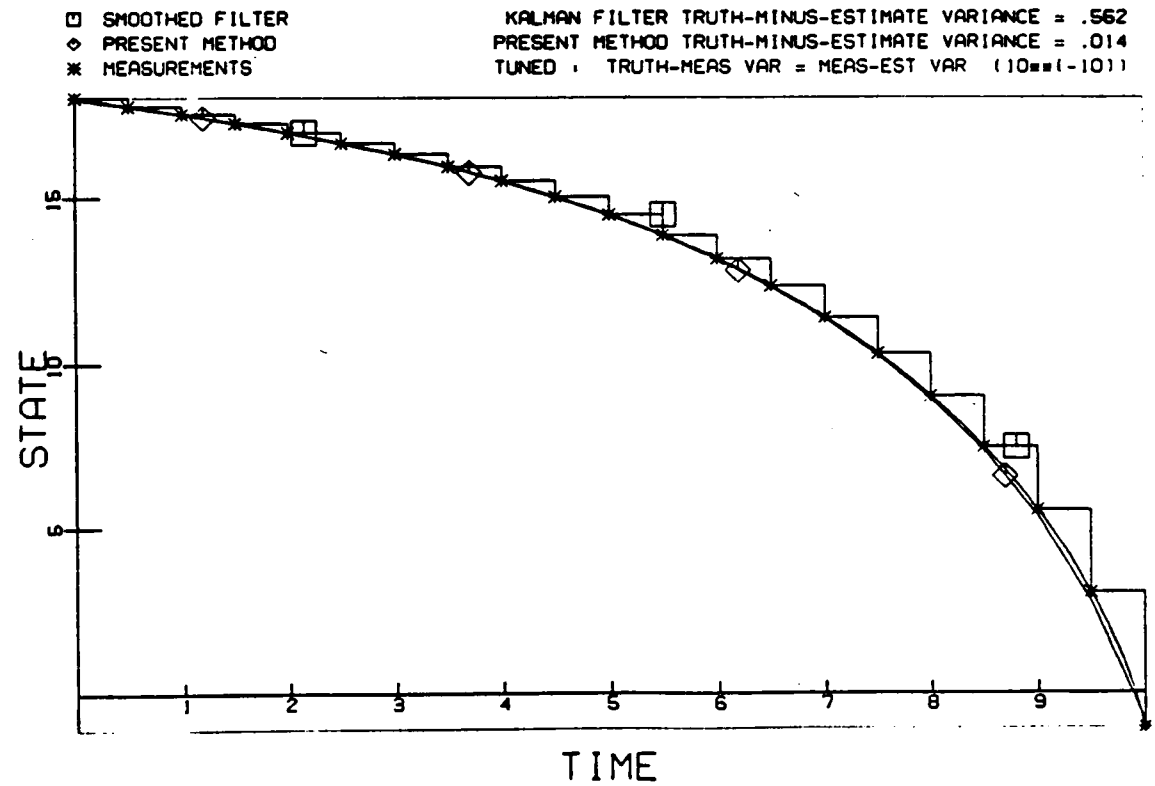


Figure 4.7-b Perfect Measurements

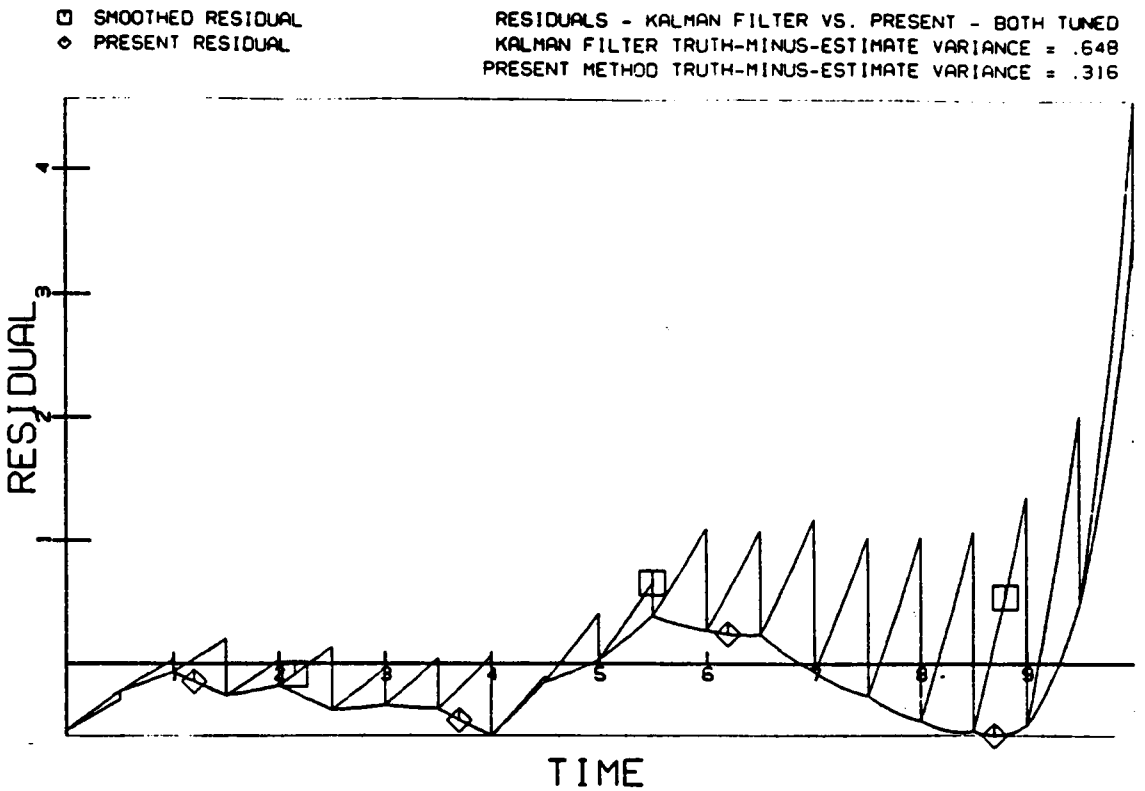


Figure 4.8-a Imperfect Measurements

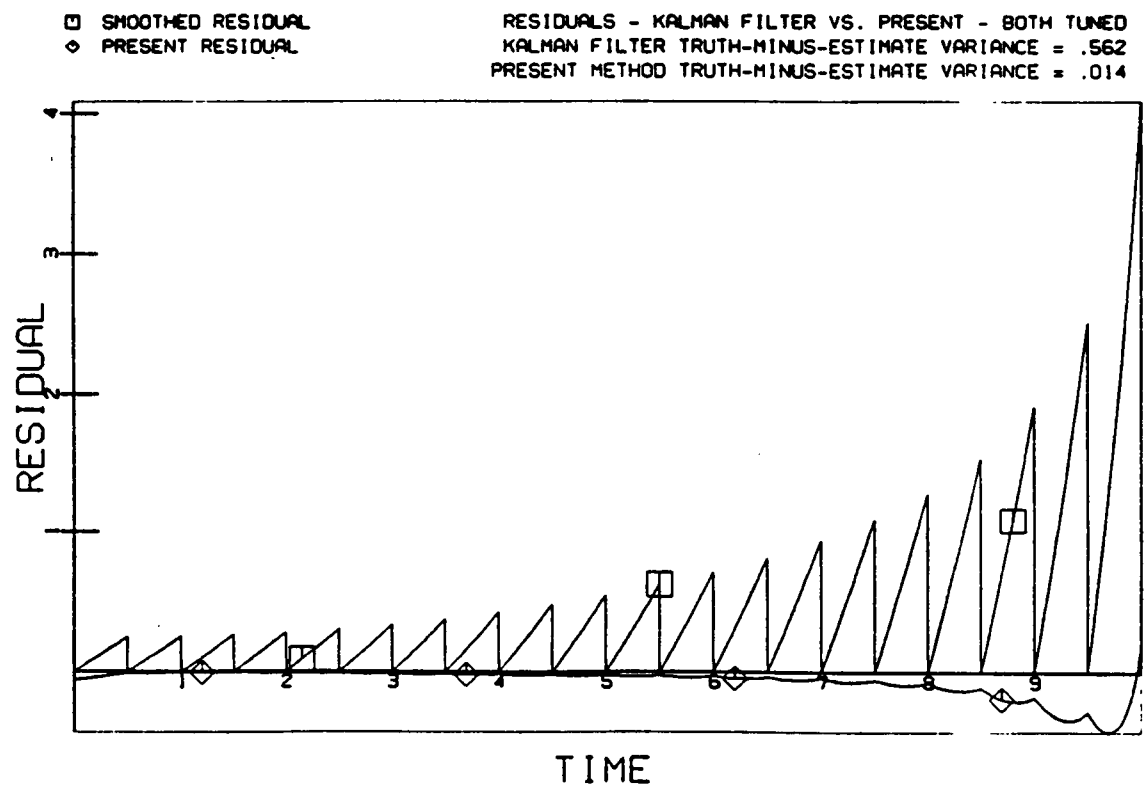


Figure 4.8-b Perfect Measurements

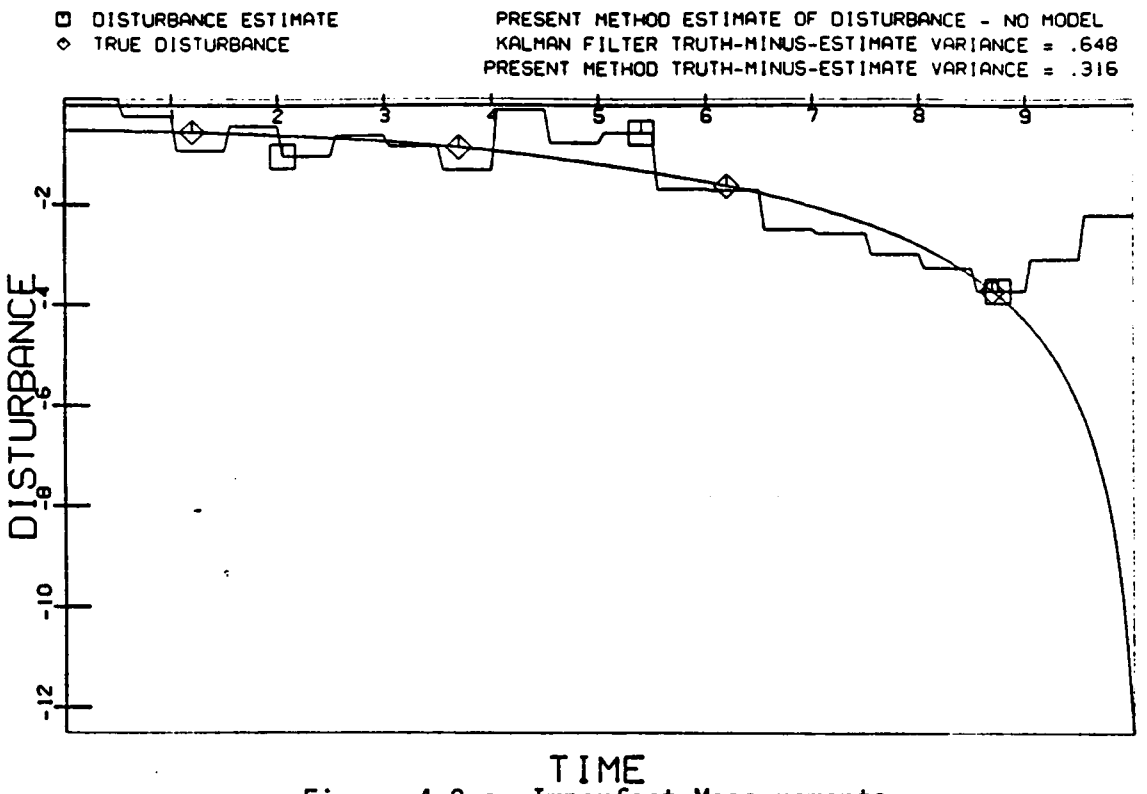


Figure 4.9-a Imperfect Measurements

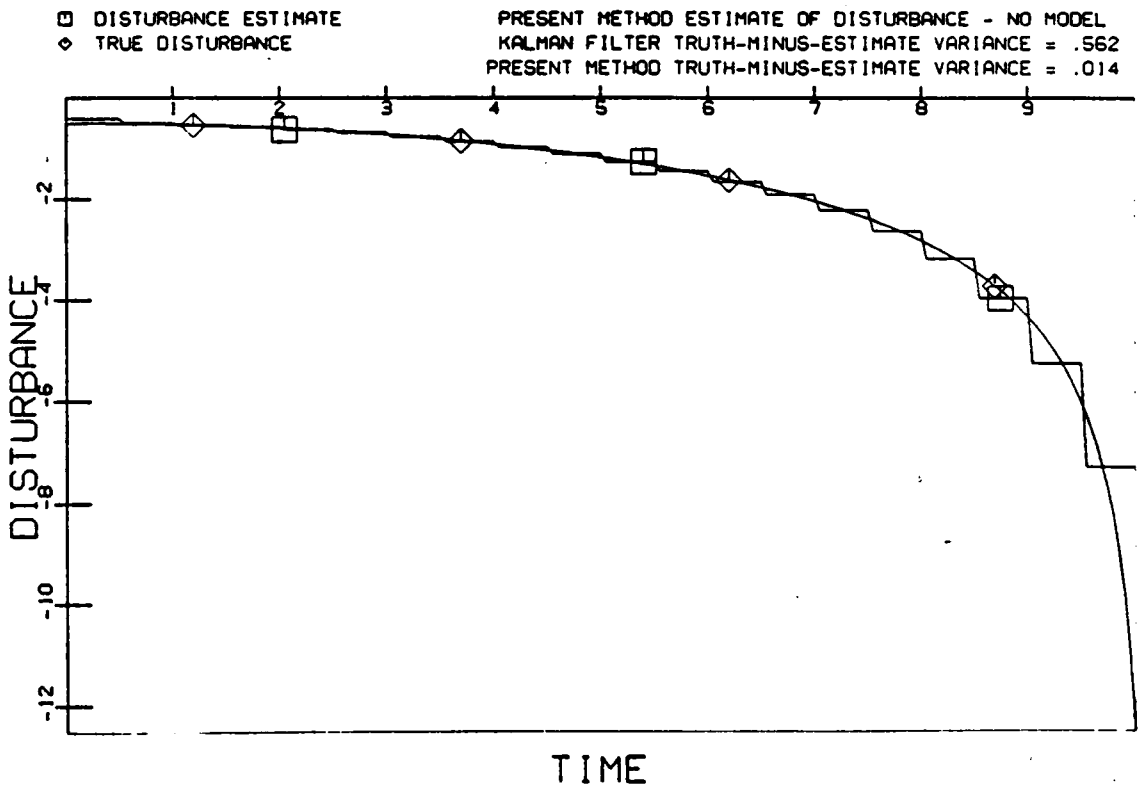


Figure 4.9-a Perfect Measurements

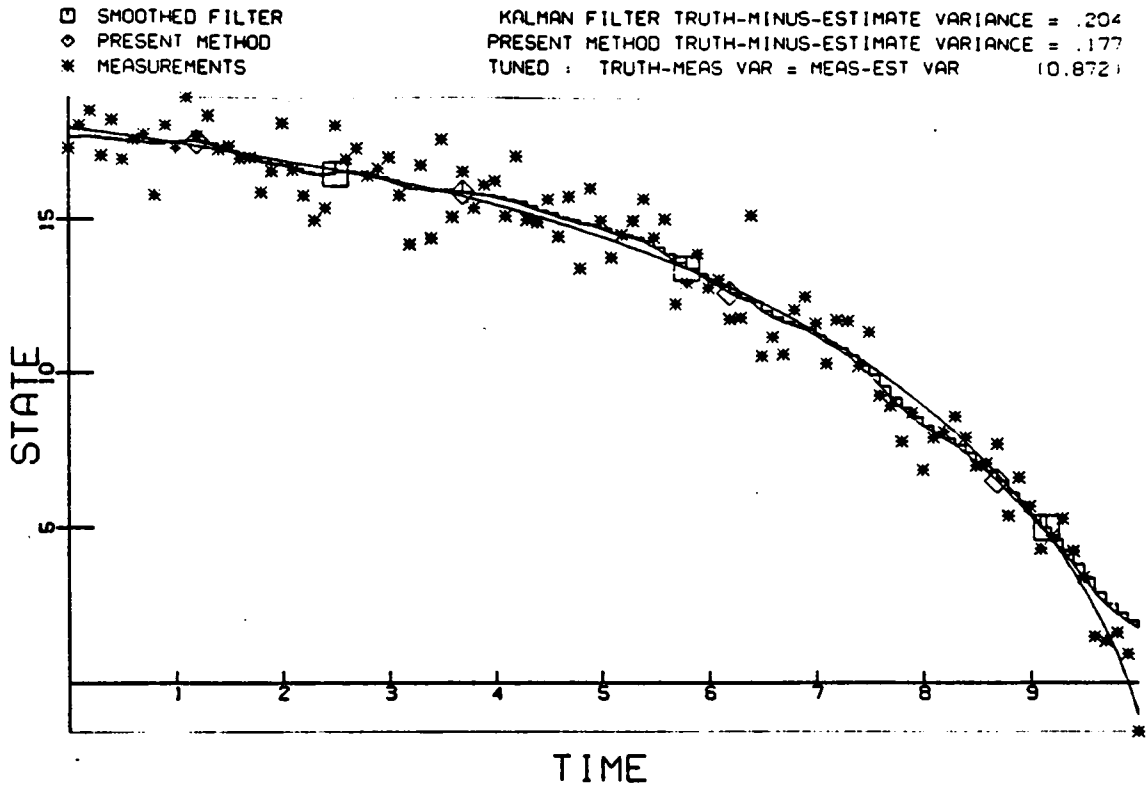


Figure 4.10-a Imperfect Measurements

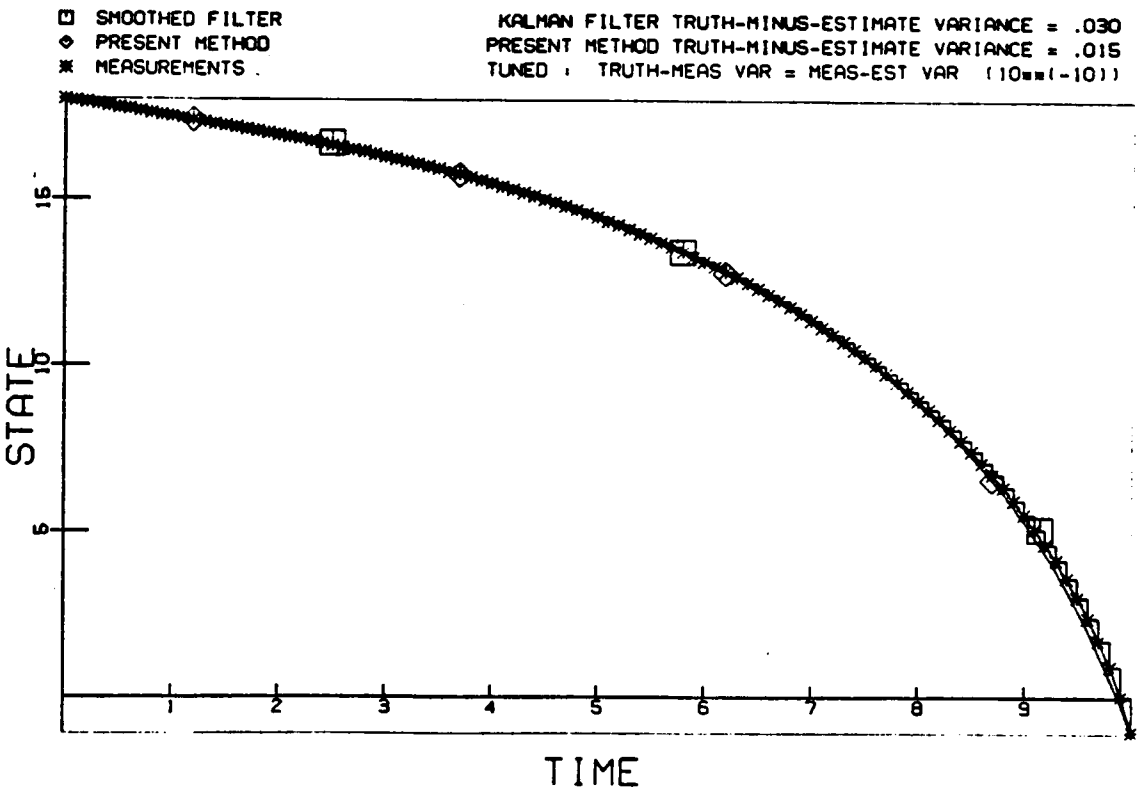


Figure 4.10-b Perfect Measurements

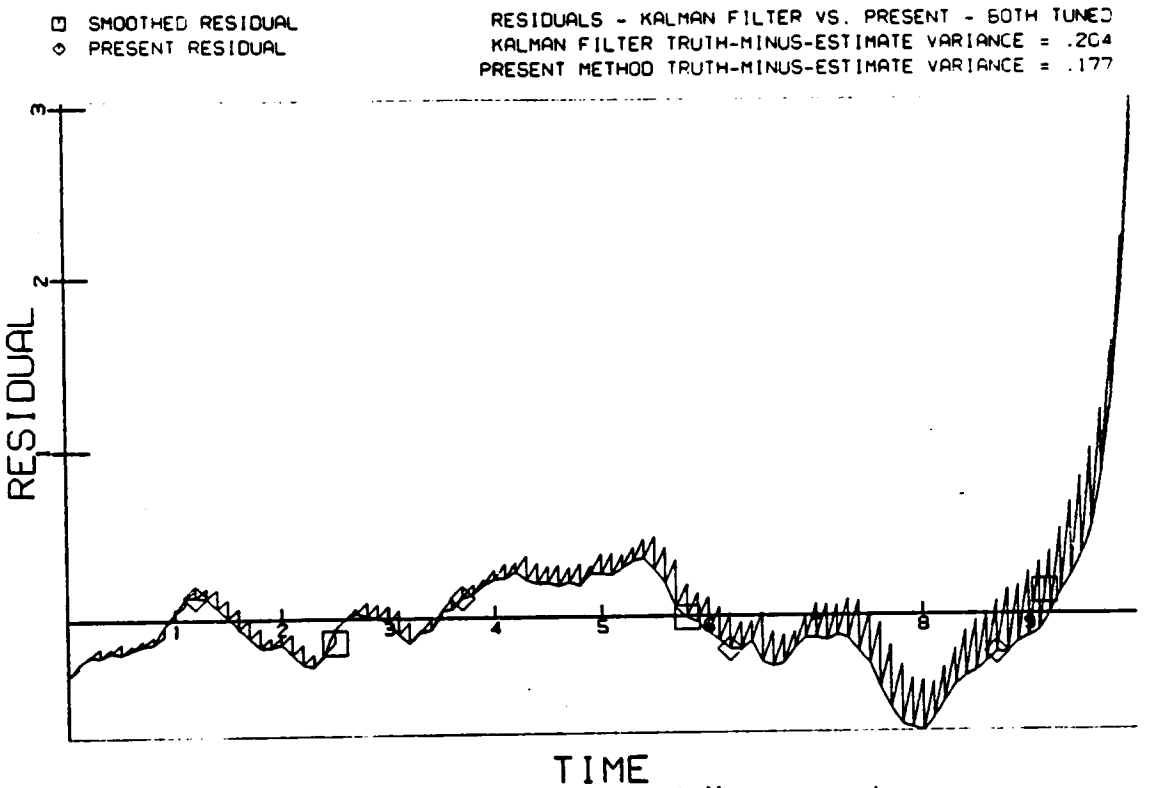


Figure 4.11-a Imperfect Measurements

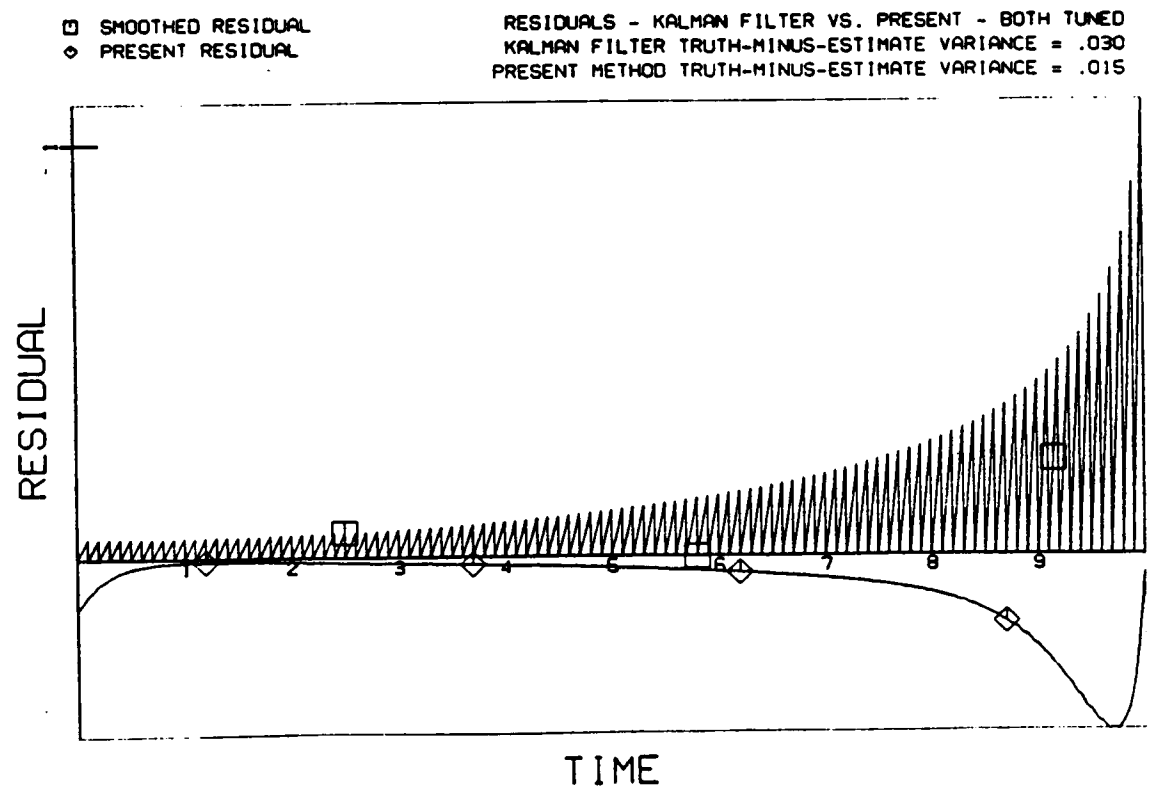


Figure 4.11-b Perfect Measurements

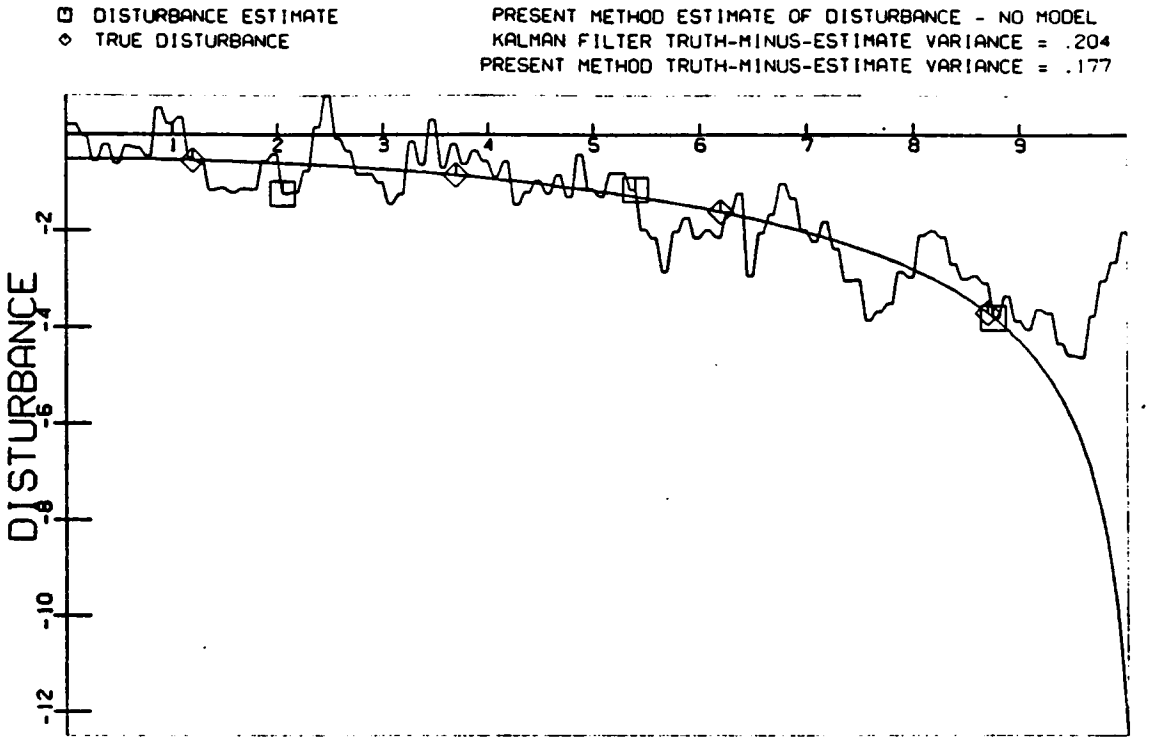


