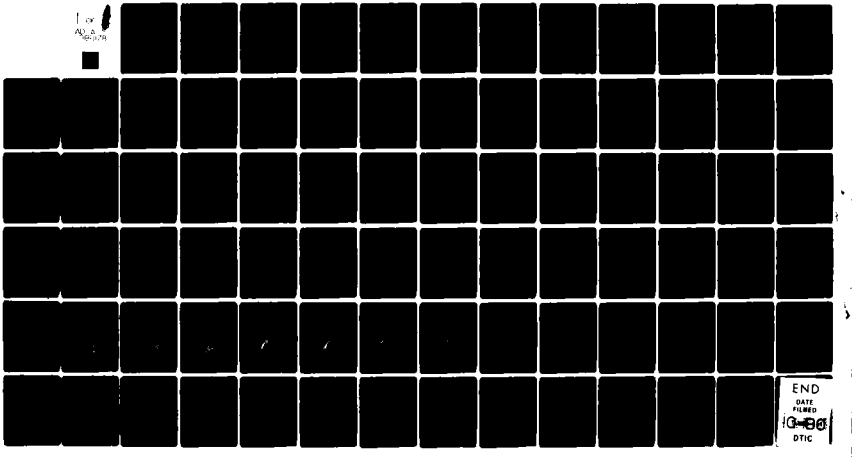
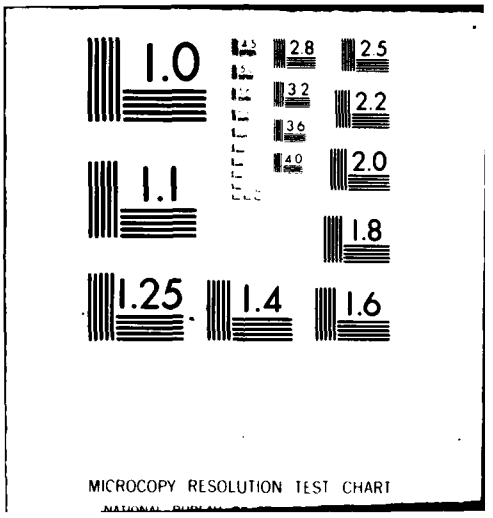
AD-A088 178 PRINCETON UNIV NJ INFORMATION SCIENCES AND SYSTEMS LAB F/G 12/1
THE OFF-LINE USE OF A SEQUENTIAL ESTIMATOR.(U)
JUL 80 T G ROBERTAZZI, S C SCHARTZ N00014-80-C-0530
UNCLASSIFIED TR-4 NL



END
DATE FILMED
10-80
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART

Report Number 4

LEVEL II

12
BS.

AD A088178

THE OFF-LINE USE OF A SEQUENTIAL ESTIMATOR

T.G. Robertazzi and S.C. Schwartz

INFORMATION SCIENCES AND SYSTEMS LABORATORY

Department of Electrical Engineering and Computer Science
Princeton University
Princeton, New Jersey 08540

JULY 1980

Technical Report

Approved for public release; distribution unlimited

Prepared for

OFFICE OF NAVAL RESEARCH (Code 436)
Statistics and Probability Branch
Arlington, Virginia 22217

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (14) TR-4	2. GOVT ACCESSION NO. AD-A088 178
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) (6) The Off-Line Use of a Sequential Estimator	5. TYPE OF REPORT & PERIOD COVERED (9) Technical Report
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s) (10) Thomas G. Robertazzi and Stuart C. Schwartz	8. CONTRACT OR GRANT NUMBER(s) (15) N00014-80-C-0530
9. PERFORMING ORGANIZATION NAME AND ADDRESS Information Sciences & Systems Laboratory Dept. of Elec. Eng. & Computer Science Princeton University, Princeton, NJ 08544	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR042-385
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE (11) July 1980
	13. NUMBER OF PAGES 87
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 82	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Sequential estimation
Extended Kalman Filter
Filter Divergence
Gauss-Newton Method