Figure 4.12-a Imperfect Measurements

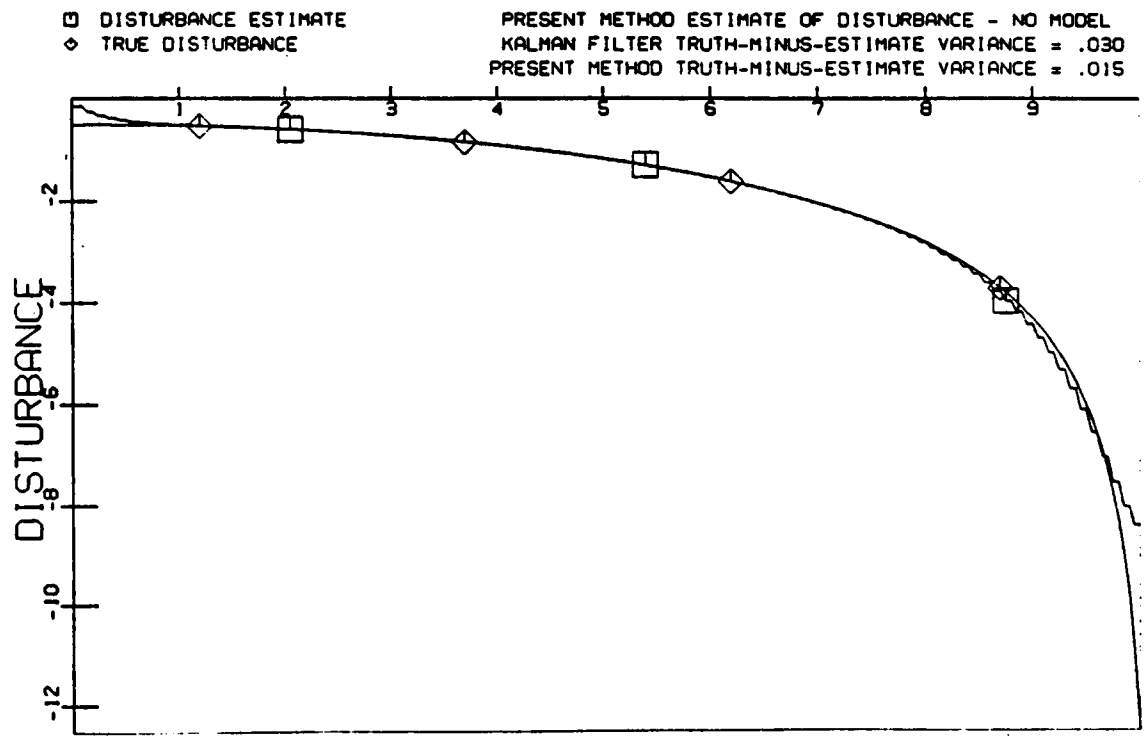


Figure 4.12-b Perfect Measurements

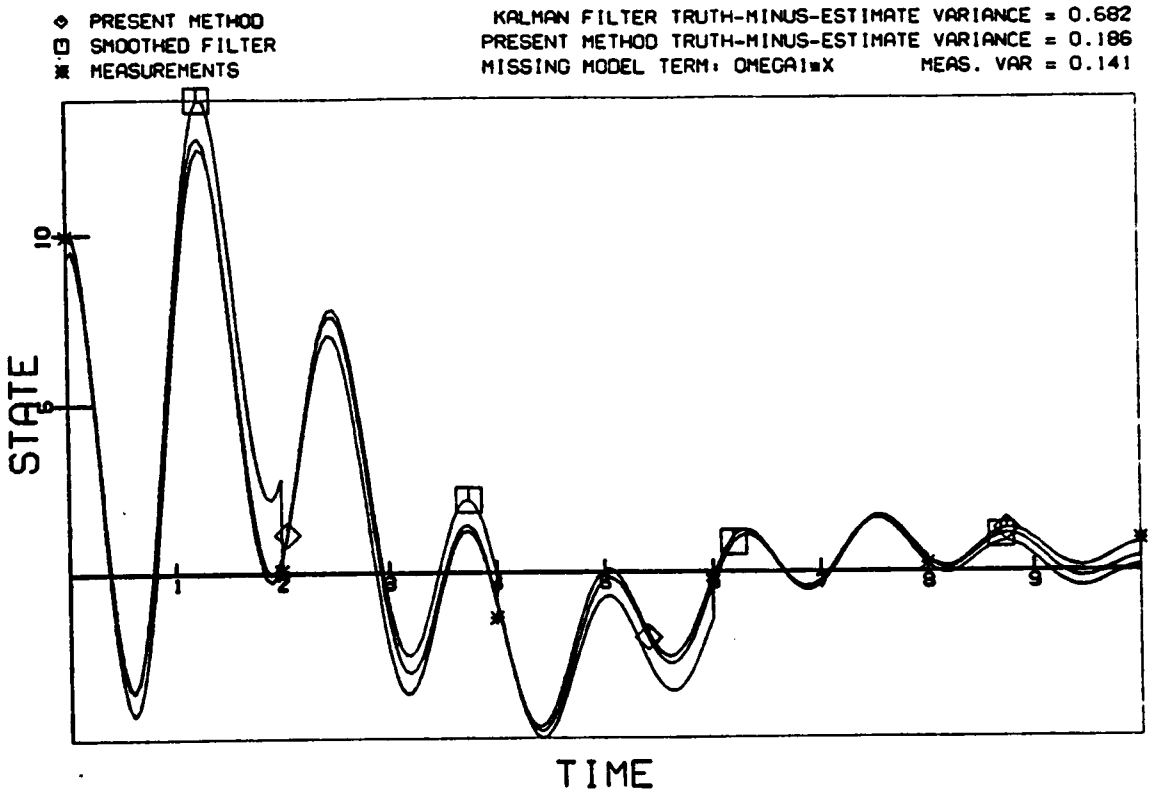


Figure 4.13-a Imperfect Measurements

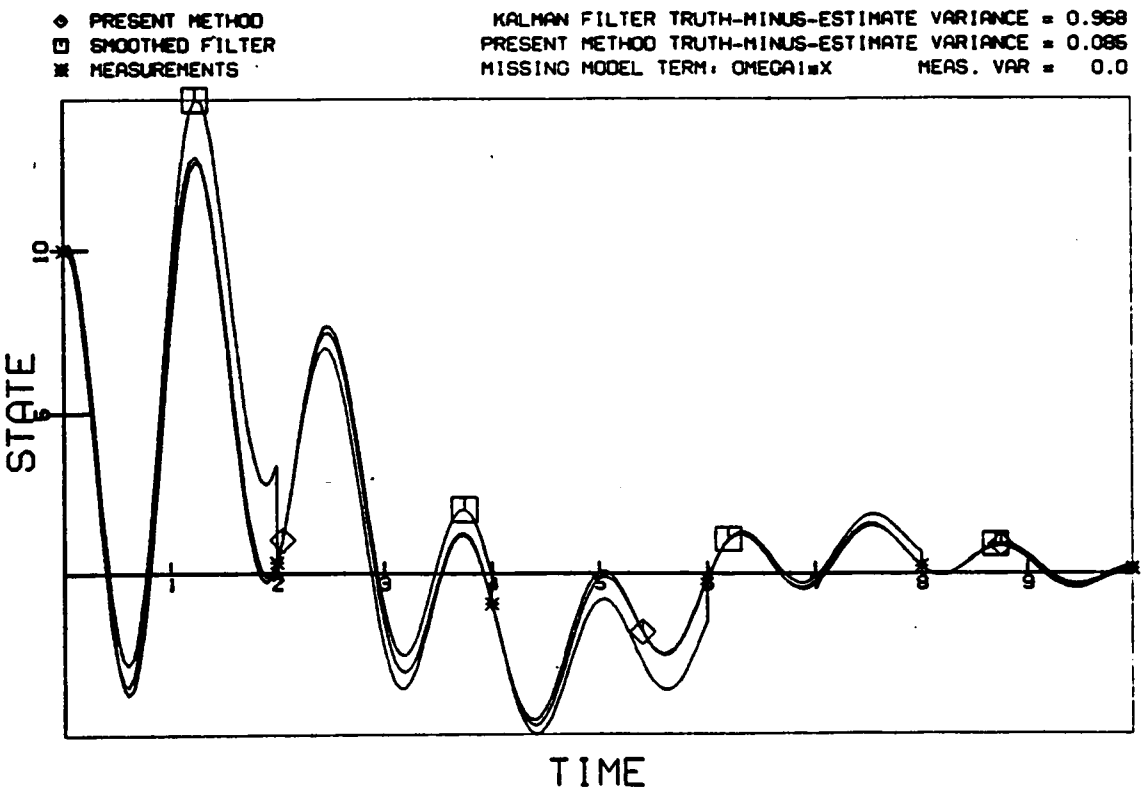


Figure 4.13-b Perfect Measurements

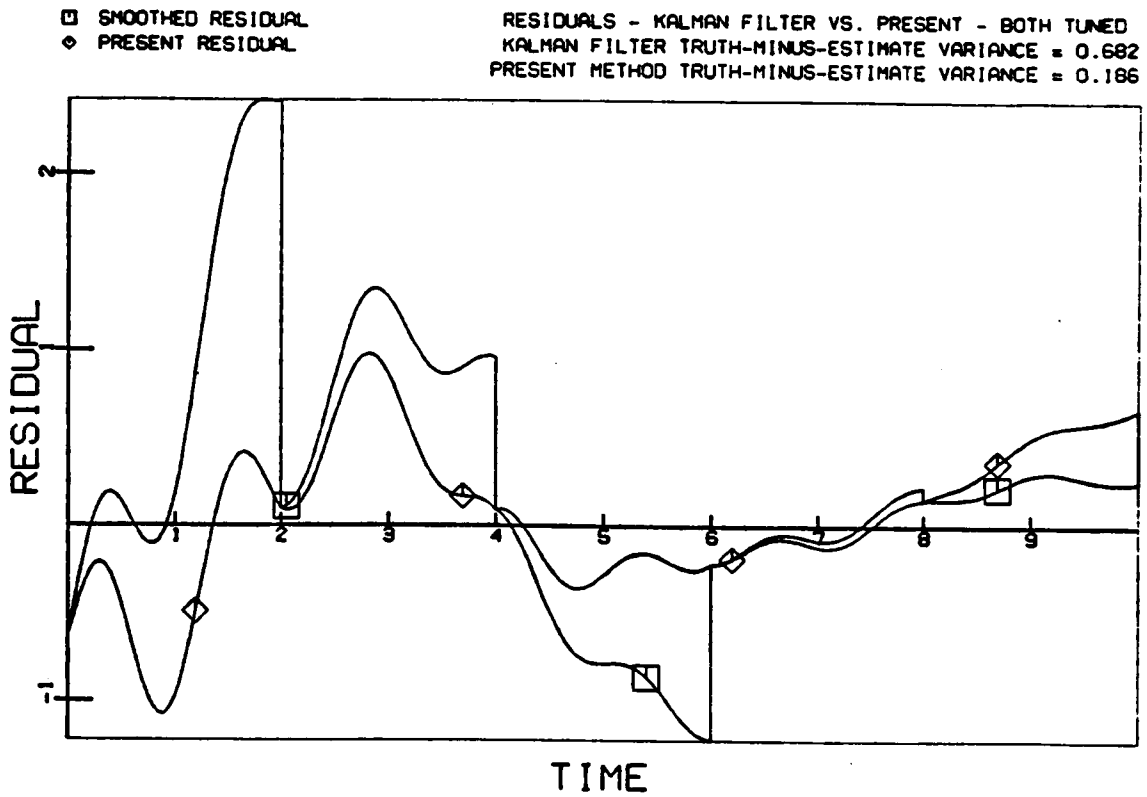


Figure 4.14-a Imperfect Measurements

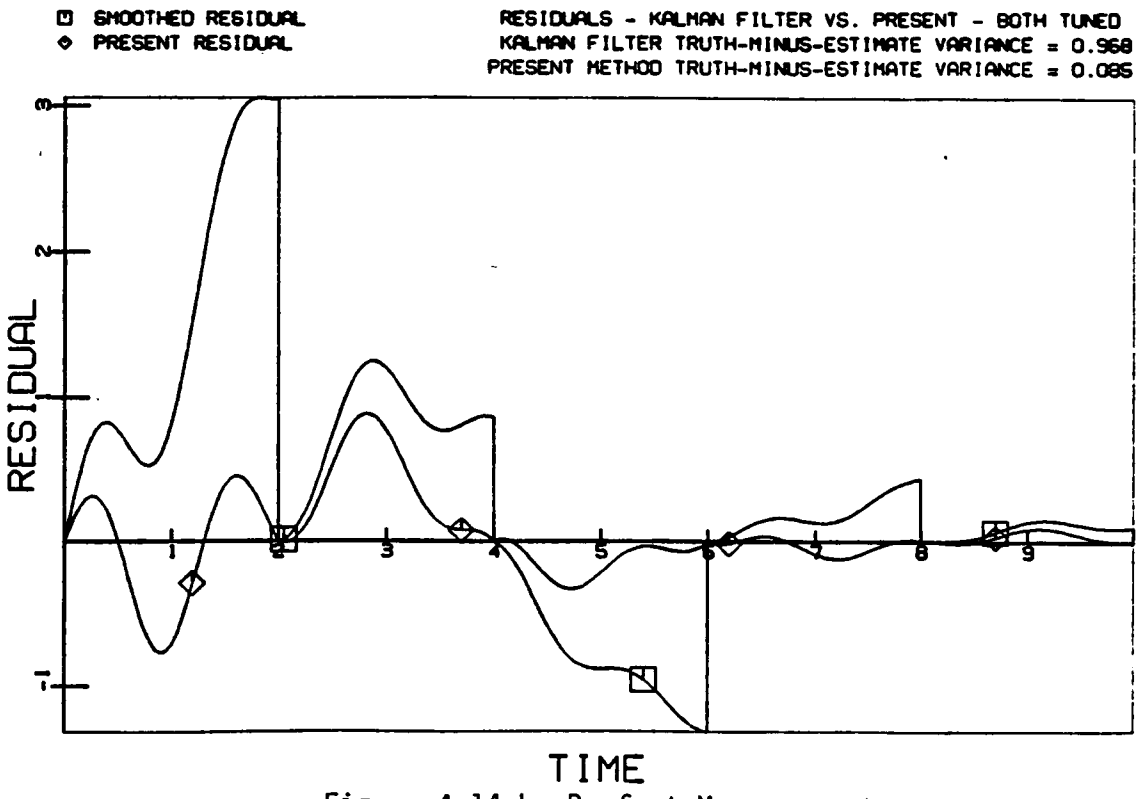


Figure 4.14-b Perfect Measurements

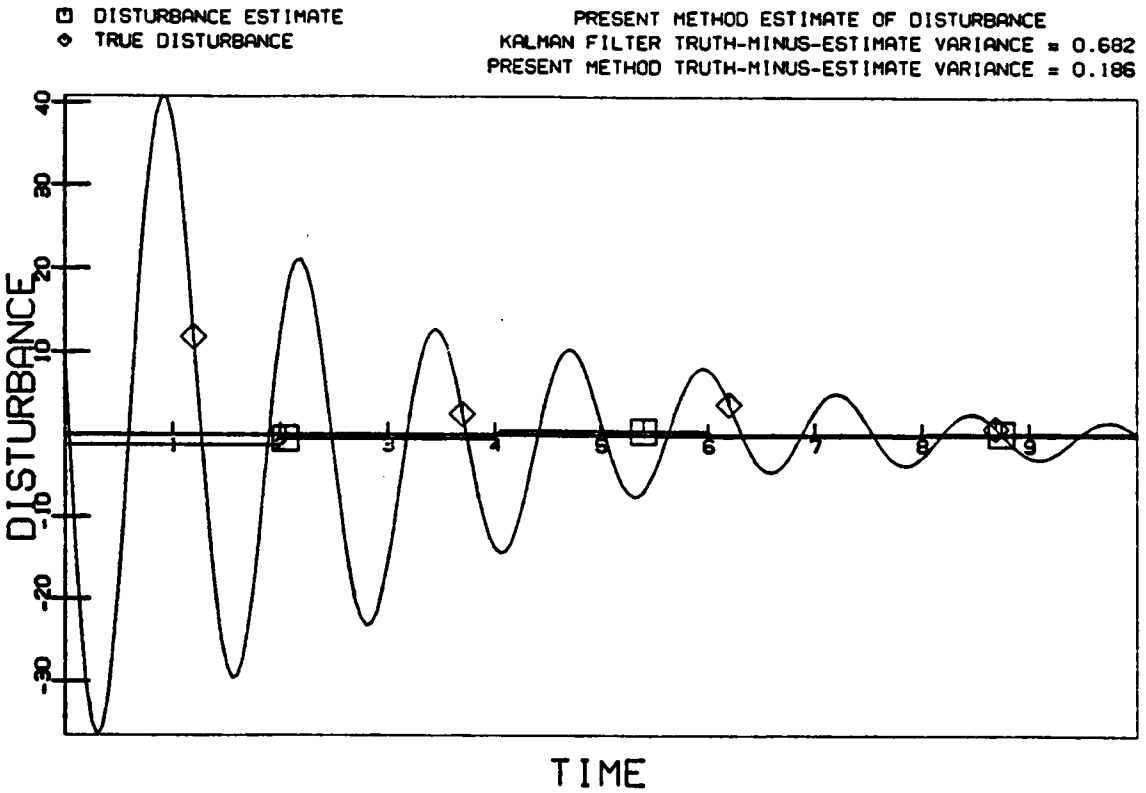


Figure 4.15-a Imperfect Measurements

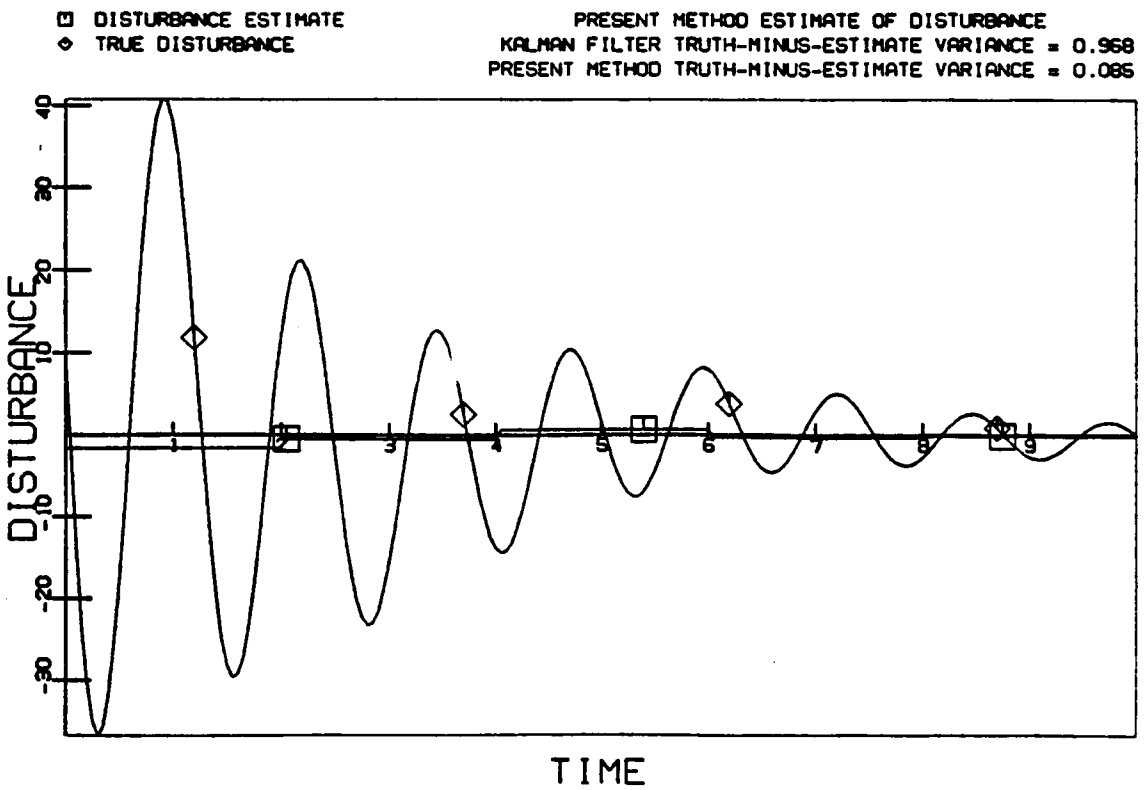


Figure 4.15-b Perfect Measurements

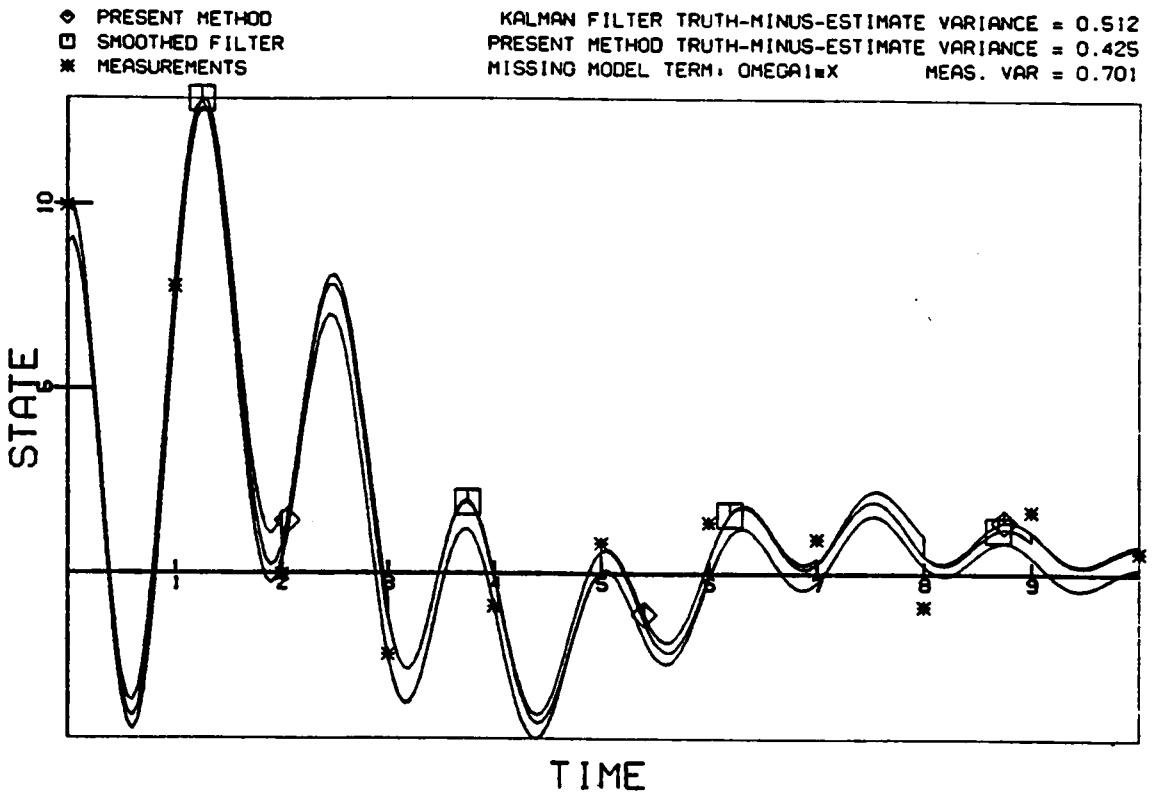


Figure 4.16-a Imperfect Measurements

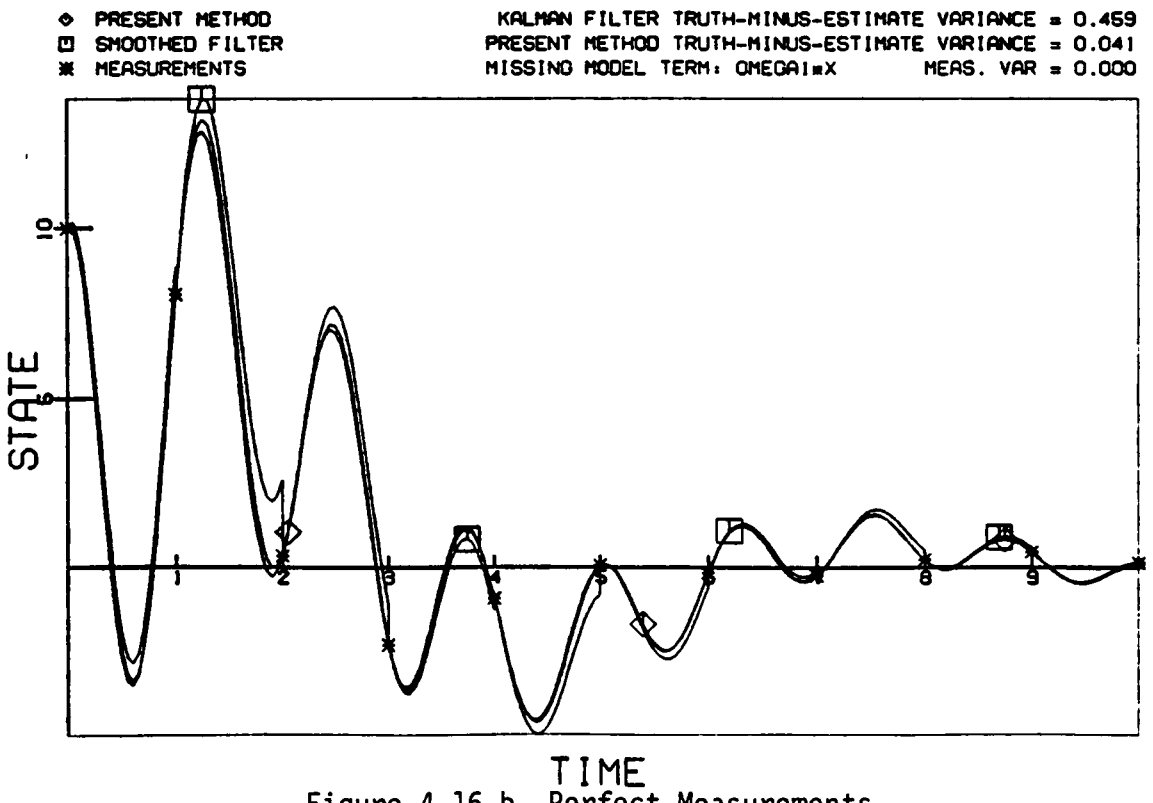


Figure 4.16-b Perfect Measurements