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The straightforward application of a sequential estimator to nonlinear regression (curve fitting) problems is generally not possible when good a priori parameter estimates are not available and also when minimizing the error over a local portion of the data does not insure that it is minimized globally. However, a sequential estimator may be easily utilized to perform off-line processing in such a situation. The key is to process the measure-

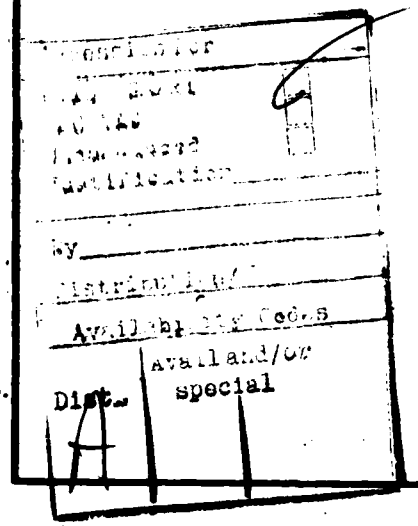
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ments in a random order rather than the causal order in which they occur.

The off-line use of an extended Kalman filter is illustrated in terms of a particular application. This technique is essentially a sequential version of the Gauss-Newton minimization procedure with relinearization being performed after each measurement is processed. Ficticious measurement noise is necessary to prevent filter divergence and is included in a very simple manner.

Computational savings over more conventional iterative minimization techniques are possible if the functions and partial derivatives involved are sufficiently complex to evaluate. But there is a real question regarding the extent to which convergence can be assured. The results of simulations are presented.



THE OFF-LINE USE OF A SEQUENTIAL ESTIMATOR

by

T.G. Robertazzi and S.C. Schwartz
Department of Electrical Engineering and Computer Science
Princeton University
Princeton, NJ 08544

Abstract

The straightforward application of a sequential estimator to nonlinear regression (curve fitting) problems is generally not possible when good a priori parameter estimates are not available and also when minimizing the error over a local portion of the data does not insure that it is minimized globally. However, a sequential estimator may be easily utilized to perform off-line processing in such a situation. The key is to process the measurements in a random order rather than the causal order in which they occur.

The off-line use of an extended Kalman filter is illustrated in terms of a particular application. This technique is essentially a sequential version of the Gauss-Newton minimization procedure with relinearization being performed after each measurement is processed. Ficticious measurement noise is necessary to prevent filter divergence and is included in a very simple manner.

Computational savings over more conventional iterative minimization techniques are possible if the functions and partial derivatives involved are sufficiently complex to evaluate. But there is a real question regarding the extent to which convergence can be assured. The results of simulations are presented.

TABLE OF CONTENTS

Chapter 1

The Problem Statement

We will restrict our attention to nonlinear models of the form

$$y(\theta_k) = h(\theta_k) + V_k \ , \ k=1,2...N \tag{1a}$$

$$h(\theta_k) = f(\theta_k)(c_1 + c_2\theta_k + \cdots + c_n\theta_k^{n-1}) \tag{1b}$$

where

i) $f(\theta)$ is a known continuous function that is even and monotonically decreasing in θ.

ii) $\lim_{|\theta|\to\infty} \theta^{n-1} f(\theta) = 0$

iii) θ is a function of time and some parameters whose values are unknown. The c's are also unknown.

iv) n is known.

v) The number of data points available and the signal-to-noise ratio are such that a curve fitting technique, that is estimating parameters by minimizing some error criteria, is viable.

For the purposes of this paper, it will also be assumed that

vi) V_k is a Gaussian random variable $\sim n(0,\sigma^2)$, σ^2 known.

vii) $\theta_k = S(t_k - T_0)$, where S and T_0 are unknown scale and location parameters, respectively.

One wishes to estimate the parameters, $c_1, c_2 \ldots c_n, S, T_0$. This is a nonlinear (in the parameters) optimization problem. For off-line processing, there are several well-known iterative mini-

mization techniques (IMT), such as those of the descent type, that are obvious possibilities.

Moreover, the nonlinear model in question is a 'separable' model; its parameters can be separated into two groups, those which appear linearly (c_1, c_2, c_3) and those which appear nonlinearly (S, T_0). This class of nonlinear models is significant both because of the wide variety of applications in which it appears and because it's structure can be exploited in off-line optimization to achieve computational efficiency [1].

We will first, however, investigate the suitability of sequential estimation of the parameters. By 'sequential estimation' is meant calculating an updated estimate as each sampled measurement of $y(\theta)$, that is $y(\theta_k)$, arrives. This calculation is made using only the previous estimate and the current observation.

The reason for considering sequential estimation is that in at least one application, the estimation of ship movement by means of a fixed sensor measuring magnetic field intensity, the signal can arrive over a period of a minute or so with the spacing between sampled points being on the order of tenths of a second. Not only may enough time be available for real time sequential processing but it may also be desirable to obtain reasonably accurate estimates as early as possible.

Chapter 2

Some Difficulties With Sequential Processing

The model of (1) can be put in the Kalman estimation framework with the state vector, $\underline{x}_k$, comprised of the parameters to be estimated. Then:

$$\begin{aligned} \underline{x}_{k+1} &= \underline{x}_k \\ y_k &= h_k(\underline{x}_k) + v_k \end{aligned} \tag{2}$$

where there is no system noise present.

Nonlinear sequential estimators, such as the extended Kalman filter, are most successful when an initial reference "trajectory", about which one can linearize, is known to a high degree of confidence. For this discussion and for the latter examples we will assume that initial nonlinear parameter estimates accurate to about a factor of two are available.

The main difficulty that is encountered in attempting to sequentially estimate the parameters of (1) is that minimizing the error (between the observations and those predicted using the estimated parameters) over a local section of the waveform does not guarantee a good fit over the entire waveform.

In particular, for a local section of the waveform many sets of parameter estimates will produce almost equally good fits.

To illustrate this point, consider the model of (1) with

$$\begin{aligned} f(\theta_k) &= e^{-\theta_k^2} \\ S &= .014 \\ T_0 &= 200. \\ C1 &= 1.0 \\ C2 &= 1.0 \\ C3 &= -.1 \\ \sigma^2 &= 10^{-4} \\ N &= 200 \end{aligned} \tag{3}$$

Fig. 1. is a plot of $y(\theta_k)$.

For different values of (S,T_0), the least square estimate of (c_1,c_2,c_3) were calculated using the noise corrupted waveform of Fig. 1. The sum of the square errors between the $y(\theta_k)$ and the model using each such set of $(\hat{S},\hat{T}_0,\hat{c}_1,\hat{c}_2,\hat{c}_3)$ was also calculated. Because of the least square error criterion, the linear parameter estimates, $\hat{c}_1,\hat{c}_2,\hat{c}_3$, are a function of the chosen $(\hat{S},\hat{T}_0)$ Thus a three-dimensional plot of the error versus various values of $(\hat{S},\hat{T}_0)$ is informative.

Moreover, Golub has established that there is a direct relationship between the extrema in this plot and those in the space of all the parameters.

The theorem, which is proven in Golub [1], deals with square error optimization applied to separable models. One can define two error functions. The first $e_1^2(\hat{\underline{x}})$, is the error as a function of both linear and nonlinear parameters. Now let $\underline{x}$ be partioned into linear and nonlinear parameters, $\underline{x}^T=[\underline{x}_{LIN}^T,\underline{x}_{NL}^T]^T$. The second error function, $e_2^2(\hat{\underline{x}}_{NL})$, utilizes the fact that for a given nonlinear parameter estimate, $\hat{\underline{x}}_{NL}$, the least square estimate of the linear parameters, $\hat{\underline{x}}_{LIN\ LS}$ is unique. Since the linear parameter estimates are then a function of the nonlinear parameter estimates, $e_2^2(\hat{\underline{x}}_{NL})$ is a function of only the nonlinear parameters.