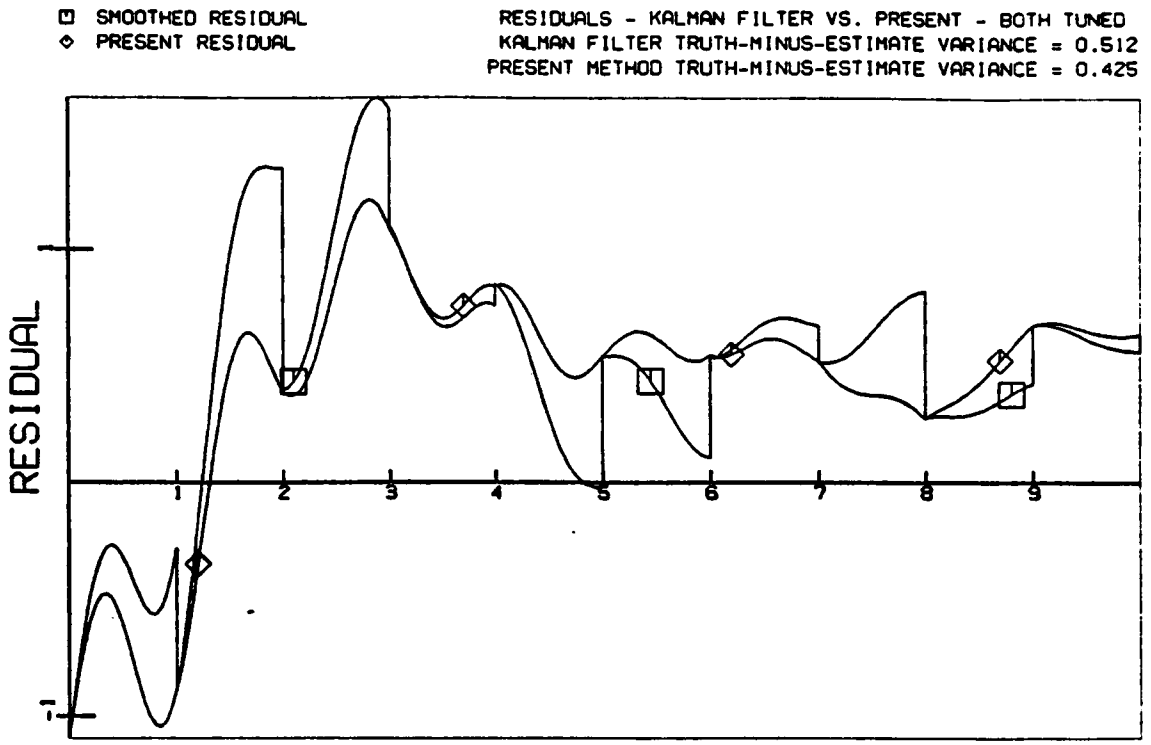


Figure 4.17-a Imperfect Measurements

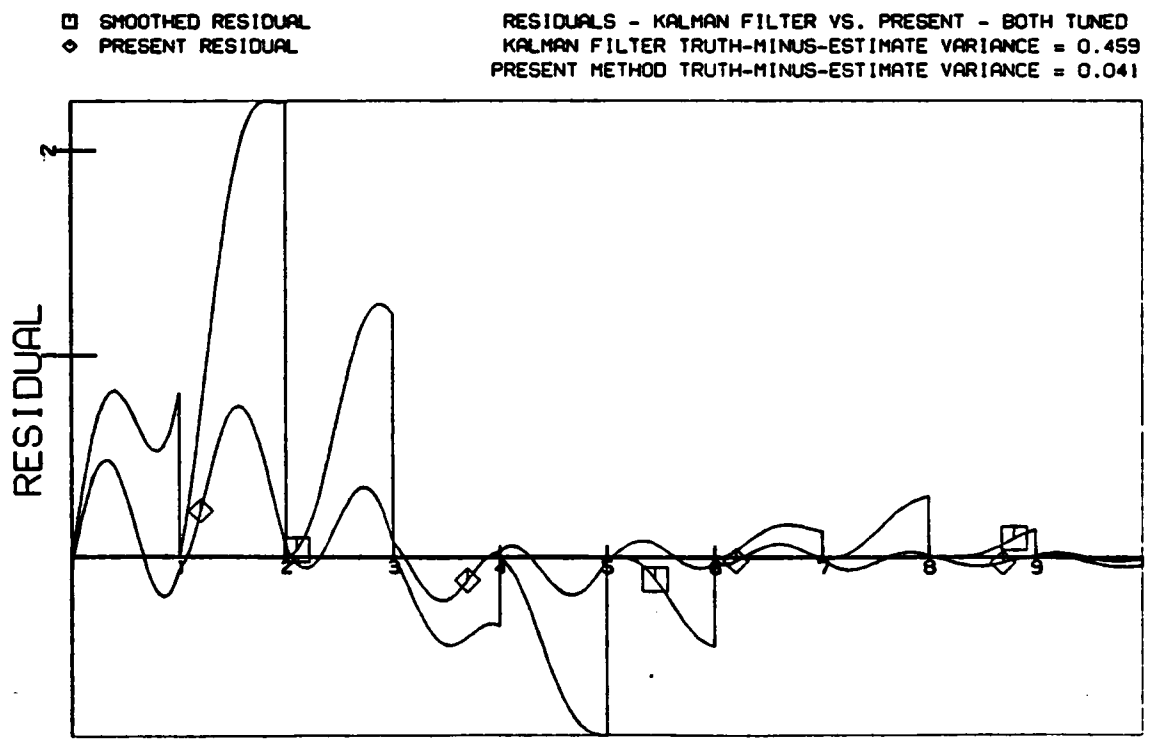


Figure 4.17-b Perfect Measurements

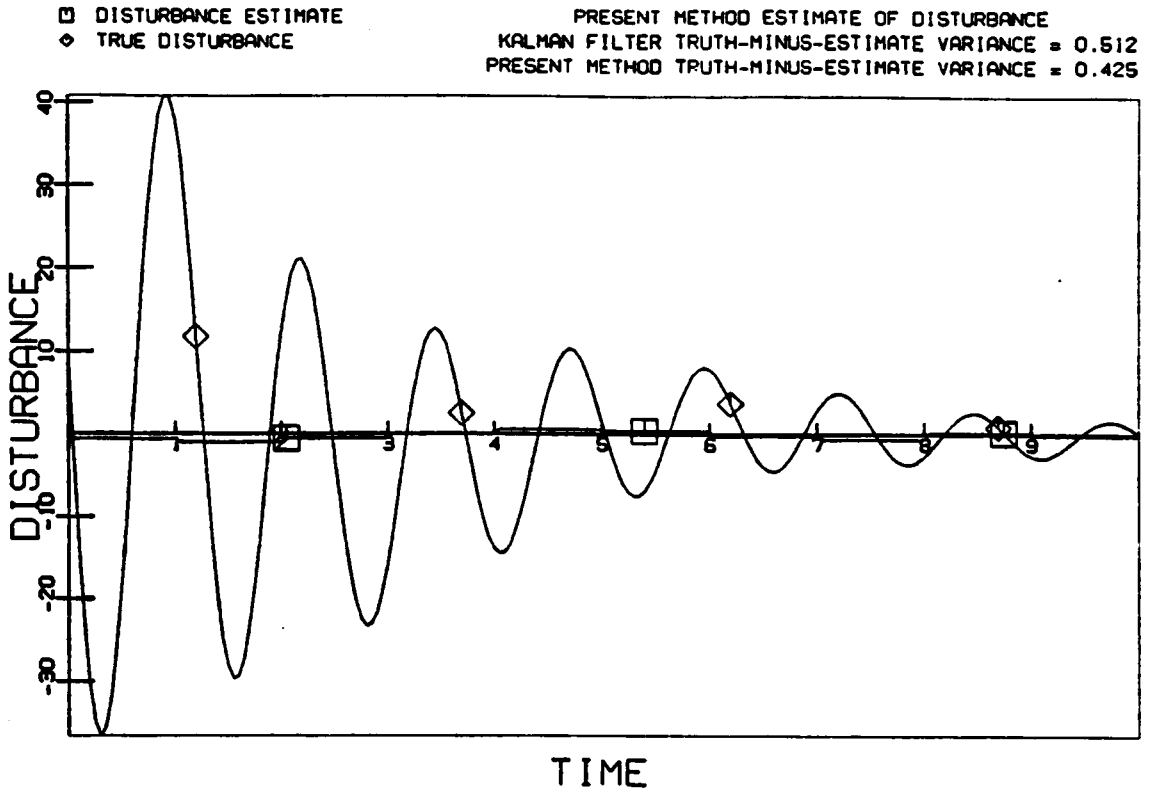


Figure 4.18-a Imperfect Measurements

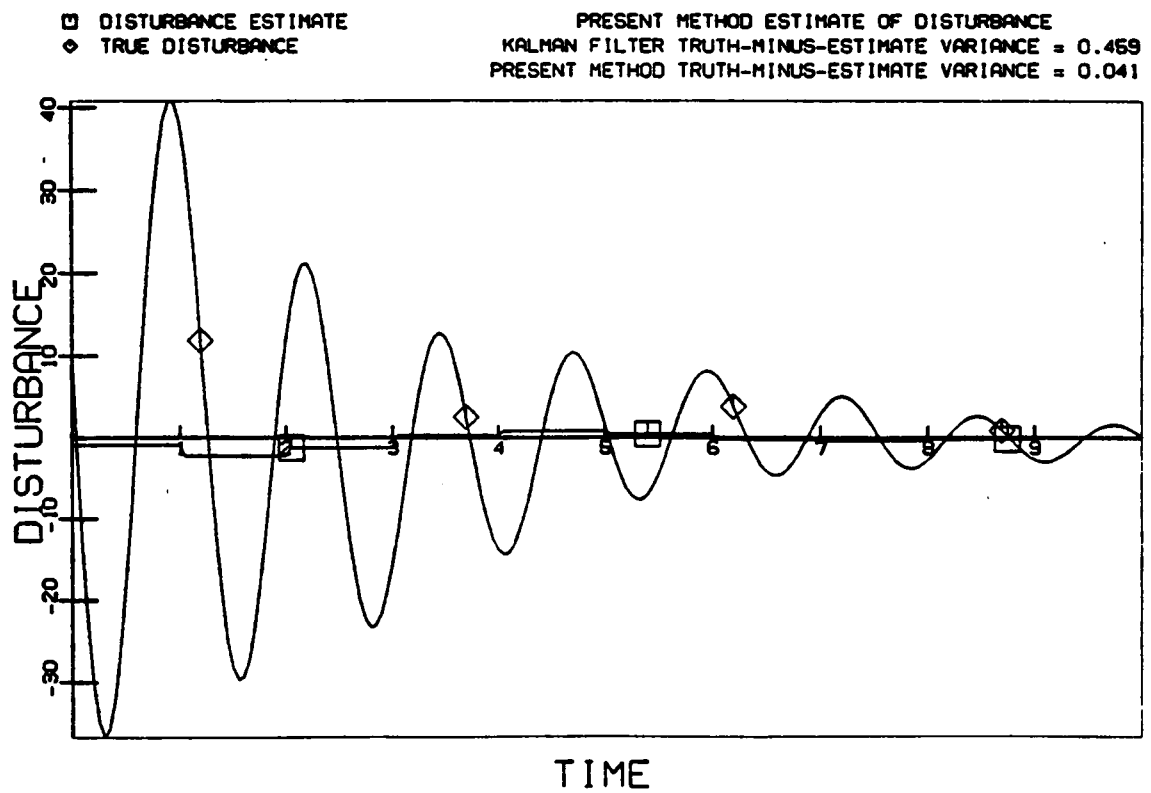


Figure 4.18-b Perfect Measurements

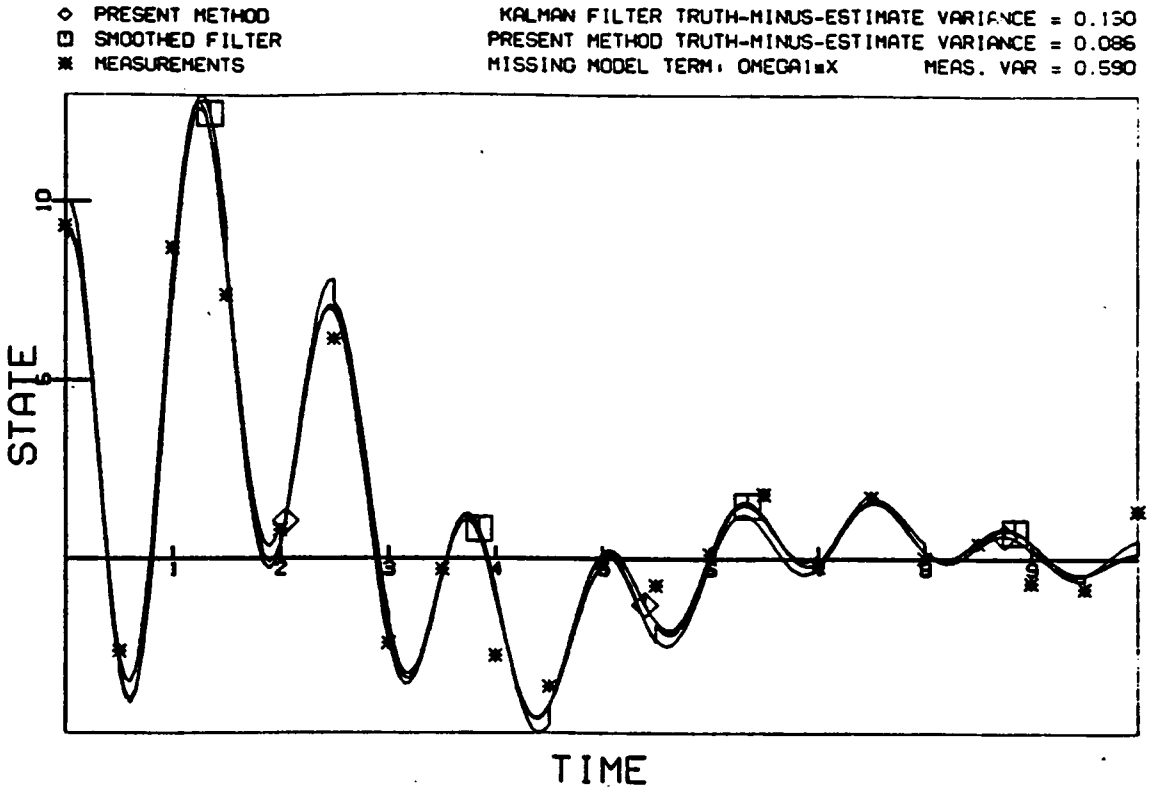


Figure 4.19-a Imperfect Measurements

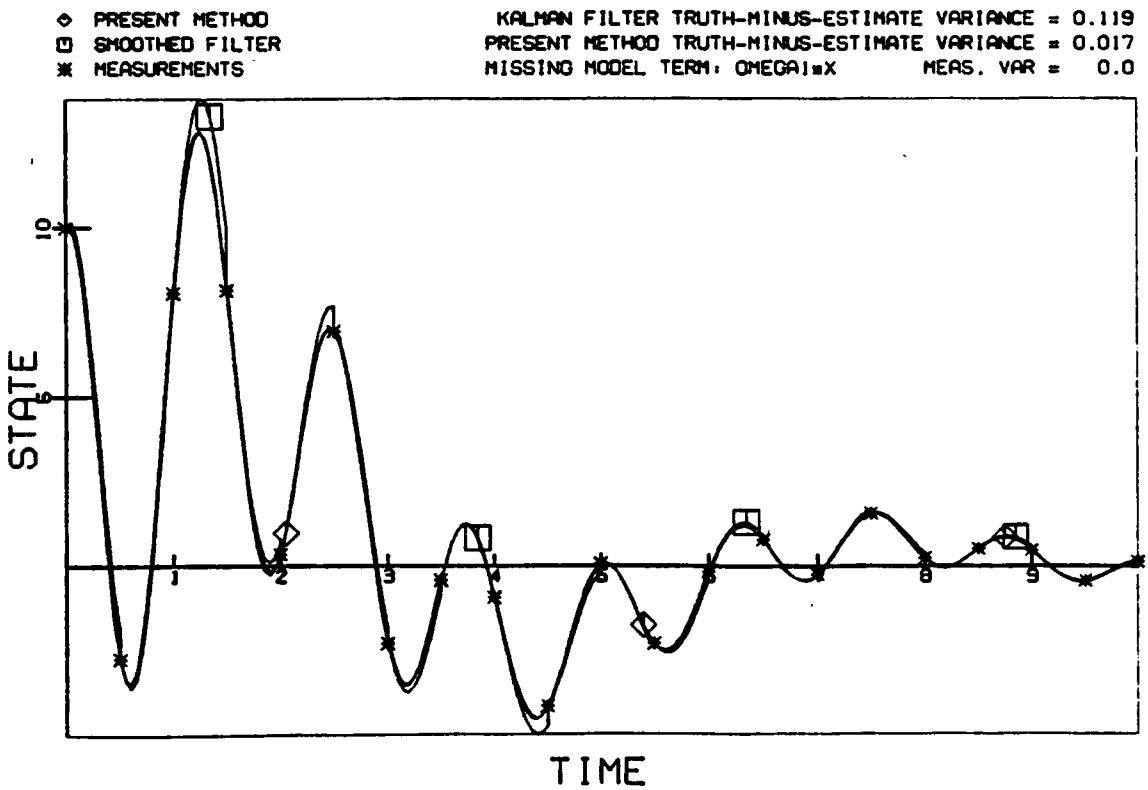


Figure 4.19-b Perfect Measurements

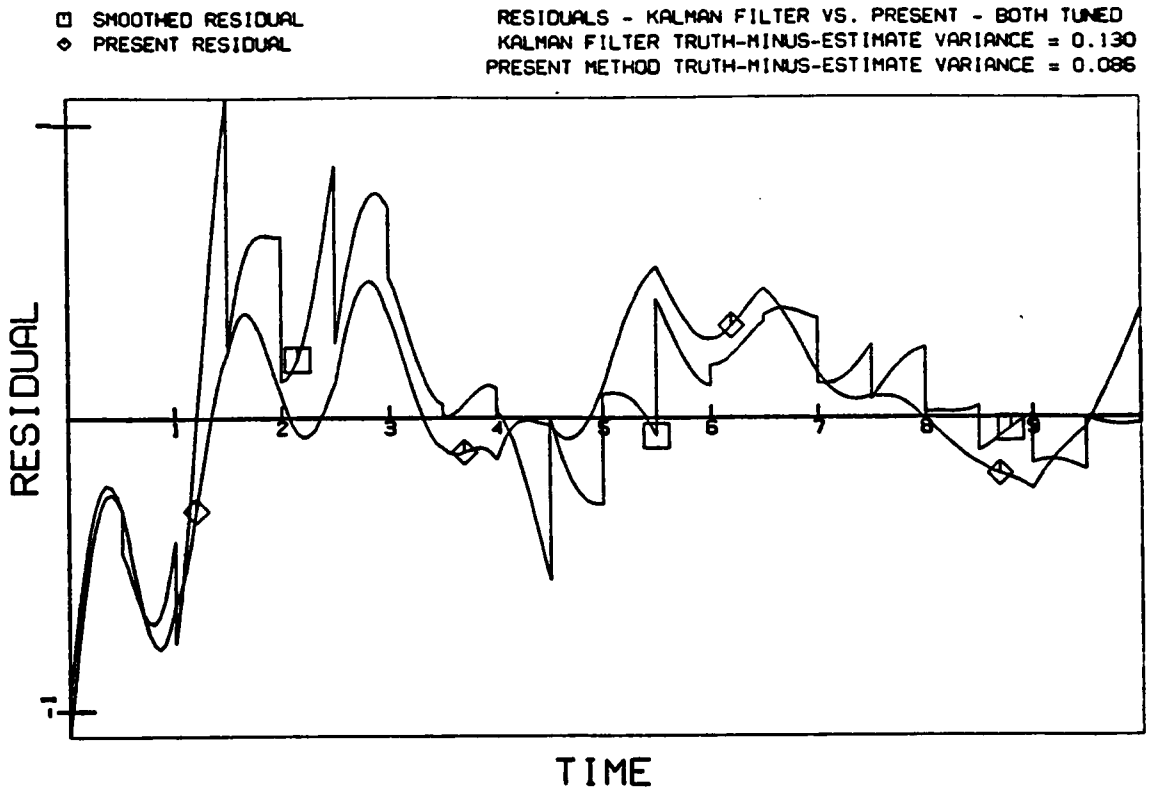


Figure 4.20-a Imperfect Measurements

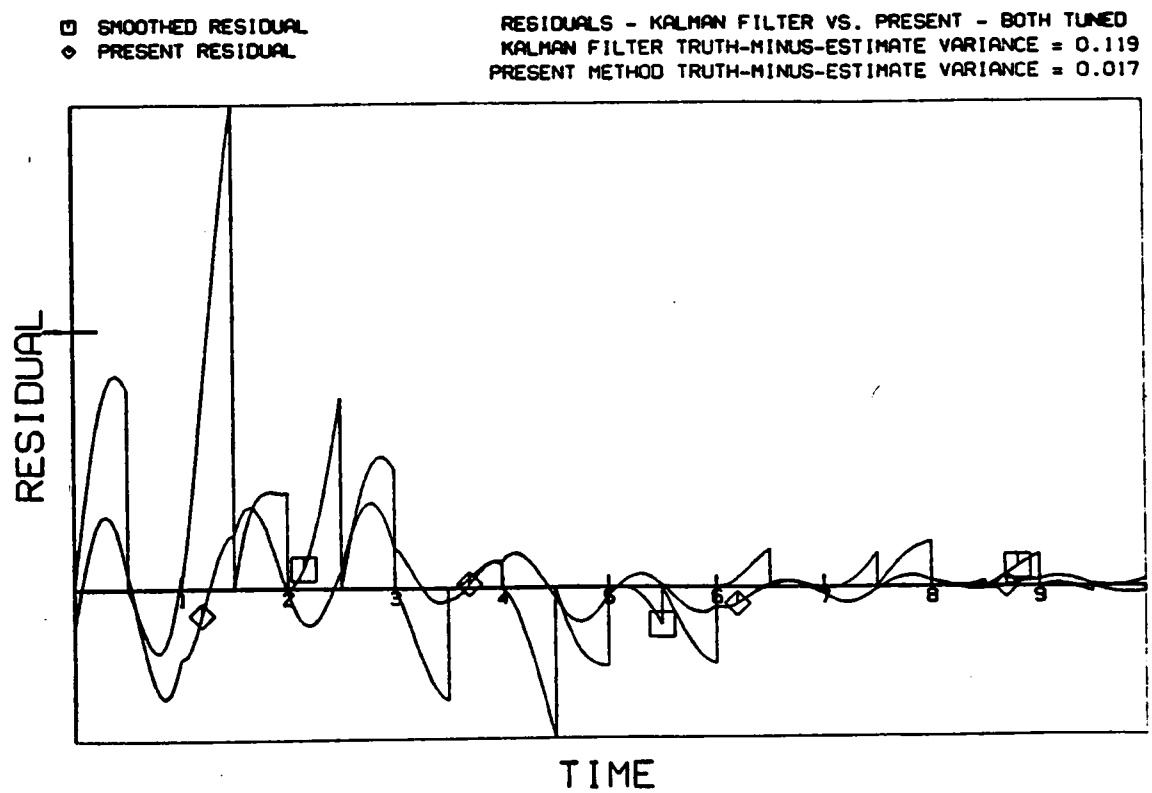


Figure 4.20-b Perfect Measurements

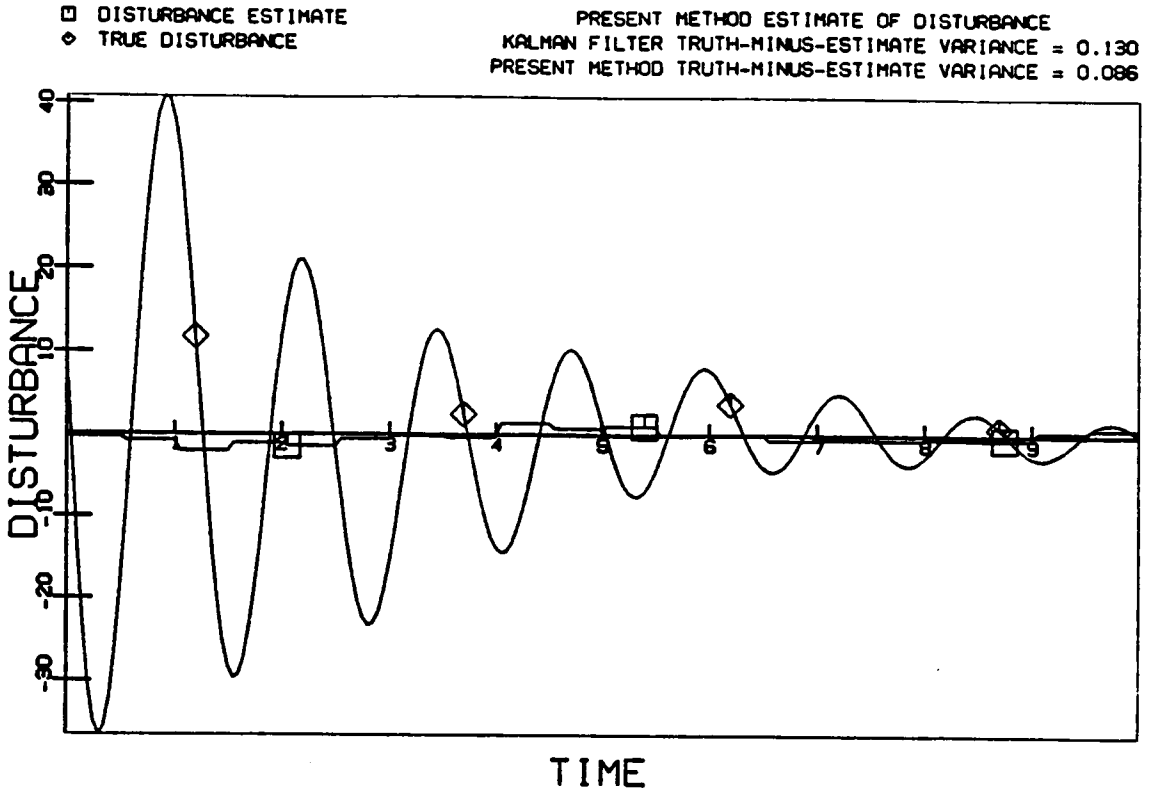


Figure 4.21-a Imperfect Measurements

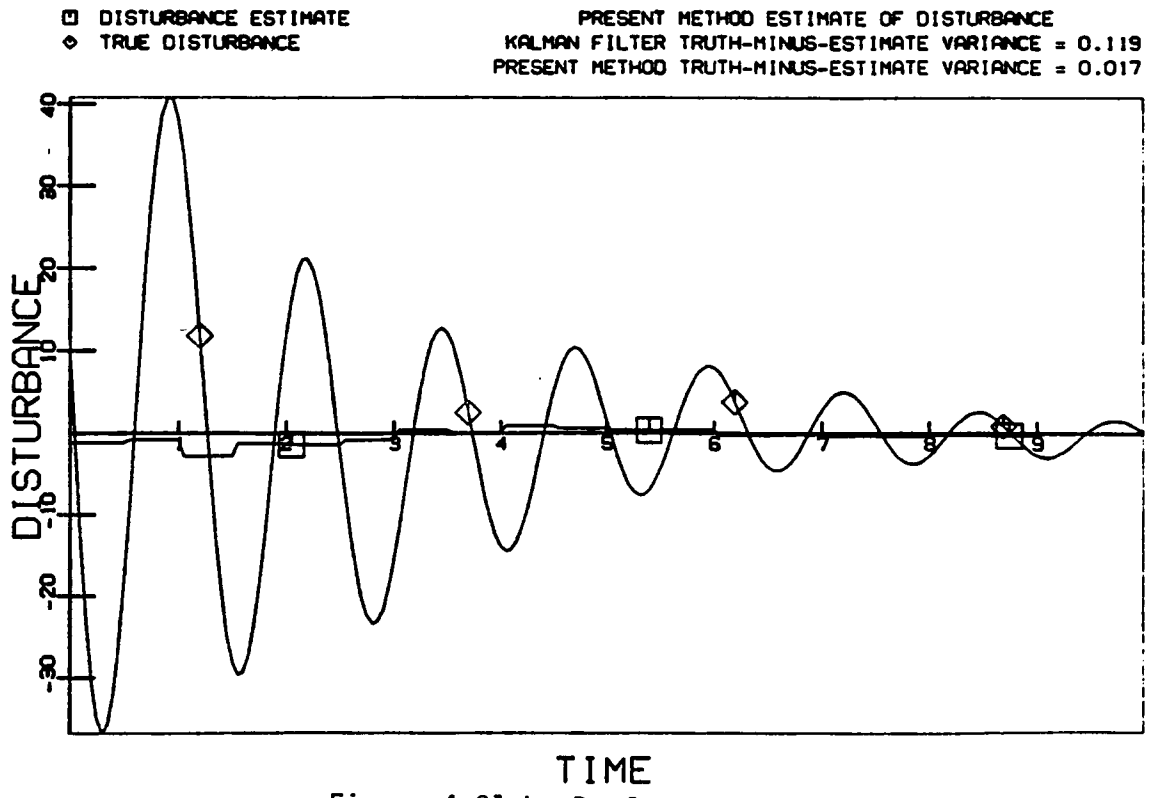


Figure 4.21-b Perfect Measurements

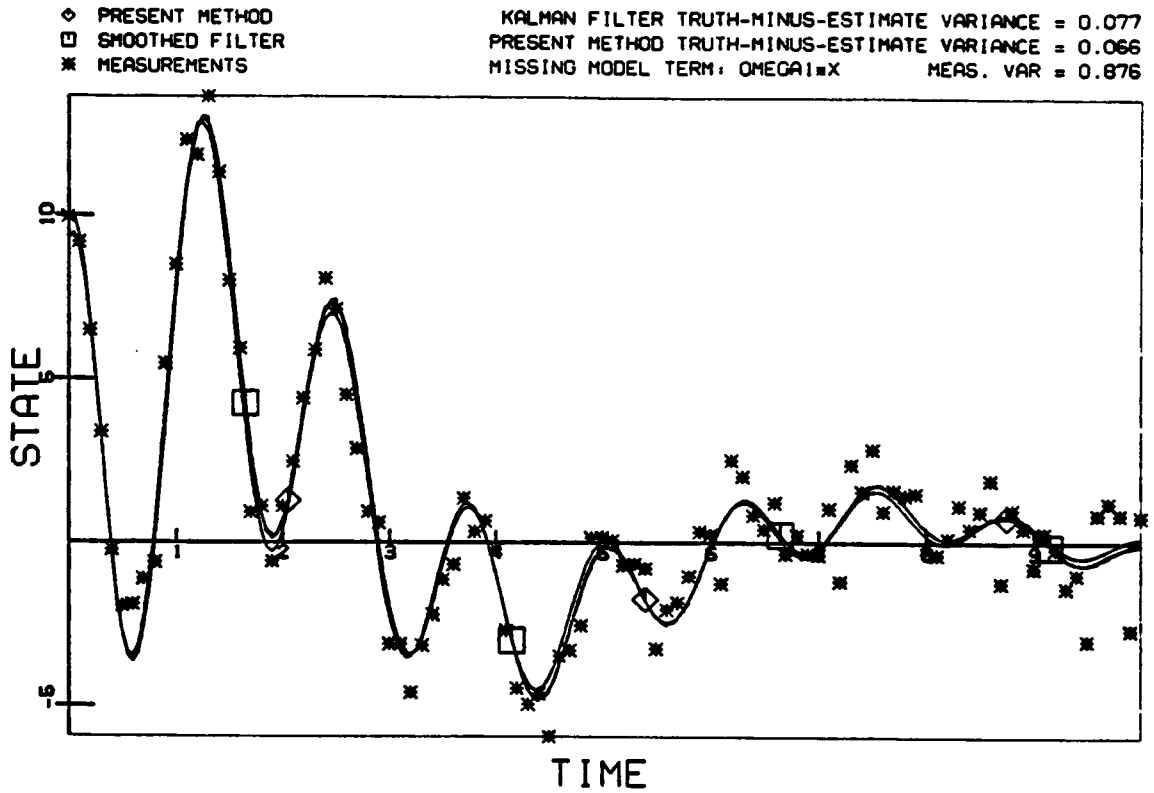


Figure 4.22-a Imperfect Measurements

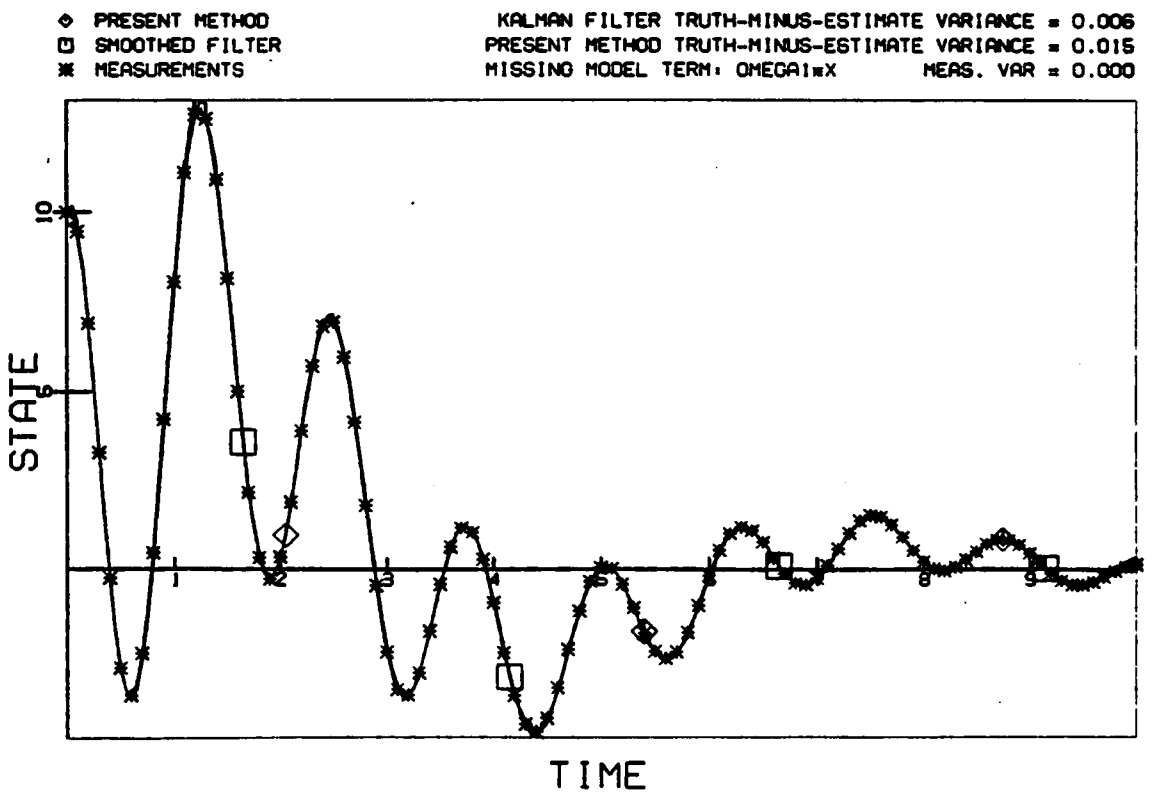


Figure 4.22-b Perfect Measurements

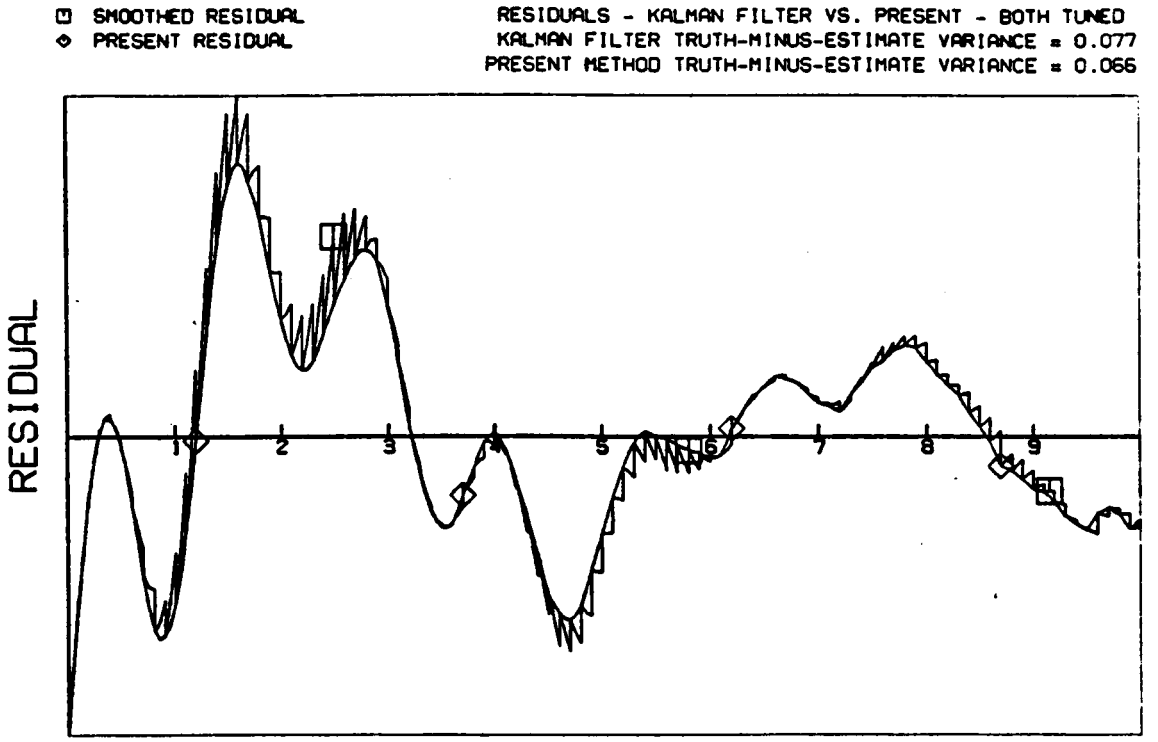


Figure 4.23-a Imperfect Measurements

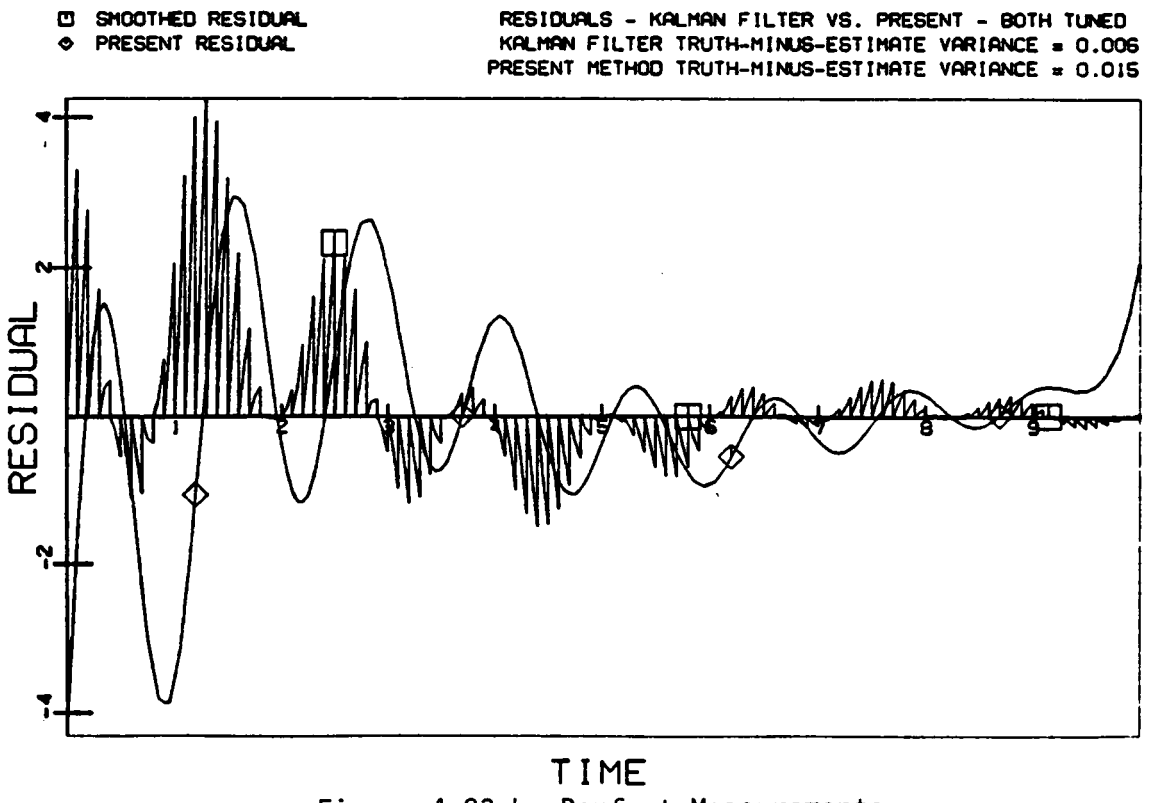


Figure 4.23-b Perfect Measurements

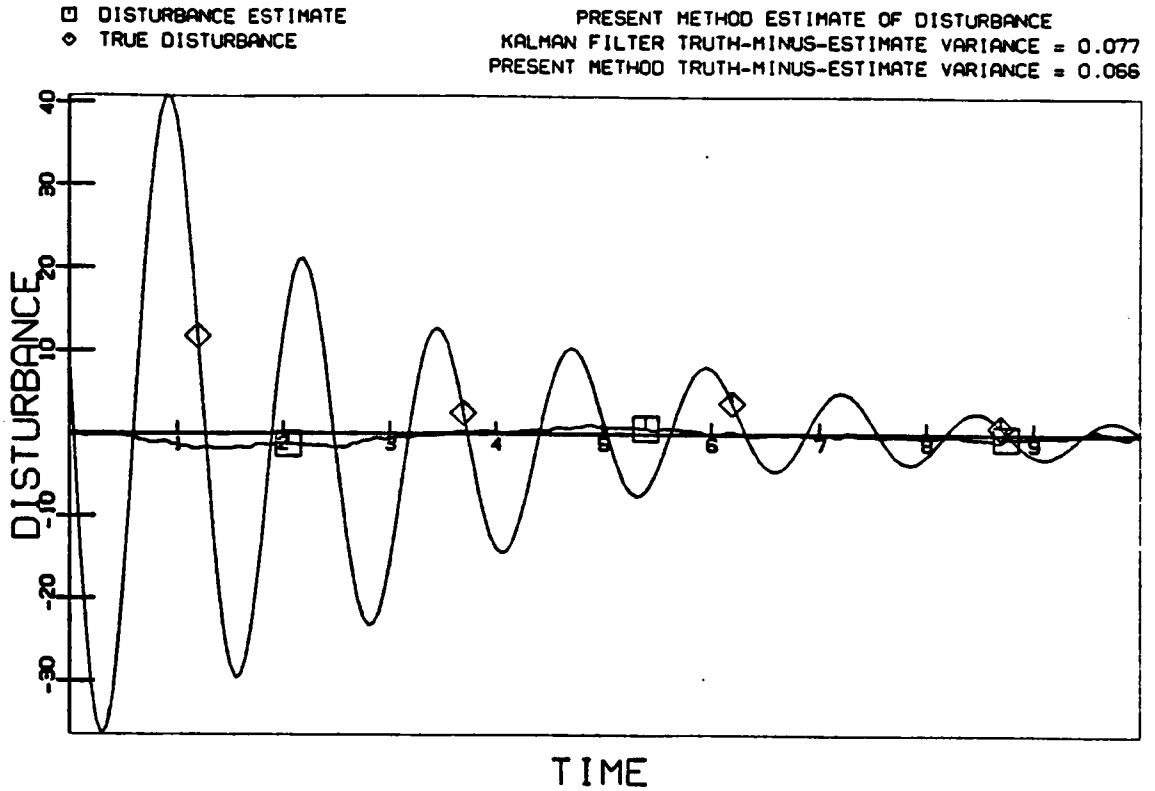


Figure 4.24-a Imperfect Measurements

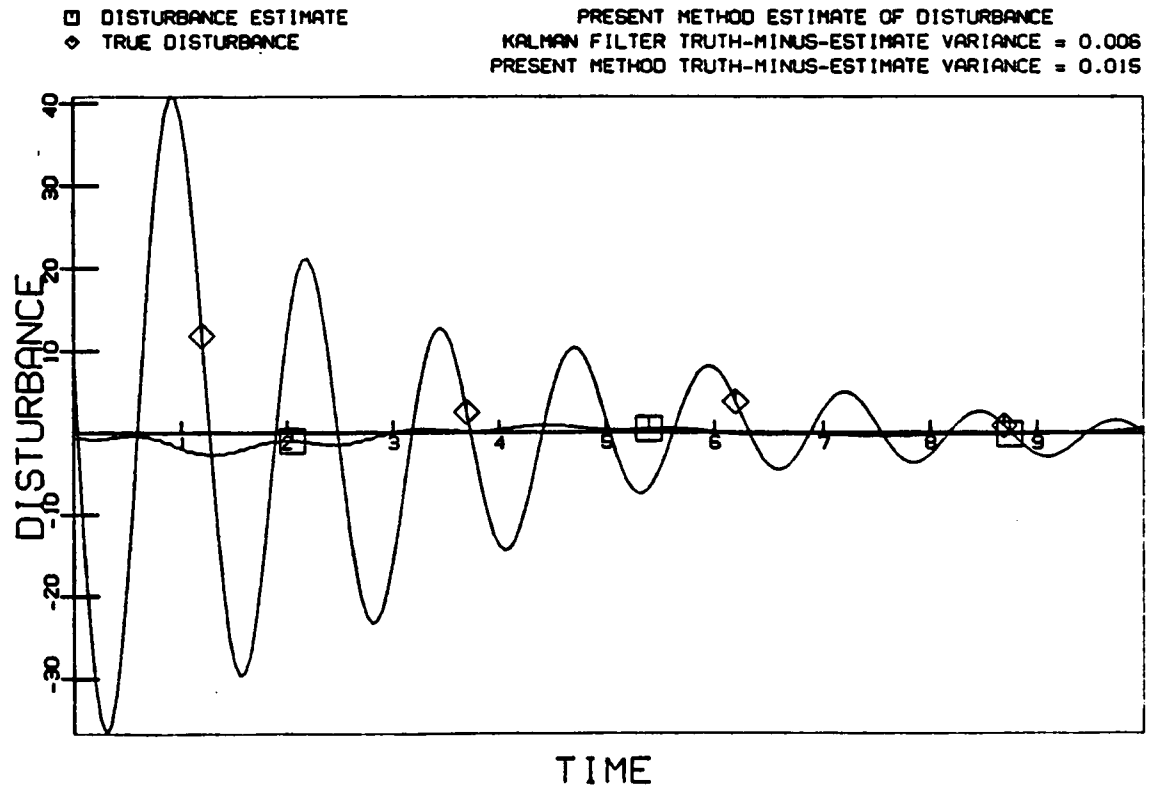


Figure 4.24-b Perfect Measurements

CHAPTER 5

SPACECRAFT ATTITUDE ESTIMATION

5.1 Introduction

In this chapter, some example problems in estimation of the attitude history of a rigid spacecraft are solved by both the present method and an extended Kalman filter algorithm.