For our purposes, the theorem is relevant as it establishes that for a small enough neighborhood in the nonlinear parameter space, Ω, where the matrix of basis functions has constant rank (see Appendix B) the following are true:

i) if $\hat{\underline{\alpha}}_{NL}$ is a critical point (or global minimizer in Ω) of $e_2^2(\hat{\underline{\alpha}}_{NL})$, then $(\hat{\underline{\alpha}}_{LIN\ LS}^T, \hat{\underline{\alpha}}_{NL}^T)^T$ is also a critical point (or global minimizer for $\hat{\underline{\alpha}}_{NL} \epsilon \Omega$) of $e_1^2(\hat{\underline{\alpha}})$.

ii) if $(\hat{\underline{\alpha}}_{LIN\ LS}^T, \hat{\underline{\alpha}}_{NL}^T)^T$ is a global minimizer in $e_1^2(\hat{\underline{\alpha}})$ for $\hat{\underline{\alpha}}_{NL} \epsilon \Omega$, then $\hat{\underline{\alpha}}_{NL}$ is a global minimizer of $e_2^2(\hat{\underline{\alpha}}_{NL})$ in Ω.

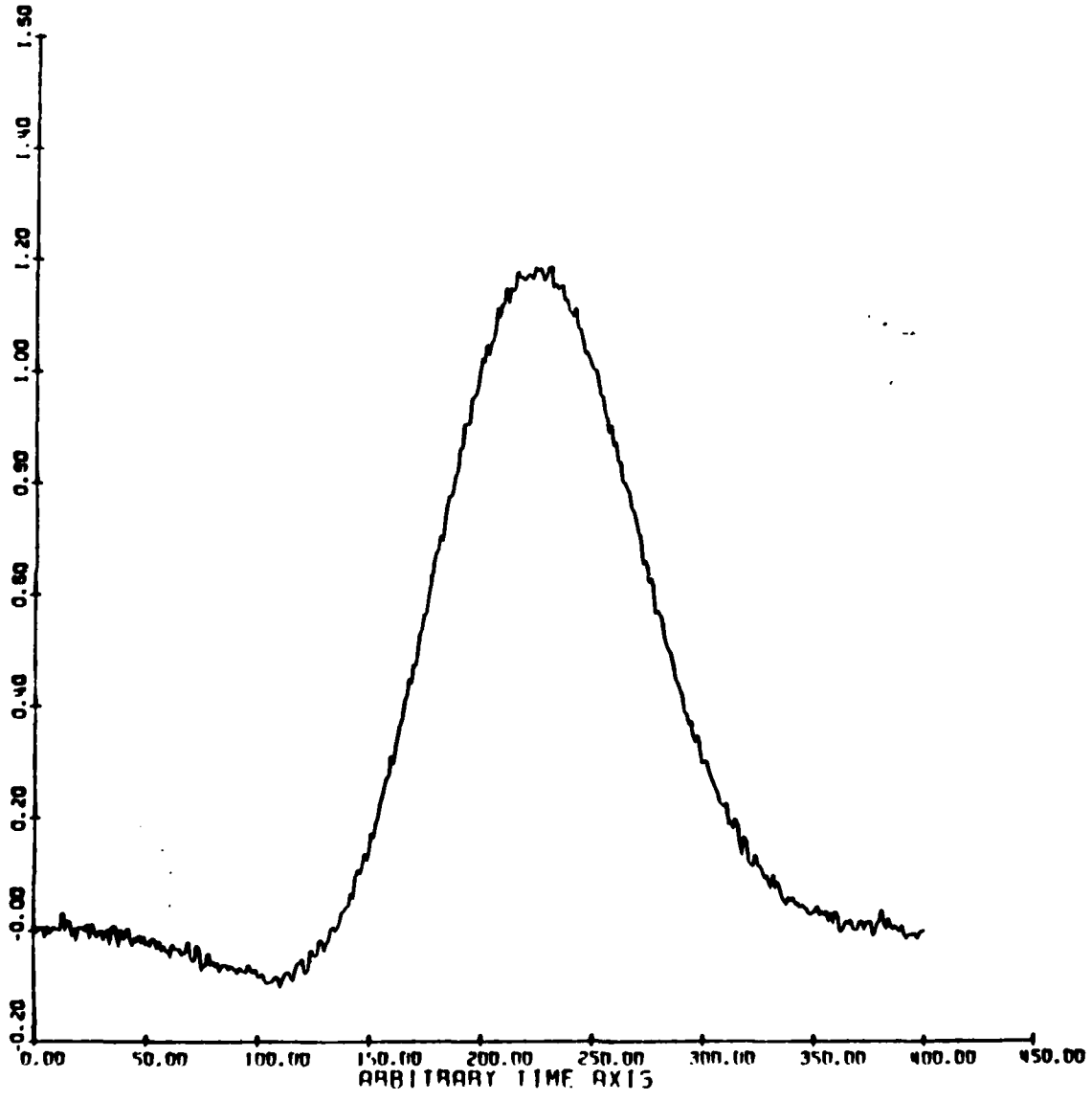
Practically, (i) implies that a local minimum found in one of the previous error surfaces, which are functions only of the nonlinear parameters, would in fact be a minimum in the space of all the parameters. This is the space in which the RSKF, as well as iterative minimization techniques, would operate.