The governing equations for the general three-dimensional rotation of a rigid body are given in Appendix A, along with coordinate system definitions. These equations are incorporated into a "truth model" program which is used to provide the measurements for the test cases. The details of the truth model program are also given in Appendix A.

For the present examples, the following assumptions are made:

(1) Angular velocity measurements are available at a frequency of twenty sets per second; each set includes $\tilde{\omega}_1$, $\tilde{\omega}_2$, and $\tilde{\omega}_3$.

(2) Attitude angle measurements are available at a frequency of one set every ten seconds; each set includes $\tilde{\psi}$, $\tilde{\theta}$, and $\tilde{\phi}$.

(3) The time span of interest is zero to thirty seconds; thus, there will be 601 angular velocity measurements sets and four attitude measurement sets.

(4) Measurement errors are zero mean and gaussian.

The unmodeled disturbance terms in the present method are added to the angular acceleration equations only. Thus, although there are six states and six governing equations, only three unmodeled effects terms are included. The reason for this is that forces (moments) are only

present in the acceleration equations; the velocity equations are based upon geometry and are considered to be exact.

There is an infinite number of possible test cases for spacecraft attitude problems, since every set of initial conditions produces a unique state history. The test cases presented in this chapter are not necessarily representative of any known or planned spacecraft orbit; however, they are all calculated using the exact equations of motion presented in Appendix A. The measurement accuracy and frequency are not varied in this problem; instead, values are used which are representative of existing hardware. For the angular velocity measuring devices, a sampling rate of twenty times per second and an error variance of $(1 \text{ arc-sec/sec})^2$ $((5 \times 10^{-6} \text{ rad/sec})^2)$ are used. For the attitude angle measuring devices, the sampling rate is once every ten seconds with an error variance of $(5 \text{ arc-sec})^2$ $((25 \times 10^{-6} \text{ rad})^2)$.

The figures which appear in this chapter show either the true orbit or the estimated residuals. Both algorithms generate such accurate solutions that they are indistinguishable from the truth in a plot, so no attempt is made to plot the state estimates. The variances reported in the figure captions are the averages of the three quantities plotted (either angles or angular velocities).

5.2 The Examples

The examples compare the accuracy of estimates obtained using both the present method and the extended Kalman filter on various problems which include unmodeled disturbances in the angular acceleration equations. However, to provide a "starting point", an orbit with no

unmodeled disturbance is shown in Fig. (5.1). The extended Kalman filter and present method estimates are both highly accurate; the residuals are plotted in Fig. (5.2). The present method obtains estimates with lower variances than the extended Kalman filter for both attitude and angular velocity.

In Figs. (5.3)-(5.7), the same initial conditions are used, but an unmodeled disturbance is added to the $\dot{\omega}_2$ governing equation as

$$\dot{\omega}_2 = \frac{(I_3 - I_1)}{I_2} \omega_1\omega_3 + A \sin \Omega t \qquad (5.3.1)$$

Figure (5.3) shows the residuals for the case $A = 10^{-7}$, $\Omega = 1$. Note that this magnitude of disturbance is 1/50 the standard deviation of the measurement noise. The variances of the estimates for both methods are appropriately the same as in the no disturbance example.

Figure (5.4) shows the residuals for the case $A = 10^{-6}$, which is 1/5 the standard deviation of the measurement noise. Again, the present method estimate has a smaller error variance than the extended Kalman filter.

In Fig. (5.5), the results for the case when $A = 10^{-5}$ are shown. Now the disturbance has a larger magnitude than the measurement error. The extended Kalman filter angular velocity estimate is now more accurate than the present method, reflecting the ability of the extended Kalman filter to match the measurements more accurately than the present method. The angular velocity measurements are very dense, and for this disturbance, the measurement accuracy is better than the model accuracy. Thus, by following the measurements closely, the extended Kalman filter is better able to estimate the angular velocity. Note

that the largest errors in both methods are in ω_2, which reflects the unmodeled disturbance d_2.

The present method obtains better estimates of the attitude history; however, this result will depend on the particular sample of attitude measurement errors. Since attitude is measured only four times, the extended Kalman filter estimate is sensitive to the particular measurement errors.

In Fig. (5.6), the true orbit histories for the case $A = 10^{-4}$ are shown. For this disturbance, the orbit histories have detectable differences (to the "naked eye") from the nominal histories shown in Fig. (5.1). The residuals from the estimates are shown in Fig. (5.7). The extended Kalman filter variance for angular velocity is two orders of magnitude lower than that of the present method. Again, the presence of dense, accurate measurements improves the extended Kalman filter solution. The extended Kalman filter attitude variance is also lower than that of the present method. This is a reflection of the more accurate angular velocity estimates, but is still somewhat dependent on the particular set of attitude measurement errors.

Finally, results are shown when unmodeled terms are added to each of the three angular acceleration equations as

$$\begin{aligned}
\dot{\omega}_1 &= \frac{I_2 - I_3}{I_1}\,\omega_2\omega_3 + A_1 \sin \Omega_1 t \\
\dot{\omega}_2 &= \frac{I_3 - I_1}{I_2}\,\omega_3\omega_1 + A_2 \sin \Omega_2 t \qquad (5.3.2) \\
\dot{\omega}_3 &= \frac{I_1 - I_2}{I_3}\,\omega_1\omega_2 + A_3 \sin \Omega_3 t
\end{aligned}$$

In Fig. (5.8), the orbit histories obtained with $A_1 = A_2 = A_3 = 10^{-5}$ and $\Omega_1 = \Omega_2 = \Omega_3 = 1$ are shown. The residuals of the estimates are shown in Fig. (5.9). The extended Kalman filter angular velocity estimate has lower variance than the present method estimate, again reflecting the dense, accurate measurements. The present method attitude estimate has lower variance than the extended Kalman filter attitude estimate; again, this depends upon the particular error sample.

5.3 Summary

In this chapter, the present method is compared with an extended Kalman filter for performance in estimating some spacecraft attitude histories. Various levels of disturbances are added to the angular acceleration equations, including cases well below the measurement noise and cases well above the measurement noise. In those cases where the disturbance amplitude is less than the measured error amplitude the present method obtains more accurate estimates of both attitude and angular velocity than does the extended Kalman filter. When the disturbance amplitudes are higher than the measurement error amplitudes, the extended Kalman filter is able to obtain more accurate angular velocity estimates, due to the dense, accurate measurements. The results for attitude estimation are less conclusive, due to the low density of measurements which limits the extended Kalman filter accuracy and makes it more sensitive to he particular sample of measurement errors.

The present method does not obtain accurate recoveries of the true unmodeled disturbance terms. Apparently, the presence of unmodeled

effect terms in each angular velocity equation allows the true disturbance to be "spread out"; consequently, the recovered disturbance estimates are not accurate. Nevertheless, the attitude estimates are comparable to the extended Kalman filter estimates, although the rate estimates are not.

The examples of this chapter are presented for illustrative purposes and do not constitute a comprehensive comparison for the spacecraft attitude history estimation problem. Many variations of both algorithms are available; it is beyond the scope of the present work to investigate all of these options. For example, in this study the "tuning parameters" (Q in the filter, w in the present method) are taken to be scalar multiples of the identity matrix although in general, they may be arbitrary positive definite matrices. For practical problems, longer estimation time intervals are more realistic. The additional measurements of the longer intervals may help to alleviate the small sample anomalies in the attitude measurements; otherwise, a "monte carlo" study may be required in order to obtain conclusive comparisons. All of these issues remain to be addressed in other studies.

As a final note, the plots of the residuals highlight the jump discontinuities of the filter estimates. Although the present method estimates are not always as accurate, they are smooth.

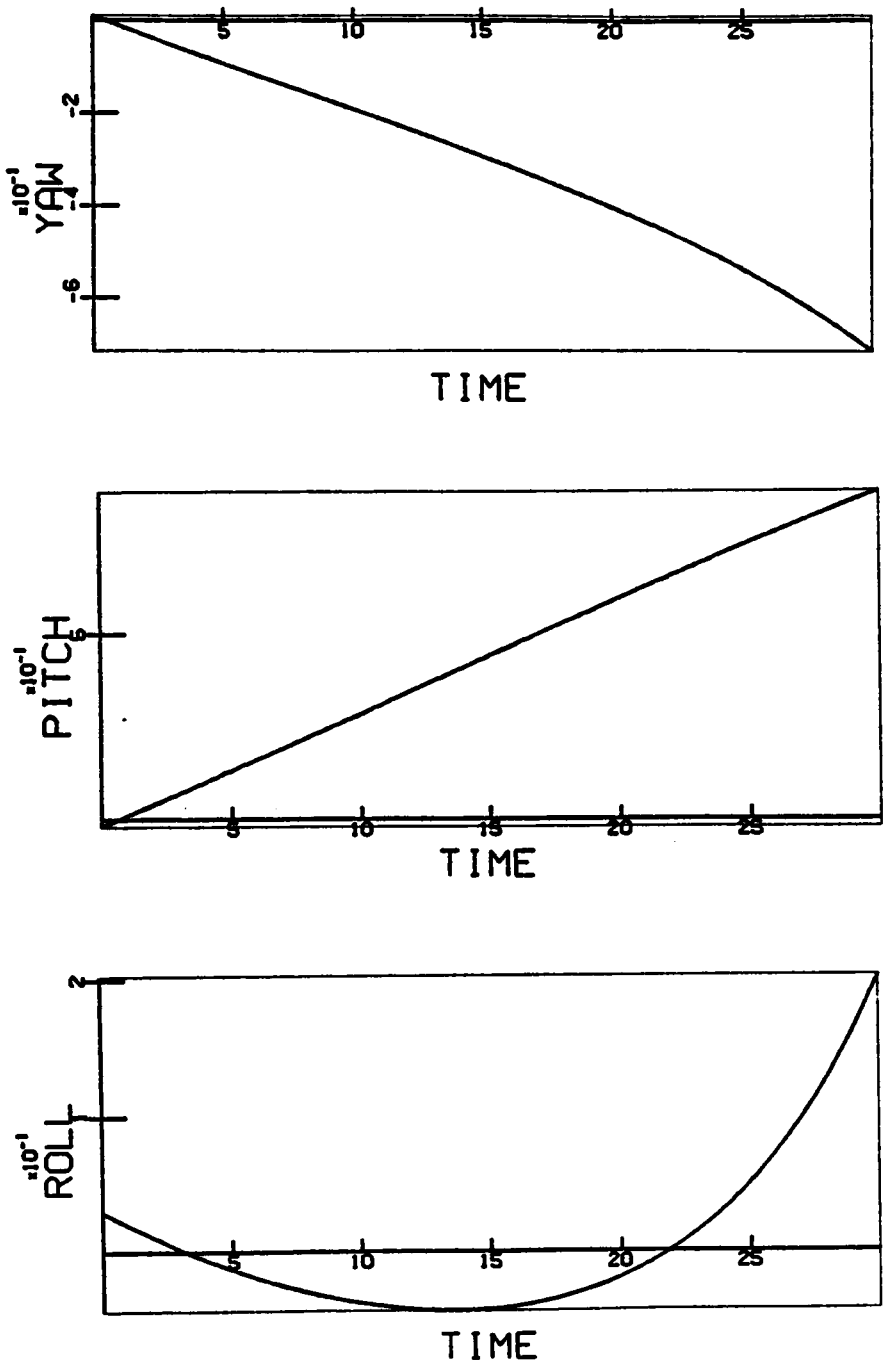

5
10
15
20
25
-2
-4
-6
x10⁻¹
YAW
TIME
6
x10⁻¹
PITCH
TIME
2
x10⁻¹
ROLL
TIME

Figure 5.1a

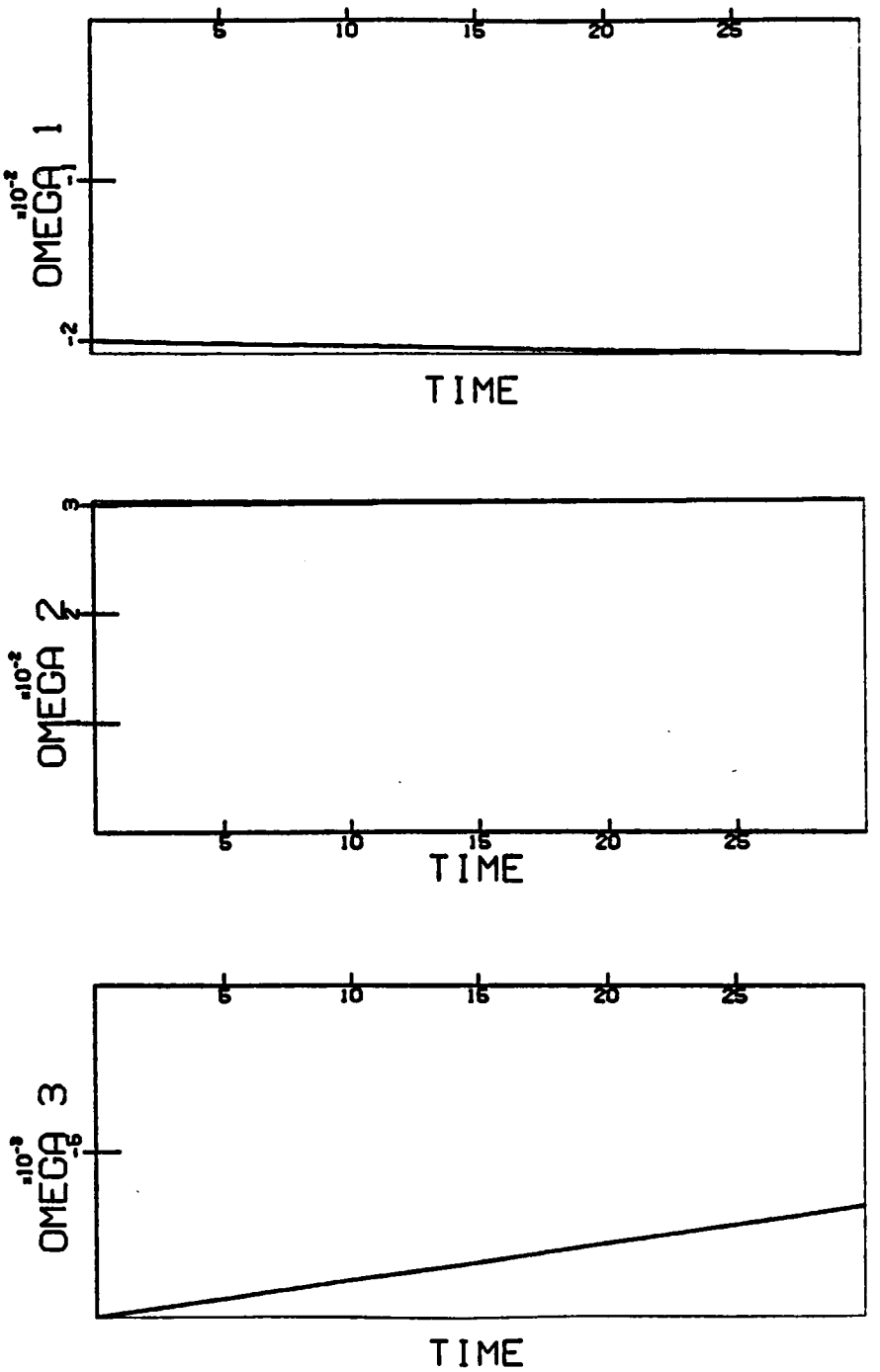
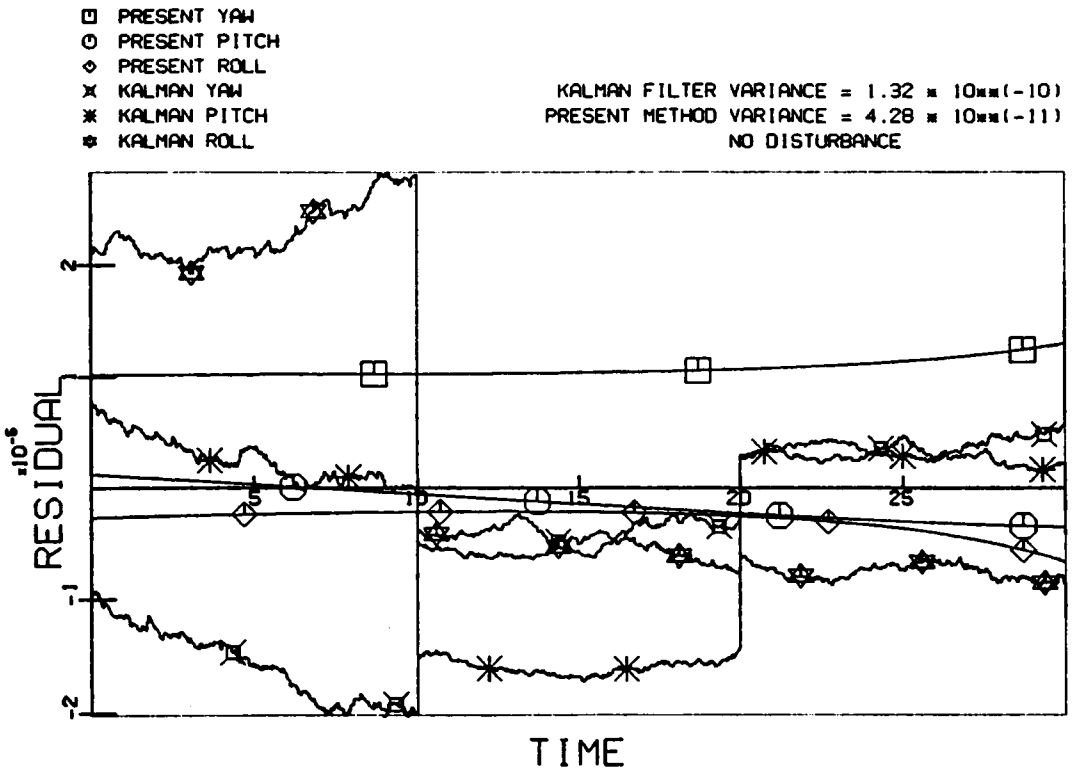

5
10
15
20
25
OMEGA 1
.10⁻²
-1
-2
TIME
3
2
OMEGA 2
.10⁻²
5
10
15
20
25
TIME
5
10
15
20
25
OMEGA 3
.10⁻³
-5
TIME

Figure 5.1b


PRESENT YAW
PRESENT PITCH
PRESENT ROLL
KALMAN YAW
KALMAN PITCH
KALMAN ROLL
KALMAN FILTER VARIANCE = 1.32 * 10**(-10)
PRESENT METHOD VARIANCE = 4.28 * 10**(-11)
NO DISTURBANCE
RESIDUAL
*10**-5
2
1
-1
-2
5
10
15
20
25
TIME

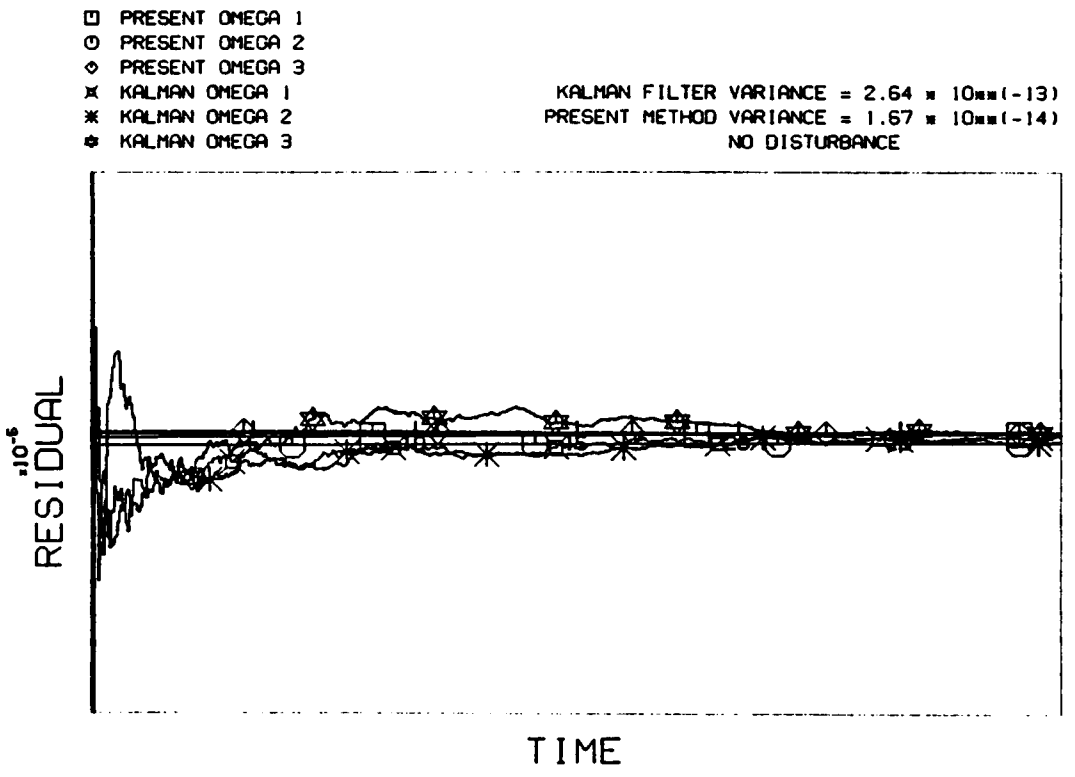
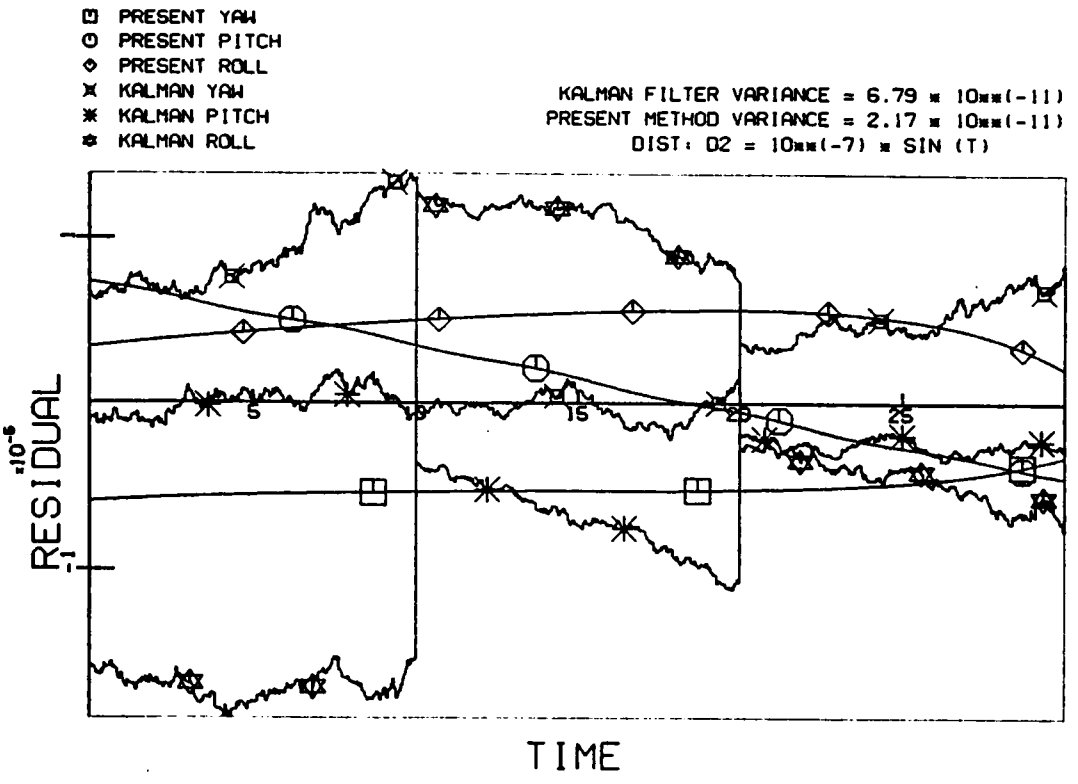

PRESENT OMEGA 1
PRESENT OMEGA 2
PRESENT OMEGA 3
KALMAN OMEGA 1
KALMAN OMEGA 2
KALMAN OMEGA 3
KALMAN FILTER VARIANCE = 2.64 * 10**(-13)
PRESENT METHOD VARIANCE = 1.67 * 10**(-14)
NO DISTURBANCE
RESIDUAL
*10**-5
TIME

Figure 5.2


PRESENT YAW
PRESENT PITCH
PRESENT ROLL
KALMAN YAW
KALMAN PITCH
KALMAN ROLL
KALMAN FILTER VARIANCE = 6.79 * 10**(-11)
PRESENT METHOD VARIANCE = 2.17 * 10**(-11)
DIST: D2 = 10**(-7) * SIN (T)
RESIDUAL
x10**-6
-1
5
10
15
20
25
TIME