In particular, several plots were made, Figs. 2-5, using the first 100,200,300 and 400 sampled values of the waveform, respectively. The error was normalized by $\sum_{k=1}^{N} y^2(\theta_k)$ and $-\log_{10}$ of this error was actually plotted so that the z axis is a logarithmic scale and a peak corresponds to an error minimum.

As one might expect, with only the first 100 sampled values accesible (Fig. 2), the error surface is quite flat. There is no clearly optimum set of $(S, \underline{T}_0)$. As more data is available (Figs. 3-5), the optimal $(S, \underline{T}_0)$ does indeed develop at about ($S=.014$, $T_0=200$), while local minima decrease in significance.

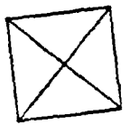
A similar set of plots (Figs. 6-9) were made using $f(\theta_k) = (1+\theta_k^2)^{-5/2}$. This function arises in a particular application that is explained in some detail in Chapter 4. The parameters of (25) were used in generating these plots. With only 100 sampled values available, once again the error surface is rather flat.

This situation especially mitigates against the use of estimators that implicitly assume that the current parameter estimate adequately summarizes the information concerning the parameters



DATA USING THE MODEL OF EQUATION 3

Figure 1



FIRST 100 SAMPLES ARE ACCESSIBLE

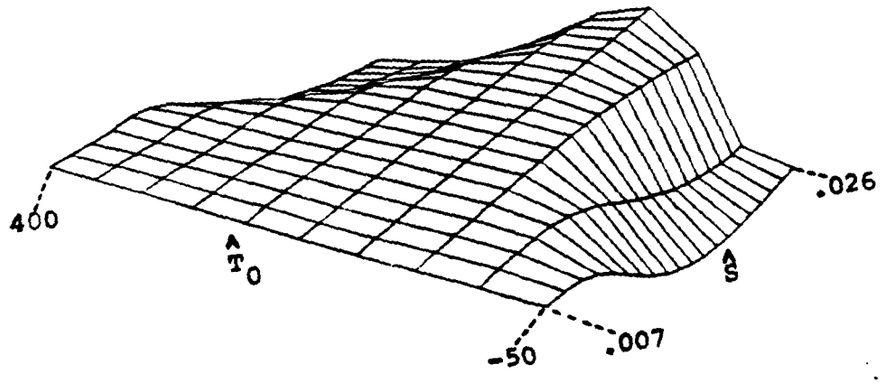


Figure 2

All 3-D plots utilize the same orientation and the same independent variable scaling.

01/19/95
15:32:35

FIRST 200 SAMPLES ARE ACCESSIBLE

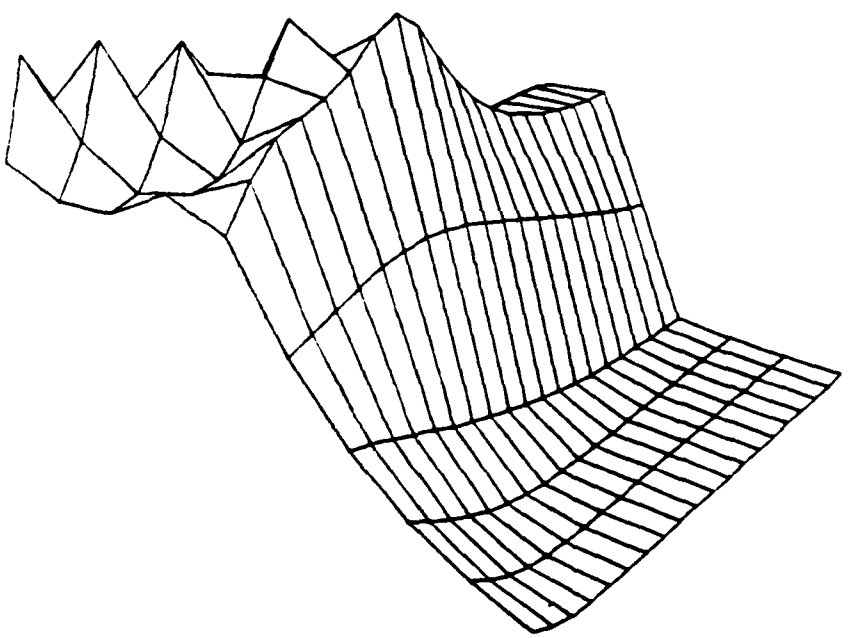


Figure 3

FIRST 300 SAMPLES ARE ACCESSIBLE

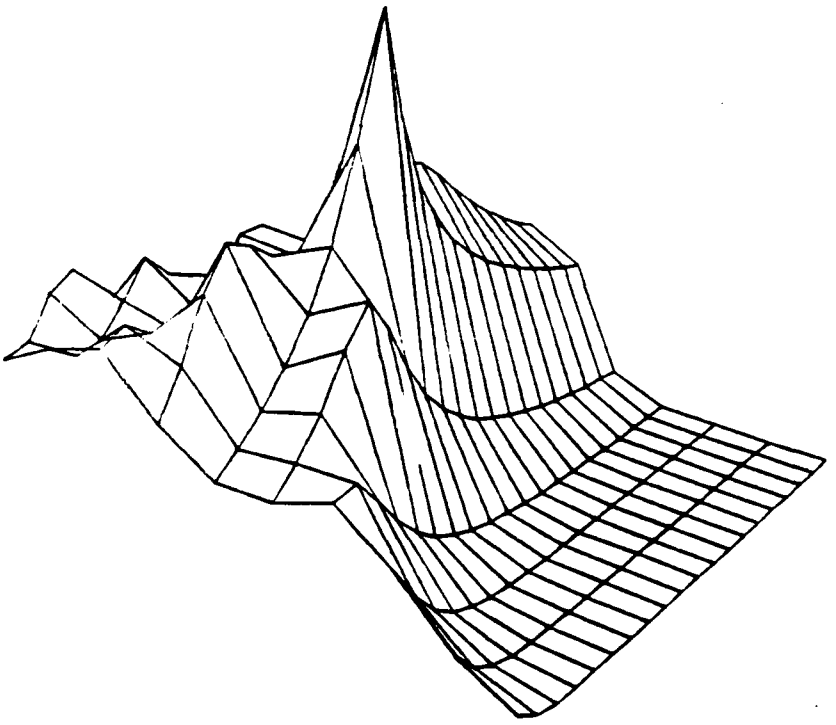


Figure 4

01/10/80
15.34.03

FIRST 400 SAMPLES ARE ACCESSIBLE