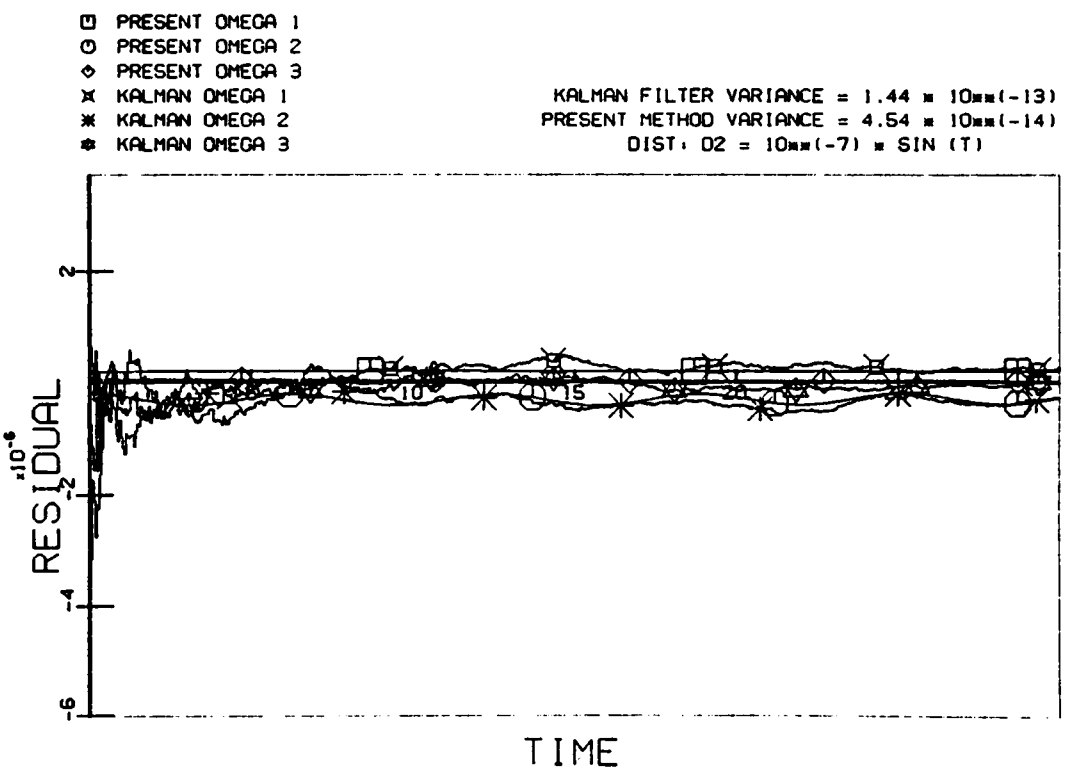
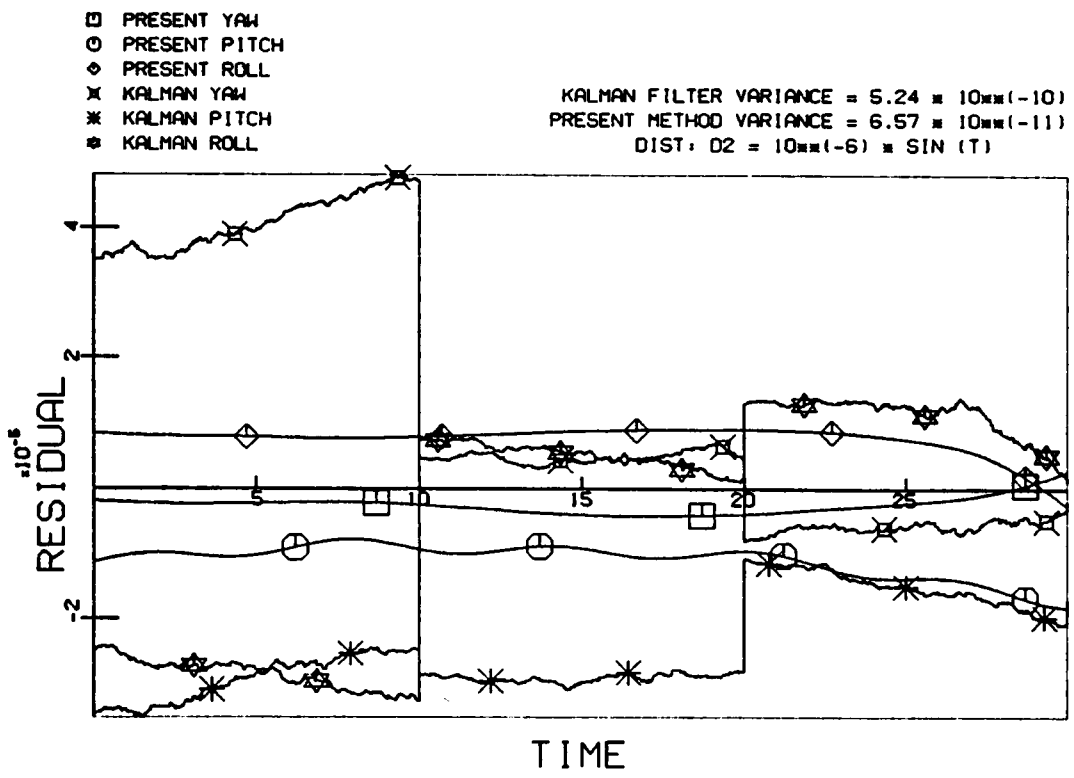

PRESENT OMEGA 1
PRESENT OMEGA 2
PRESENT OMEGA 3
KALMAN OMEGA 1
KALMAN OMEGA 2
KALMAN OMEGA 3
KALMAN FILTER VARIANCE = 1.44 * 10**(-13)
PRESENT METHOD VARIANCE = 4.54 * 10**(-14)
DIST: D2 = 10**(-7) * SIN (T)
RESIDUAL
x10**-6
2
-2
-4
-6
10
15
20
TIME

Figure 5.3


PRESENT YAW
PRESENT PITCH
PRESENT ROLL
KALMAN YAW
KALMAN PITCH
KALMAN ROLL
KALMAN FILTER VARIANCE = 5.24 * 10**(-10)
PRESENT METHOD VARIANCE = 6.57 * 10**(-11)
DIST: D2 = 10**(-6) * SIN (T)
RESIDUAL
x10^-6
4
2
-2
5
10
15
20
25
TIME

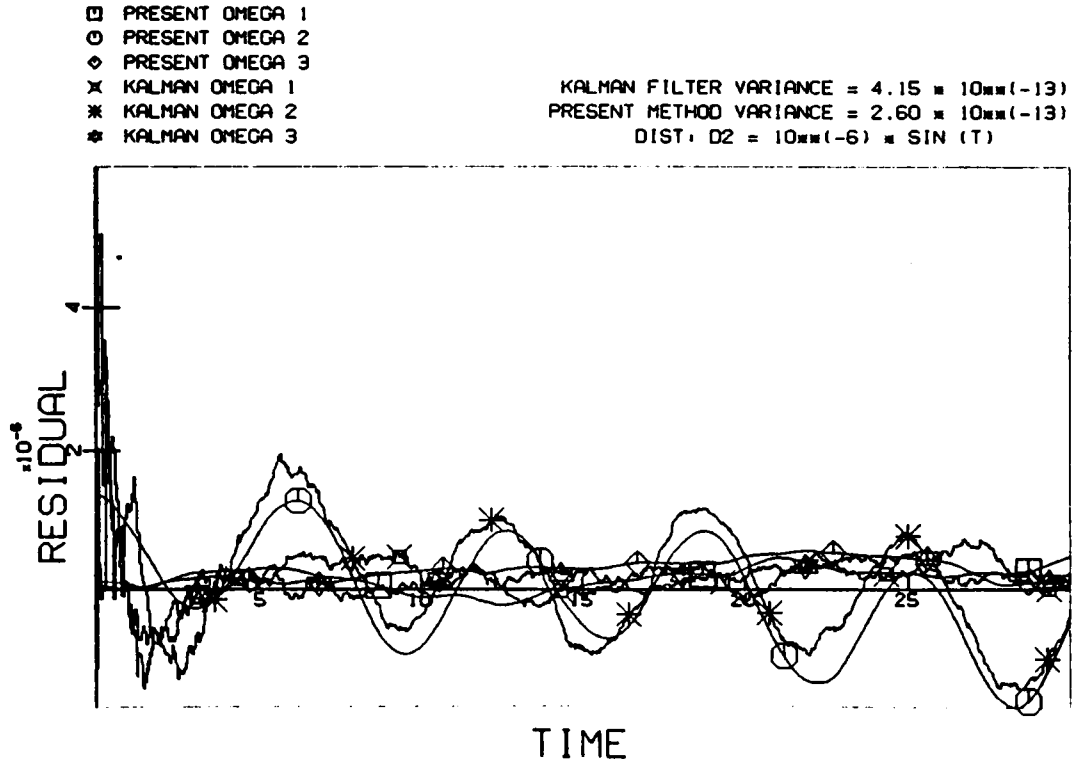
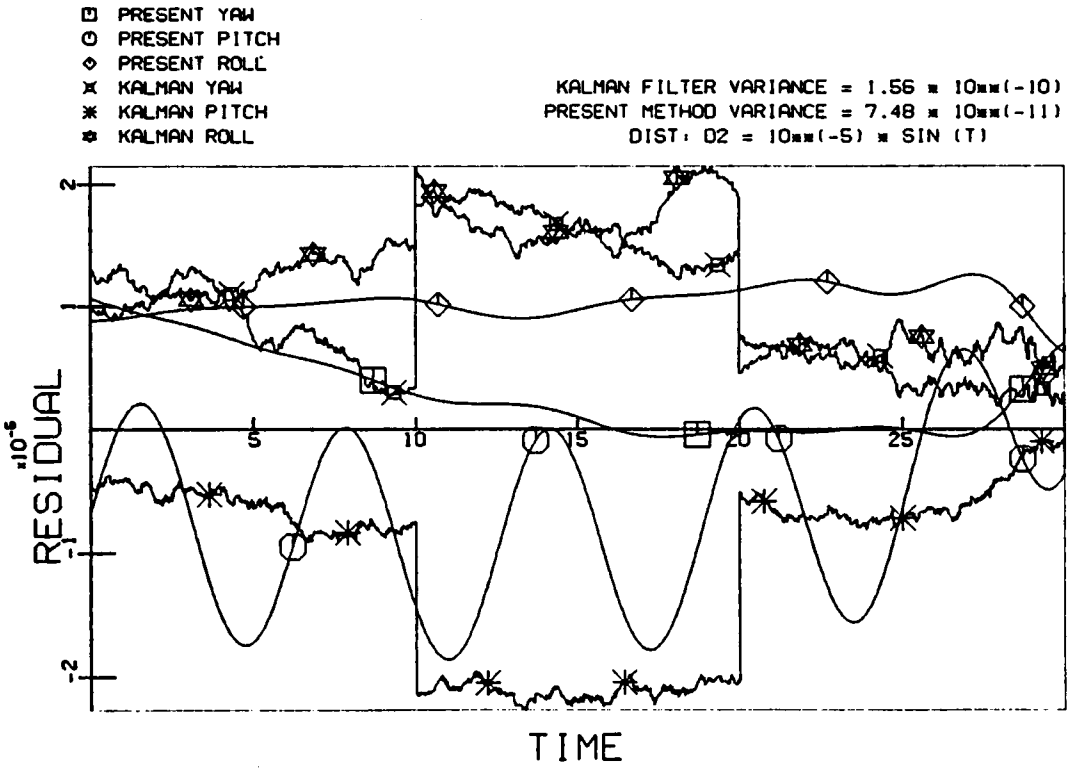

PRESENT OMEGA 1
PRESENT OMEGA 2
PRESENT OMEGA 3
KALMAN OMEGA 1
KALMAN OMEGA 2
KALMAN OMEGA 3
KALMAN FILTER VARIANCE = 4.15 * 10**(-13)
PRESENT METHOD VARIANCE = 2.60 * 10**(-13)
DIST: D2 = 10**(-6) * SIN (T)
RESIDUAL
x10^-6
4
2
5
10
15
20
25
TIME

Figure 5.4


PRESENT YAW
PRESENT PITCH
PRESENT ROLL
KALMAN YAW
KALMAN PITCH
KALMAN ROLL
KALMAN FILTER VARIANCE = 1.56 * 10**(-10)
PRESENT METHOD VARIANCE = 7.48 * 10**(-11)
DIST: D2 = 10**(-5) * SIN (T)
RESIDUAL
*10**-5
2
1
-1
-2
5
10
15
20
25
TIME

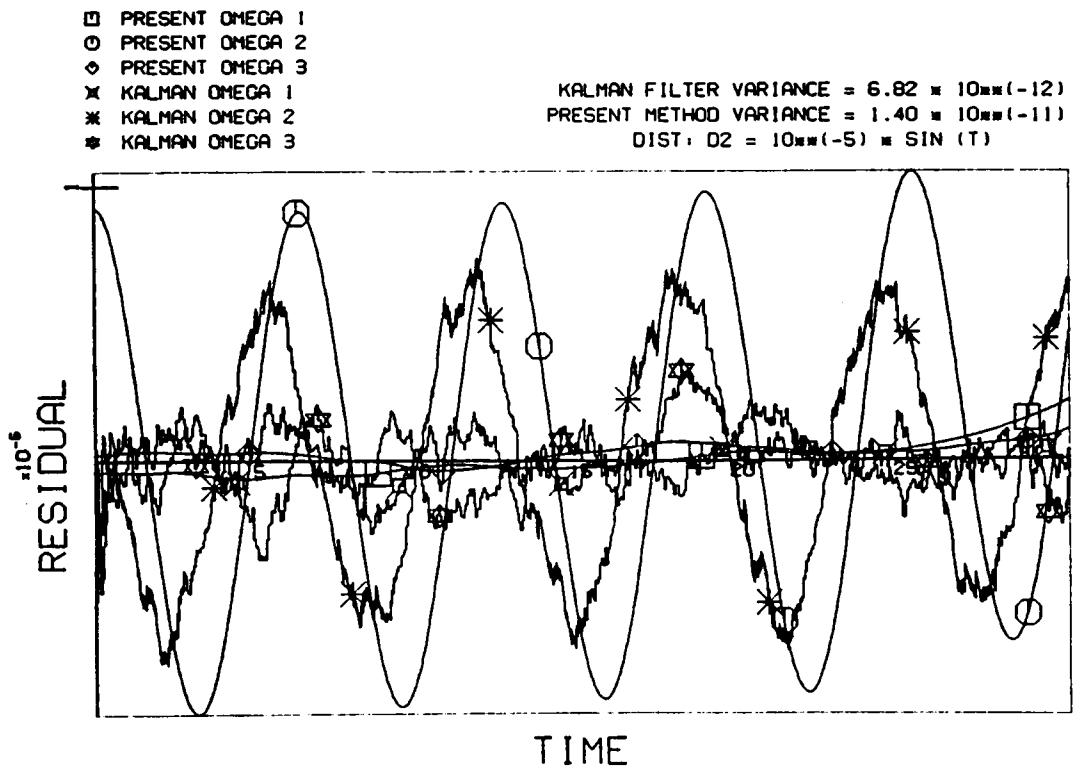

PRESENT OMEGA 1
PRESENT OMEGA 2
PRESENT OMEGA 3
KALMAN OMEGA 1
KALMAN OMEGA 2
KALMAN OMEGA 3
KALMAN FILTER VARIANCE = 6.82 * 10**(-12)
PRESENT METHOD VARIANCE = 1.40 * 10**(-11)
DIST: D2 = 10**(-5) * SIN (T)
RESIDUAL
*10**-5
TIME

Figure 5.5

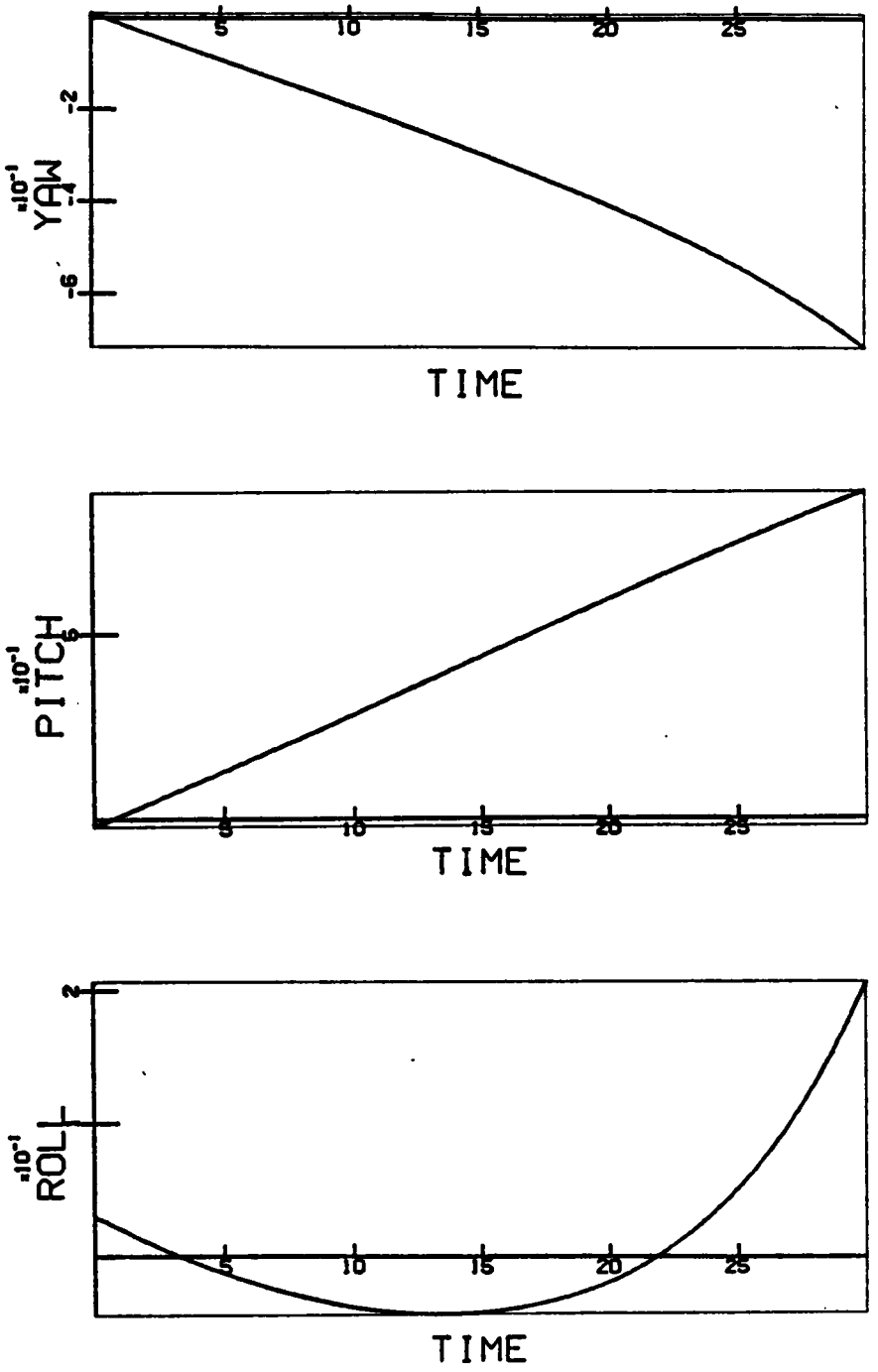


Figure 5.6a

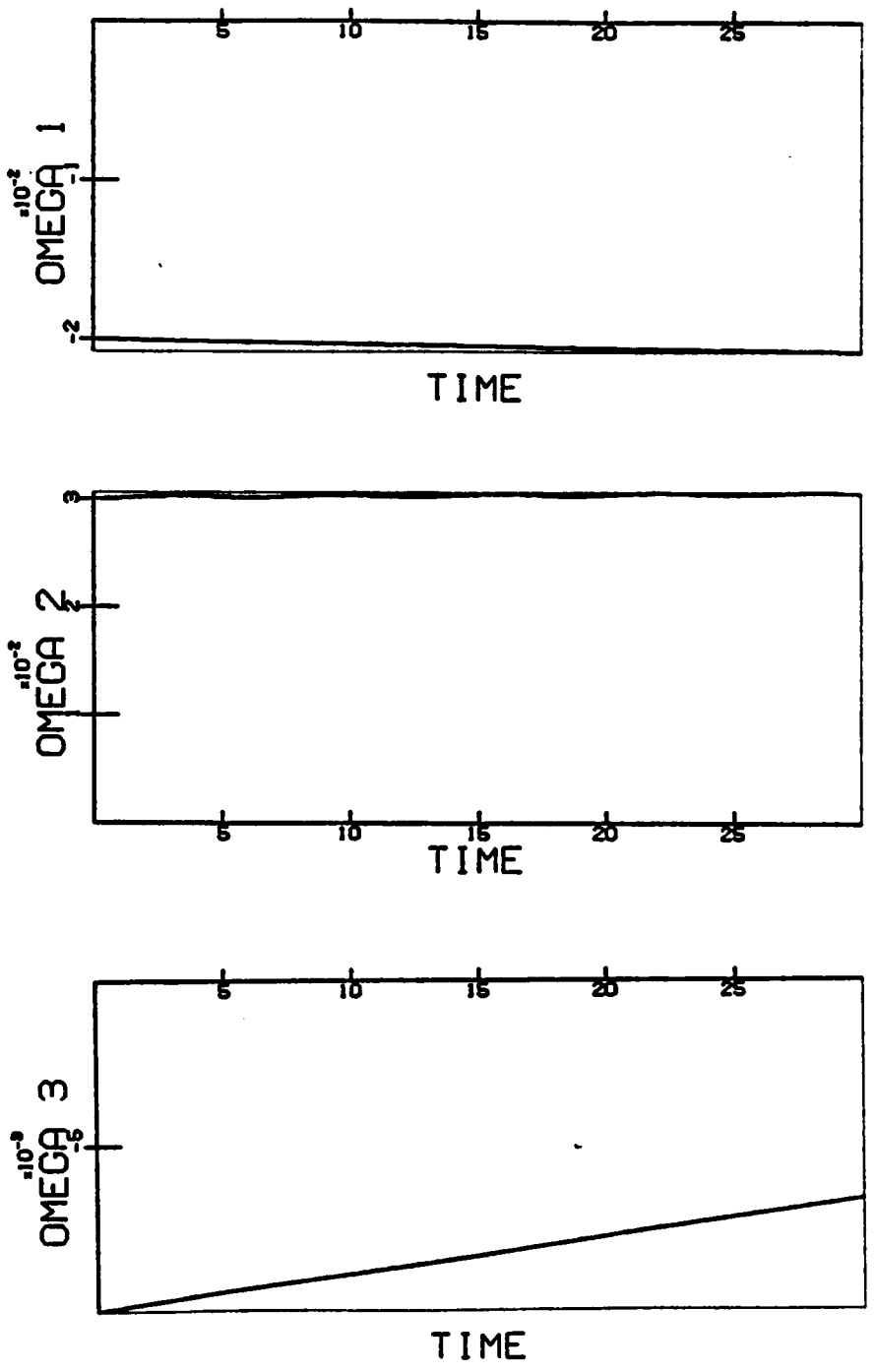
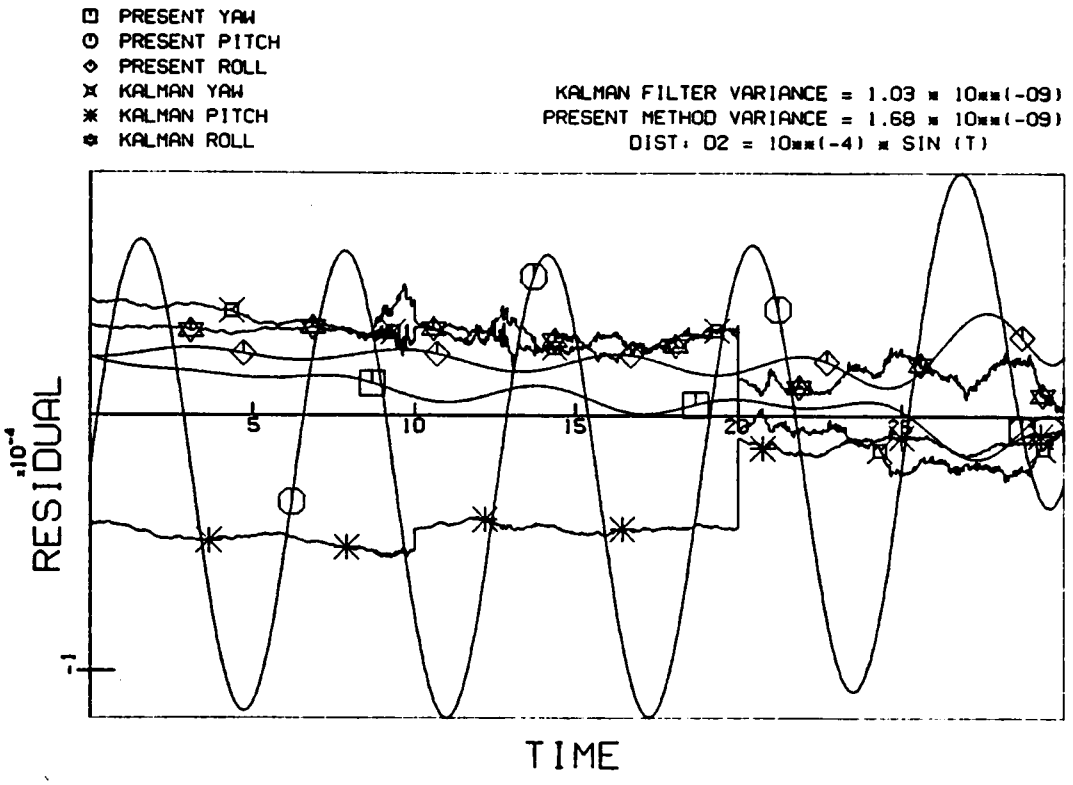

5
10
15
20
25
OMEGA 1
*10**-2
1
-2
TIME
3
2
OMEGA 2
*10**-2
1
5
10
15
20
25
TIME
5
10
15
20
25
OMEGA 3
*10**-3
-5
TIME

Figure 5.6b


PRESENT YAW
PRESENT PITCH
PRESENT ROLL
KALMAN YAW
KALMAN PITCH
KALMAN ROLL
KALMAN FILTER VARIANCE = 1.03 * 10**(-09)
PRESENT METHOD VARIANCE = 1.68 * 10**(-09)
DIST: D2 = 10**(-4) * SIN (T)
RESIDUAL
.10**-4
5
10
15
20
-1
TIME

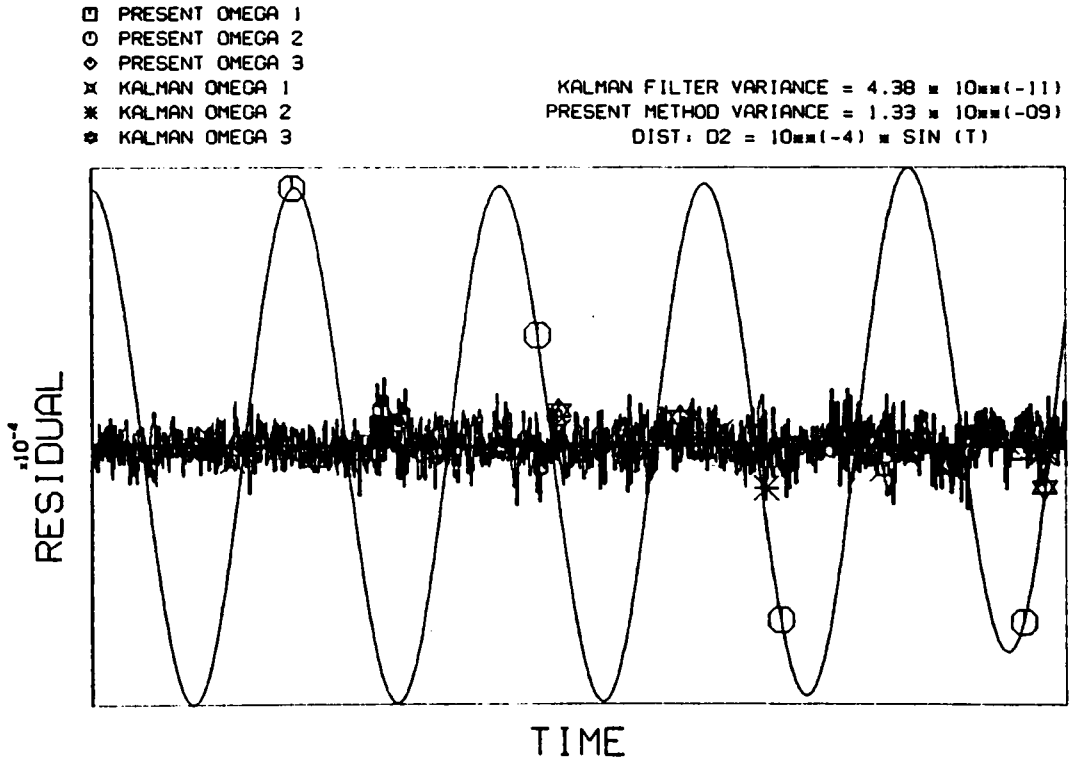

PRESENT OMEGA 1
PRESENT OMEGA 2
PRESENT OMEGA 3
KALMAN OMEGA 1
KALMAN OMEGA 2
KALMAN OMEGA 3
KALMAN FILTER VARIANCE = 4.38 * 10**(-11)
PRESENT METHOD VARIANCE = 1.33 * 10**(-09)
DIST: D2 = 10**(-4) * SIN (T)
RESIDUAL
.10**-4
TIME

Figure 5.7

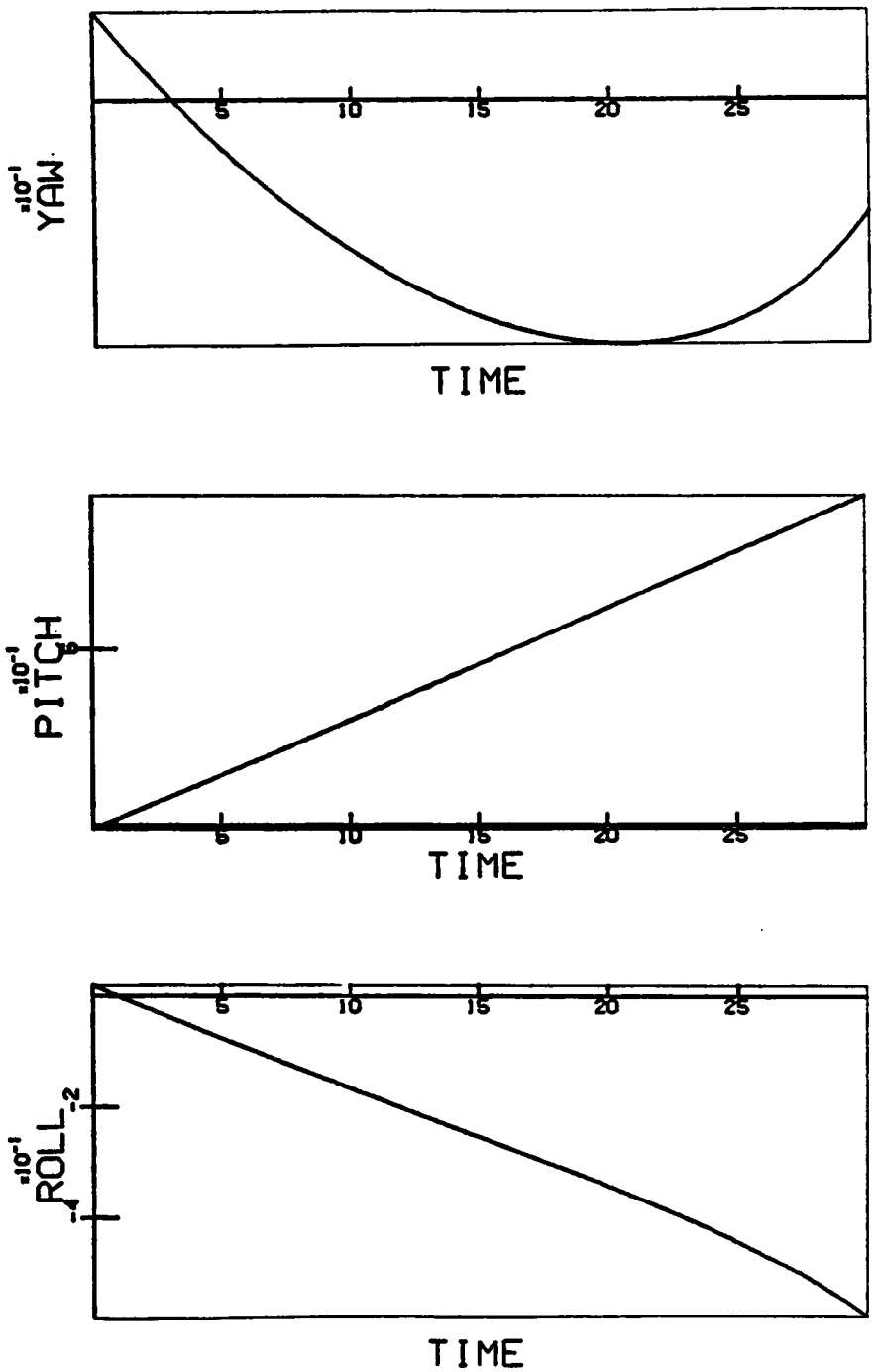


Figure 5.8a

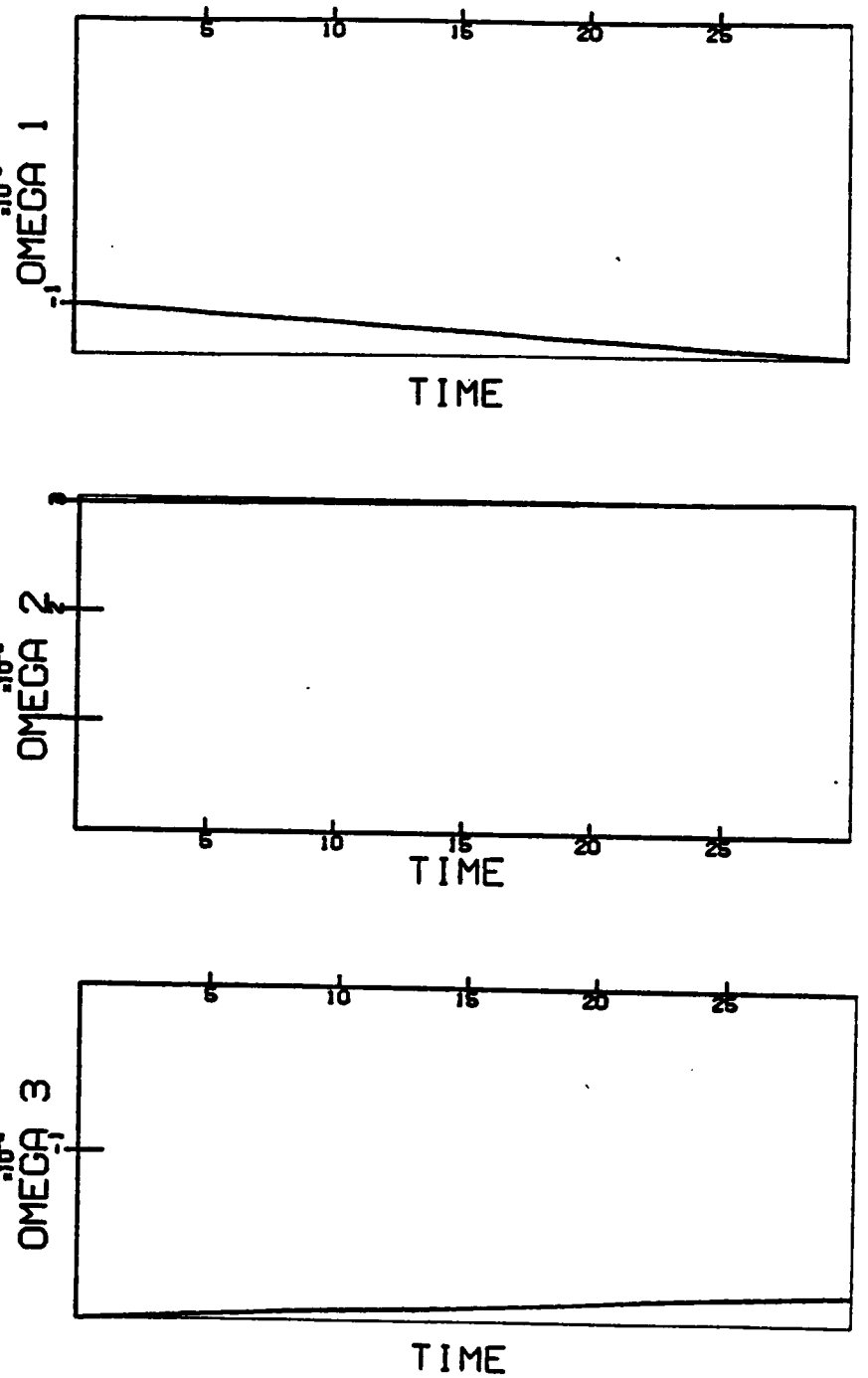
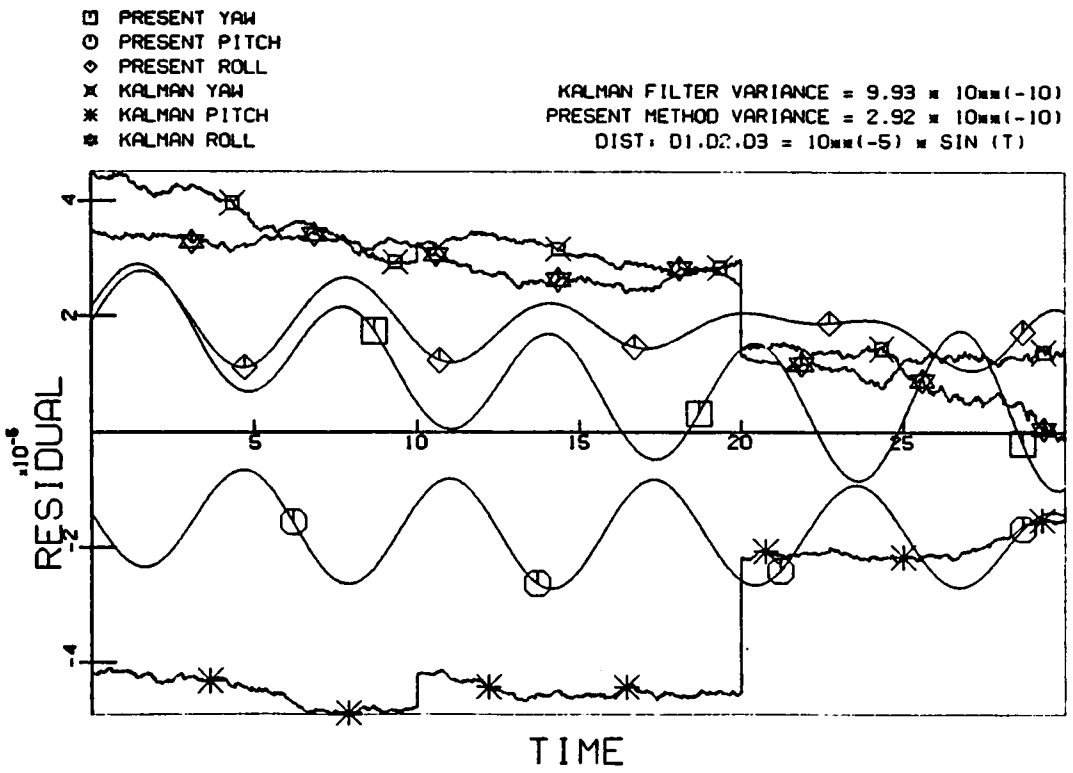


Figure 5.8b


PRESENT YAW
PRESENT PITCH
PRESENT ROLL
KALMAN YAW
KALMAN PITCH
KALMAN ROLL
KALMAN FILTER VARIANCE = 9.93 * 10**(-10)
PRESENT METHOD VARIANCE = 2.92 * 10**(-10)
DIST: D1.D2.D3 = 10**(-5) * SIN (T)
RESIDUAL
x10**-5
4
2
-2
-4
5
10
15
20
25
TIME

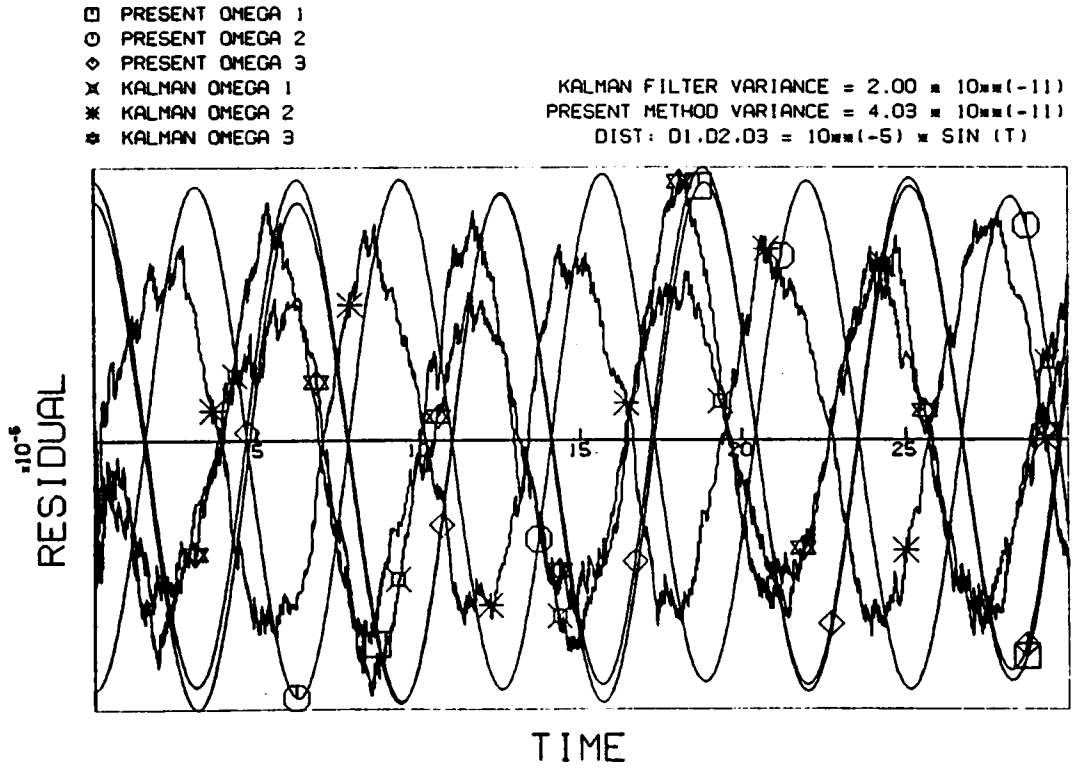

PRESENT OMEGA 1
PRESENT OMEGA 2
PRESENT OMEGA 3
KALMAN OMEGA 1
KALMAN OMEGA 2
KALMAN OMEGA 3
KALMAN FILTER VARIANCE = 2.00 * 10**(-11)
PRESENT METHOD VARIANCE = 4.03 * 10**(-11)
DIST: D1.D2.D3 = 10**(-5) * SIN (T)
RESIDUAL
x10**-5
5
15
20
25
TIME

Figure 5.9

CHAPTER 6

SUMMARY AND CONCLUSIONS

6.1 Summary

The most commonly used estimation strategies are the Kalman filter-type algorithms. These algorithms account for model errors in the system state governing equations via "process noise", which is assumed to be a white noise (zero mean). In the discussions and derivations of Chapter 3, the net effect of this approach was shown to be an increased dependence of the estimate on the measurement; no improvement in the model is obtained. Thus, this approach is heavily dependent on the frequency and the accuracy of the measurements. The results of some simple example problems in Chapter 4 show this dependence clearly.

A novel estimation strategy based on concepts from optimal control theory was derived in Chapter 2. This new method explicitly accounts for unmodeled effects in the system governing equations. Not only is the model improved, but an estimate of the disturbance is also obtained.

Furthermore, the present method yields a continuous estimated state history; the Kalman filter-type algorithms produce jump discontinuities at the measurement times. For many practical problems, jump discontinuities in the states are physically unrealizable (e.g., position estimation). Thus, the Kalman filter estimates are essentially undefined in small regions around each measurement time.

The two methods were compared for estimate accuracy on some simple example problems in Chapter 4, and on some spacecraft attitude estimation problems in Chapter 5.

6.2 Conclusions

The present method obtains a solution of much higher accuracy than the Kalman filter-type methods, for problems in which both model errors and low measurement frequency are present. The results of Chapter 4, where both measurement frequency and accuracy are varied, indicate that the Kalman filter-type algorithms are extremely sensitive to measurement frequency in the presence of unmodeled disturbances. Apparently, measurement in the presence of large model errors, frequency is more important than measurement accuracy. The new method shows a clear advantage when dealing with large model errors and sparse data, and is obviously competitive for the other extreme of dense data and small model errors.

The present method accuracy is also improved as the measurement frequency and accuracy are increased. However, it is not nearly as sensitive as the Kalman filter-type algorithms to the size of the model error and to the frequency of the measurements.

If the measurements are dense and accurate enough, the Kalman filter-type algorithms are superior to the present method algorithm if model error is present (although not presented in this report, both methods give essentially perfect estimates if the models are perfect). This is due to the fact that the Kalman filter-type methods may easily be forced to match the measurements perfectly. However, the present method depends on a nonzero measurement-minus-estimate residual to drive the unmodeled effect; thus, it cannot "correct" itself unless it is in error. A numerical singularity is present in the limit of zero residuals. Consequently, it cannot match every measurement perfectly.

The accuracy of the disturbance estimate of the present method is highly variable. If no model is present and the measurements are dense and accurate, the disturbance recovery may be excellent - e.g., Sections 2.5 and 4.2. However, as the accuracy and frequency of the measurements diminishes, the disturbance recovery is degraded. The spacecraft attitude examples and the examples of Section 4.3 indicate that if the model is reasonably accurate, the unmodeled disturbance recovery is essentially nonexistent. The disturbance term is penalized in the to-be-minimized cost functional, and the results indicate that a sufficiently accurate estimate may be obtained by adding much less disturbance than the actual model error. This is especially true in the "accurately modeled" spacecraft attitude examples, where the presence of large disturbance terms in each angular acceleration equation seems to combine to give an accurate solution even though the magnitudes of the estimated disturbances are smaller than of the actual disturbances. The issue of uniqueness and stationarity are difficult generally and are not addressed in the work. In addition, in this coupled multi-state example, the presence of a model error in one state seems to spawn unmodeled disturbance recovery in the coupled states; thus, the single disturbance may be "spread" over multiple unmodeled disturbance terms, none of which resemble the actual disturbance.

The usefulness of the Kalman filter-type algorithms in estimation during the past twenty five years has been enormous. The Kalman methods have provided an important tool to solve estimation problems outside the scope of classical batch estimation algorithms, specifically those involving significant model errors. They perform extremely well on

systems with good models, and especially systems with dense, accurate measurements. However, the present method has been shown to outperform a typical Kalman filter algorithm on a variety of test problems where large or even modest model errors are present, including some but not all of the densely and accurately measured spacecraft attitude estimation examples. For problems with large model errors and sparse measurements, especially, the present method outperforms any known approach. Certainly more work is necessary to document case studies in other application areas, and to solidify some of the concepts. More theoretical and numerical work is needed especially on the following issues: (i) the consequences of non-stationarity of residuals and measurement errors, (ii) avoiding singular behavior near the limit of zero measurement residuals, and (iii) more rigorous arguments to justify the covariance constraints and minimum model error arguments used to define the algorithms. The results of this work indicate that the present method shows considerable promise as a competitor to the Kalman filter-type estimation strategies. The present method allows, for the first time, batch estimation in the presence of significant unknown model errors.

REFERENCES

1. Alspach, D. L., "Nonlinear Filters In Feedback Control," 6th Symposium on Nonlinear Estimation Theory and Applications, San Diego, CA, Sept. 15-17, 1975.

2. Tse, E., Bar-Shalom, Y., and Meier, L., "Wide-Sense Adaptive Dual Control of Stochastic Nonlinear Systems," IEEE Trans. A.C., Vol. AC-18, pp. 48-108, 1973.

3. Automatica, The Journal of the International Federation of Automatic Control, Pergamon Press, September 1984.

4. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problem," Trans. ASME, J. Basic. Engr., Ser. D., Vol. 82, pp. 34-45, 1960.

5. Kalman, R. E., and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," Trans. ASME, J. Basic Engr., Ser. D, Vol. 83, p. 95, 1961.

6. Junkins, J. L., An Introduction to Optimal Estimation of Dynamical Systems, Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands, 1978.

7. Gelb, A. A. (Editor), Applied Optimal Estimation, The M.I.T. Press, Cambridge, MA, 1974.

8. Anderson, B. D. O., Optimal Filtering, Prentice-Hall, Englewood Cliffs, NJ, 1979.

9. Rhodes, I. B., "A Tutorial Introduction to Estimation and Filtering," IEEE Trans. A.C., Vol. AC-16, pp. 688-706, 1971.

10. Chang, C. B., and Tabaczynski, J. A., "Application of State Estimation to Target Tracking," IEEE Trans. A.C., Vol. AC-29, pp. 98-109, 1984.

11. Nishimura, T., "Modeling Errors in Kalman Filters," Chapter 4 in Theory and Applications of Kalman Filtering, C. T. Leondes, editor, AGARDO-graph 139, 1970.

12. Tukey, J. W., "A Survey of Samplings from Contaminated Distributions," pp. 448-485 in Contributions to Probability and Statistics, I. Olkin, editor, Stanford University Press, Stanford, CA, 1960.

13. Mehra, R. K., "On the Identification of Variances and Adaptive Kalman Filtering," IEEE Trans. A.C., Vol. AC-15, pp. 175-184, 1970.

14. Fitzgerald, R. J., "Divergence of the Kalman Filter," IEEE Trans. A.C., Vol. AC-16, p. 736, 1971.

15. Huber, P. J., "The 1972 Wald Lecture: Robust Statistics: A Review," Annals of Mathematical Statistics, Vol. 43, p. 1041, 1972.

16. Breza, M. J., and Bryson, A. E., "Minimum-Variance Steady-State Filters with Eigenvalue Constraints," 5th Symposium on Nonlinear Estimation Theory and its Applications, San Diego, CA, 1974.

17. Jazwinski, A. H., "Adaptive Filtering," Automatica, Vol. 5, pp. 475-485, 1969.

18. Mehra, R. K., "Approaches to Adaptive Filtering," IEEE Trans. A.C., Vol. 17, pp. 693-698, 1972.

19. Chin, L., "Advances in Adaptive Filtering," in Control and Dynamics Systems, C. T. Leondes editor, Vol. 15, pp. 278-356, Academic Press, NY, 1979.

20. Tsai, C. and Kurz, L., "An Adaptive Robustizing Approach to Kalman Filtering," Automatica, Vol. 19, pp. 279-288, 1983.

21. Dee, Cohn, Dalcher, and Ghil, "An Efficient Algorithm for Estimating Noise Covariances in Distributed Systems," IEEE Trans. A.C., Vol. AC-30, p. 1057, 1985.

22. Bar-Shalom, Y., Tse, E., and Dressler, R., "Adaptive Estimation in the Presence of Non-Stationary Noises with Unknown Statistics-Applications to Maneuvering Targets," Proceedings of the Fourth Symposium on Nonlinear Estimation Theory and its Applications, San Diego, CA, 1973.

23. Alspach, D. L., and Sorensen, H. W., "Recursive Bayesian Estimation Using Gaussian Sums," Automatica, Vol. 7, p. 465, 1971.

24. Masreliez, C. J., and Martin, R. D., "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter," IEEE Trans. A.C., Vol. AC-22, p. 361, 1977.

25. Carew, B., and Belanger, P. R., "Identification of Optimal Filter Steady-State Gain for Systems with Unknown Noise Covariance," IEEE Trans. A.C., Vol. AC-18, pp. 582-589, 1973.

26. Bryson, A. E., and Ho, Y. C., Applied Optimal Control, Blaisdell, Waltham, Mass., 1969.

27. Kirk, D. E., Optimal Control Theory, Prentice-Hall, Englewood Cliffs, NJ, 1970.

28. Pun, L., Introduction to Optimization Practice, John Wiley & Sons, New York, 1969.

29. Goldstine, H. H., A History of the Calculus of Variations from the 17th through the 19th Century, Springer-Verlag, New York, 1980.

30. Bolza, O., Lectures on the Calculus of Variations, University of Chicago Press, Chicago, 1904.

31. Bliss, G. A., Lectures on the Calculus of Variations, University of Chicago Press, Chicago, 1946.

32. Pontryagin, L. S., "Some Mathematical Problems Arising in Connection with the Theory of Optimal Automatic Control Systems," Session of the Academy of Sciences of the USSR on Scientific Problems of the Automating Industry, October 1956, English Translation, Proceedings of the Academy of Sciences, USSR, Vol. 11, pp. 107-117, 1957.

33. Boltyanskii, V. G., Bamkrelidze, R. V., and Pontryagin, L. S., "Theory of Optimal Processes," Doklady Akad. Nauk S.S.S.R., Vol. 110, pp. 7-10, 1956.

34. Gamkrelidze, R. V., "On the Theory of Optimal Processes in Linear Systems," Doklady Akad. Nauk S.S.S.R., Vol. 116, pp. 9-11, 1957.

35. Boltyanskii, V. G., "The Maximum Principle in the Theory of Optimal Processes," Doklady Akad. Nauk S.S.S.R., Vol. 119, pp. 1070-1073, 1958.

36. Gamkrelidze, R. V., "The Theory of Processes which are Optimum for Rapid Action in Linear Systems," Izvest. Akad. Nauk S.S.S.R. Ser. Mat., Vol. 22, No. 4, 1958.

37,38,39
Rozonoer, L. I., "L. S. Pontryagin Maximum Principle in Optimal System Theory," Avtomat. i Telemekh., Vol. 20, 1959, Also in Optimal and Self-Optimizing Control, R. Oldenburger, editor, M.I.T. Press, Part I, pp. 210-224; Part II, pp. 225-241; Part III, pp. 242-257; 1966.

40. Kopp, R. E., "Pontryagin Maximum Principle," Chapter 7 in Optimization Techniques, G. Leitmann, editor, Academic Press, New York, 1962.

41. Geeving, H. P., "Continuous-Time Optimal Control Theory for Cost Functionals Including Discrete State Penalty Terms," IEEE Trans. A.C., Vol. AC-21, pp. 866-869, 1976.

42. Vadali, S. R., "Solution of the Two-Point Boundary Value Problems of Optimal Spacecraft Rotational Maneuvers," Ph.D. Dissertation, VPI&SU, December 1982.

43. Keller, H. B., Numerical Solution of Two Point Boundary Value Problems, Regional Conference Series In Applied Mathematics, No. 24, SIAM, 1976.

44. Jazwinski, A. H., Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.

45. Sage, A. P., and Melsa, J. L., Estimation Theory With Applications to Communications and Control, McGraw-Hill, New York, 1970.

46. Rauch, H. E., Tung, F., and Striebel, C. T., "Maximum Likelihood Estimates of Linear Dynamic Systems," AIAA Journal, Vol. 3, No. 8, pp. 1445-1450, 1965.

47. Junkins, J. L., and Mook, D. J., "Enhanced Spacecraft Attitude Estimation," Final Report, Contract #N60921-83-G-9-A165, VPI&SU, 1985.

48. Ketter, R. L., and Prawel, S. P., Modern Methods of Engineering Computation, McGraw-Hill, New York, 1969.

APPENDIX A - Truth Model

A.1 Overview

In order to test the algorithm developed in this work on an orbit attitude problem, and in order to compare it with the square root EKFS, a "truth model" program was created to generate specific test cases. The model utilizes established solutions for representative orbit attitude histories so that the created orbit (assumed to be "exact") is reasonably consistent with the known underlying physics. This "true orbit" can then be corrupted with known errors in order to simulate measurements for input to the estimation algorithms. The known errors may be as simple as Gaussian white noise, or they may include "unmodeled disturbances" which are combinations of deterministic functions.

The reference frame notation is established in Fig. (A.1). Note that there are three coordinate systems: an inertial system $\{\hat{\underline{n}}\}$, an orbit system $\{\hat{\underline{O}}\}$, and a body system $\{\hat{\underline{b}}\}$.

It is a physical truth that most orbits lie in a plane which is translating along the direction of the unit vector normal to the plane. This normal vector is called the "orbit normal" and is denoted $\hat{O}_2$. $\hat{O}_1$ is defined as pointing towards the center of the orbit in the translating plane, and $\hat{O}_3$ is defined to be perpendicular to $\hat{O}_1$ and $\hat{O}_2$ such that $\hat{O}_1 \times \hat{O}_2 = \hat{O}_3$. Collectively, $\hat{O}_1$, $\hat{O}_2$, and $\hat{O}_3$ are called the "orbit frame" and are denoted by $\{\hat{\underline{O}}\}$.

Since $\hat{O}_2$ is constant, the inertial frame $\{\hat{\underline{n}}\}$ is defined such that $\hat{O}_2 = \hat{n}_2$. Thus, the inertial frame-to-orbit frame transformation matrix is a simple single-axis rotation and may be represented as

$$\{\hat{\underline{O}}\} = [ON]\{\hat{\underline{n}}\}$$

where

$$[ON] = \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix}$$

The body frame is imbedded in the satellite and generally consists of a three-angle rotation away from the orbit reference frame. Typically, a 3-2-1 Euler angle sequence is used. The "1 rotation" is taken about the $\hat{O}_1$ axis and is called "yaw"; denoting the new frame $\{\hat{\underline{O}}\}'$, the "2 rotation" is taken about the $\hat{\underline{O}}_2'$ axis and is called "pitch"; denoting this new frame $\{\hat{\underline{O}}\}''$, the "3 rotation" is taken about the $\hat{O}_3''$ axis and is called "roll". Note that $\hat{b}_3$ and $\hat{O}_3''$ are equivalent. The angular velocities ω_1, ω_2 and ω_3 are measured about the $\hat{b}_1$, $\hat{b}_2$, and $\hat{b}_3$ axes, respectively, presumably by on-board rate gyros.

A.2 True Orbit Attitude and Attitude State Transition Matrix

The "true" orbit attitude history is obtained by numerically integrating the exact torque-free equations of motion for yaw (4), pitch (θ), and roll (ϕ), and ω_1, ω_2, and ω_3. These equations are (e.g., Junkins and Mook[47] (1985)):

$$\dot{x} \begin{Bmatrix} \dot{\psi} \\ \dot{\theta} \\ \dot{\phi} \\ \dot{\omega}_1 \\ \dot{\omega}_2 \\ \dot{\omega}_3 \end{Bmatrix} = \begin{Bmatrix} (\cos\phi f_1 - \sin\phi f_2)/\cos\theta \\ \sin\phi f_1 + \cos\phi f_2 \\ (-\cos\phi\tan\theta f_1 + \sin\phi\tan\theta f_2)/\cos\theta + f_3 \\ (I_2 - I_3)/I_1 \cdot \omega_2\omega_3 \\ (I_3 - I_1)/I_2 \cdot \omega_3\omega_1 \\ (I_1 - I_2)/I_3 \cdot \omega_1\omega_2 \end{Bmatrix} = \underline{F}(\underline{x},\ldots) \qquad (A.2.1)$$

where

$$f_1 = \omega_1 - \omega_o(\sin\psi\sin\theta\cos\phi + \cos\psi\sin\phi)$$

$$f_2 = \omega_2 - \omega_o(-\sin\psi\sin\theta\sin\phi + \cos\psi\cos\phi)$$

$$f_3 = \omega_3 + \omega_o(\sin\psi\cos\theta)$$

ψ = yaw

θ = pitch $\qquad\qquad \omega_o = \frac{2\pi}{\text{period}}$ = orbit angular velocity

ϕ = roll

The numerical integration is carried out using the "Kutta-Simpson 1/3 Rule" (e.g., Ketter and Prawel[48], 1969), outlined as follows:

$$\underline{x}(t) \equiv \text{statevector} \equiv \{\psi(t)\ \theta(t)\ \phi(t)\ \omega_1(t)\ \omega_2(t)\ \omega_3(t)\}^T \quad (A.2.2)$$

Thus, Eqs. (A.2.1) may be written

$$\dot{\underline{x}} = \underline{F}(\underline{x}(t),t) \quad (A.2.3)$$

The numerical integration for time step Δt is calculated as

$$\underline{x}(t + \Delta t) = \underline{x}(t) + \frac{1}{6}(F_1 + 2F_2 + 2F_3 + F_4) \quad (A.2.4)$$

where

$$\begin{aligned} F_1 &= \Delta t \cdot \underline{F}[\underline{x}(t),t] \\ \underline{F}_2 &= \Delta t \cdot \underline{F}[(\underline{x}(t) + \frac{F_1}{2}) \ , \ (t + \frac{\Delta t}{2})] \\ \underline{F}_3 &= \Delta t \cdot \underline{F}[(\underline{x}(t) + \frac{F_2}{2}) \ , \ (t + \frac{\Delta t}{2})] \\ \underline{F}_4 &= \Delta t \cdot \underline{F}[(\underline{x}(t) + \underline{F}_3) \ , \ (t + \Delta t)] \end{aligned} \quad (A.2.5)$$

The attitude state transition matrix is defined as

$$\Phi(t,t_o) \equiv \frac{\partial \underline{x}(t)}{\partial \underline{x}(t_o)} \quad (A.2.6)$$

and is governed by the differential equation

$$\dot{\Phi}(t,t_o) = A(t)\Phi(t,t_o) \quad (A.2.7)$$

where

$$A(t) \equiv \left.\frac{\partial \underline{F}}{\partial \underline{x}}\right|_{\underline{x}(t),t} \tag{A.2.8}$$

and $\underline{F}$ is defined by Eqs. (A.2.1) and (A.2.3). The initial condition for $\Phi(t,t_o)$ is clearly

$$\Phi(t_o,t_o) = I \tag{A.2.9}$$

The state transition matrix may be numerically integrated, simultaneously with the state vector, using the Kutta-Simpson 1/3 Rule. The elements of the partial derivative matrix, A(t) in Eq. (A.2.8), are given below, where by definition

$$A(t) \equiv \frac{\partial \underline{F}}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial F_1}{\partial \psi} & \frac{\partial F_1}{\partial \theta} & \cdots & \frac{\partial F_1}{\partial \omega_3} \\ \vdots & & & \vdots \\ \frac{\partial F_6}{\partial \psi} & \cdots\cdots\cdots & & \frac{\partial F_6}{\partial \omega_3} \end{bmatrix} \tag{A.2.10}$$

$$\frac{\partial F_1}{\partial \psi} = \frac{1}{\cos\theta}\left(\cos\phi \frac{\partial f_1}{\partial \psi} - \sin\phi \frac{\partial f_2}{\partial \psi}\right)$$

$$\frac{\partial F_1}{\partial \theta} = \frac{\sin\theta}{\cos^2\theta}\left(\cos\phi f_1 - \sin\phi f_2\right) + \frac{1}{\cos\theta}\left(\cos\phi \frac{\partial f_1}{\partial \theta} - \sin\phi \frac{\partial f_2}{\partial \theta}\right)$$

$$\frac{\partial F_1}{\partial \phi} = \frac{1}{\cos\theta}\left(-\sin\phi f_1 - \cos\phi f_2 + \cos\phi \frac{\partial f_1}{\partial \phi} - \sin\phi \frac{\partial f_2}{\partial \phi}\right)$$

$$\frac{\partial F_1}{\partial \omega_i} = \frac{1}{\cos\theta}\left(\cos\phi \frac{\partial f_1}{\partial \omega_i} - \sin\phi \frac{\partial f_2}{\partial \omega_i}\right)$$

$$\frac{\partial F_2}{\partial \psi} = \sin\phi \frac{\partial f_1}{\partial \psi} + \cos\phi \frac{\partial f_2}{\partial \psi}$$

$$\frac{\partial F_2}{\partial \theta} = \sin \phi \frac{\partial f_1}{\partial \theta} + \cos \phi \frac{\partial f_2}{\partial \theta}$$

$$\frac{\partial F_2}{\partial \phi} = \cos \phi f_1 - \sin \phi f_2 + \sin \phi \frac{\partial f_1}{\partial \phi} + \cos \phi \frac{\partial f_2}{\partial \phi}$$

$$\frac{\partial F_2}{\partial \omega_i} = \sin \phi \frac{\partial f_1}{\partial \omega_i} + \cos \phi \frac{\partial f_2}{\partial \omega_i}$$

$$\frac{\partial F_3}{\partial \psi} = \frac{1}{\cos\theta} (- \cos \phi \tan\theta \frac{\partial f_1}{\partial \psi} + \sin \phi \tan\theta \frac{\partial f_2}{\partial \psi}) + \frac{\partial f_3}{\partial \psi}$$

$$\frac{\partial F_3}{\partial \theta} = \frac{1}{\cos^2\theta} \{(-\cos \phi \sec^2\theta f_1 - \cos \phi \tan \theta \frac{\partial f_1}{\partial \theta} + \sin \phi \sec^2\theta f_2$$

$$+ \sin \phi \tan \theta \frac{\partial f_2}{\partial \theta}) \cos \theta + \sin \theta(- \cos \phi \tan \theta f_1$$

$$+ \sin \phi \tan \theta f_2)\} + \frac{\partial f_3}{\partial \theta}$$

$$\frac{\partial F_3}{\partial \phi} = \frac{1}{\cos\theta} (\sin \phi \tan \theta f_1 + \cos \phi \tan \theta f_2 - \cos \phi \tan\theta \frac{\partial f_1}{\partial \phi}$$

$$+ \sin \phi \tan \theta \frac{\partial f_2}{\partial \phi}) + \frac{\partial f_3}{\partial \phi}$$

$$\frac{\partial F_3}{\partial \omega_i} = \frac{\tan\theta}{\cos\theta} (- \cos \phi \frac{\partial f_1}{\partial \omega_i} + \sin \phi \frac{\partial f_2}{\partial \omega_i}) + \frac{\partial f_3}{\partial \omega_i}$$

where

$$\frac{\partial f_1}{\partial \psi} = -\omega_o(\cos \psi \sin \theta \cos \phi - \sin \psi \sin \phi)$$

$$\frac{\partial f_2}{\partial \psi} = - \omega_o(-\cos \psi \sin \vartheta \sin \phi - \sin \psi \cos \phi)$$

$$\frac{\partial f_3}{\partial \psi} = \omega_o \cos \psi \cos \theta$$

$$\frac{\partial f_1}{\partial \theta} = -\omega_o(\sin \psi \cos \theta \cos \phi)$$

$$\frac{\partial f_2}{\partial \theta} = -\omega_o(-\sin \psi \cos \theta \sin \phi)$$

$$\frac{\partial f_3}{\partial \theta} = -\omega_o \sin \psi \sin \theta$$

$$\frac{\partial f_1}{\partial \phi} = -\omega_o(-\sin \psi \sin \theta \sin \phi + \cos \psi \cos \phi)$$

$$\frac{\partial f_2}{\partial \phi} = -\omega_o(-\sin \psi \sin\theta \cos\phi - \cos\psi \sin \phi)$$

$$\frac{\partial f_3}{\partial \phi} = 0$$

$$\frac{\partial f_1}{\partial \omega_1} = 1 \qquad \frac{\partial f_2}{\partial \omega_1} = 0 \qquad \frac{\partial f_3}{\partial \omega_1} = 0$$

$$\frac{\partial f_1}{\partial \omega_2} = 0 \qquad \frac{\partial f_2}{\partial \omega_2} = 1 \qquad \frac{\partial f_3}{\partial \omega_2} = 0$$

$$\frac{\partial f_1}{\partial \omega_3} = 0 \qquad \frac{\partial f_2}{\partial \omega_3} = 0 \qquad \frac{\partial f_3}{\partial \omega_3} = 1$$

$$\frac{\partial F_4}{\partial \psi} = \frac{\partial F_4}{\partial \theta} = \frac{\partial F_4}{\partial \phi} = \frac{\partial F_4}{\partial \omega_1} = 0$$

$$\frac{\partial F_4}{\partial \omega_2} = \frac{I_2 - I_3}{I_1} \omega_3 \quad , \quad \frac{\partial F_4}{\partial \omega_3} = \frac{I_2 - I_3}{I_1} \omega_2$$

$$\frac{\partial F_5}{\partial \psi} = \frac{\partial F_5}{\partial \theta} = \frac{\partial F_5}{\partial \phi} = \frac{\partial F_5}{\partial \omega_2} = 0$$

$$\frac{\partial F_5}{\partial \omega_1} = \frac{I_3 - I_1}{I_2}\,\omega_3 \quad , \quad \frac{\partial F_5}{\partial \omega_3} = \frac{I_3 - I_1}{I_2}\,\omega_1$$

$$\frac{\partial F_6}{\partial \psi} = \frac{\partial F_6}{\partial \theta} = \frac{\partial F_6}{\partial \phi} = \frac{\partial F_6}{\partial \omega_3} = 0$$

$$\frac{\partial F_6}{\partial \omega_1} = \frac{I_1 - I_2}{I_3}\,\omega_2 \quad , \quad \frac{\partial F_6}{\partial \omega_2} = \frac{I_1 - I_2}{I_3}\,\omega_1$$

A.3 Use in the Attitude Estimation Study

The following items are calculated for use in the attitude estimation study:

1) Simulated attitude measurements at a rate of one measurement set ($\tilde{\psi}$, $\tilde{\theta}$, $\tilde{\phi}$) every 10 seconds, and
2) Simulated angular velocity measurements at a rate of one measurement set (ω_1, ω_2, ω_3) every 1/20 second.

The frequency of the two measurement sets is approximately equal to the state-of-the-art measurement hardware.

The required input to the truth model consists of the following:

1) Attitude initial conditions (measurements)

   $\psi(t_o)$, $\theta(t_o)$, $\phi(t_o)$, $\omega_1(t_o)$, $\omega_2(t_o)$, $\omega_3(t_o)$
2) Orbit angular velocity, ω_o
3) Means and variances for the measurement errors in the simulated attitude and angular velocity measurement sets.

The true orbit attitude history is obtained by integrating the orbit attitude governing equations, Eqs. (A.3.1). The measurement sets are obtained by adding Gaussian random sequences to the true values at the measurement times. The mean and variance for attitude and the mean

and variance for angular velocity are used to determine the appropriate Gaussian random number sequences. The simulated measurement is then created as

$$\begin{matrix}\text{Measured} \\ \text{Value}\end{matrix} = \begin{matrix}\text{True} \\ \text{Value}\end{matrix} + \begin{matrix}\text{Random} \\ \text{Number}\end{matrix}$$

and written out for input to the attitude estimation programs.

In addition, "unmodeled disturbances" consisting of deterministic functions may be added to the angular acceleration equations as

$$\begin{aligned}
\dot{\omega}_1 &= \frac{I_2-I_3}{I_1}\,\omega_2\omega_3 + d_1(t) \\
\dot{\omega}_2 &= \frac{(I_3-I_1)}{I_2}\,\omega_3\omega_1 + d_2(t) \\
\dot{\omega}_3 &= \frac{(I_1-I_2)}{I_3}\,\omega_1\omega_2 + d_3(t)
\end{aligned} \qquad \text{(A.3.1)}$$

where $d_1(t)$, $d_2(t)$, and $d_3(t)$ are used to represent unmodeled dynamic effects. The $d_i(t)$ functions are known. However, the attitude estimation algorithms are "unaware" of these unmodeled effects, i.e., the model employed by the estimation algorithms does not include these known functions. In this way, the estimation algorithms can be compared for their ability to deal with deterministic unmodeled effects as well as white noise in the measurements.

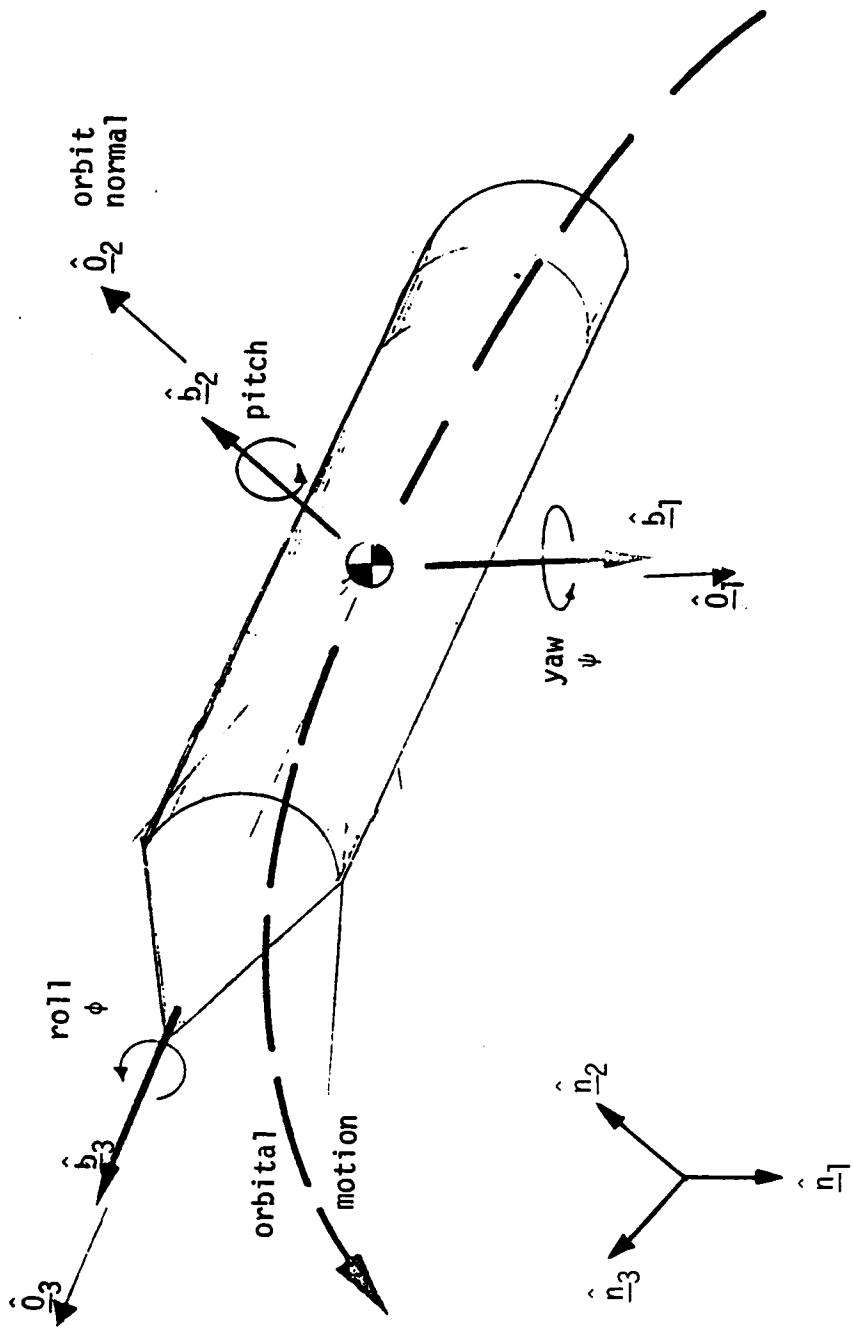

orbit normal
$\hat{o}_2$
$\hat{b}_2$
pitch
$\hat{b}_1$
$\hat{o}_1$
yaw
ψ
roll
ϕ
$\hat{b}_3$
$\hat{o}_3$
orbital motion
$\hat{n}_2$
$\hat{n}_1$
$\hat{n}_3$

Figure A.1

The vita has been removed from the scanned document

MEASUREMENT COVARIANCE-CONSTRAINED ESTIMATION IN THE PRESENCE OF POORLY MODELED DYNAMIC SYSTEMS

by

Daniel Joseph Mook

(ABSTRACT)

An optimal estimation strategy is developed for post-experiment estimation of discretely measured dynamic systems which accounts for system model errors in a much more rigorous manner than Kalman filter-smoother type methods. The Kalman filter-smoother type methods, which currently dominate post-experiment estimation practice, treat model errors via "process noise", which essentially shifts emphasis away from the model and onto the measurements. The usefulness of this approach is subject to the measurement frequency and accuracy.

The current method treats model errors by use of an estimation strategy based on concepts from optimal control theory. Unknown model error terms are explicitly included in the formulation of the problem and estimated as a part of the solution. In this manner, the estimate is improved; the model is improved; and an estimate of the model error is obtained. Implementation of the current method is straightforward, and the resulting state trajectories do not contain jump discontinuities as do the Kalman filter-smoother type estimates.

Results from a number of simple examples, plus some examples from spacecraft attitude estimation, are included. The current method is shown to obtain significantly more accurate estimates than the Kalman filter-smoother type methods in many of the examples. The difference in accuracy is accentuated when the assumed model is relatively poor and when the measurements are relatively sparse in time and/or of low

accuracy. Even for some well-modeled, densely measured applications, the current method is shown to be competitive with the Kalman filter-smoother type methods.