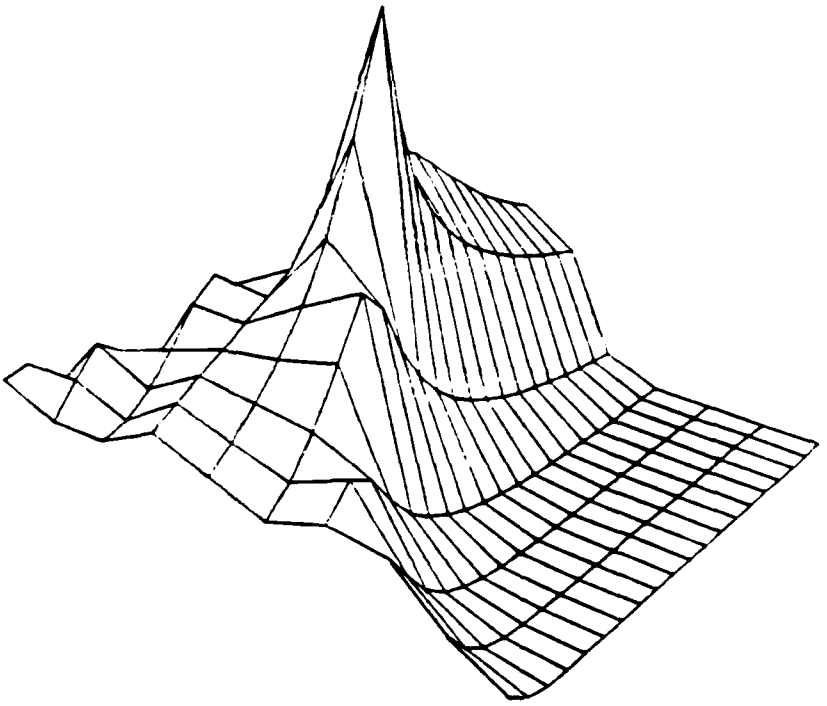
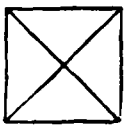


Figure 5

PROFILE
04/01/79



12/09/79
15:56:22

FIRST 100 SAMPLES ARE ACCESSIBLE

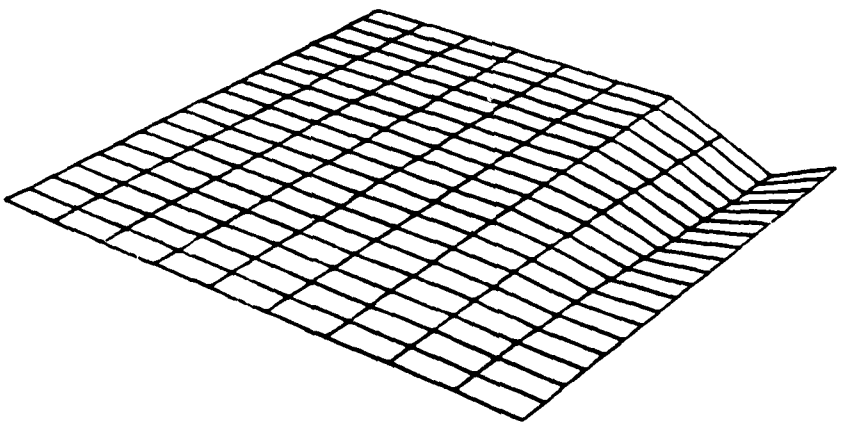


Figure 6

REVUE
04/C/79

FIRST 200 SAMPLES ARE ACCESSIB_E

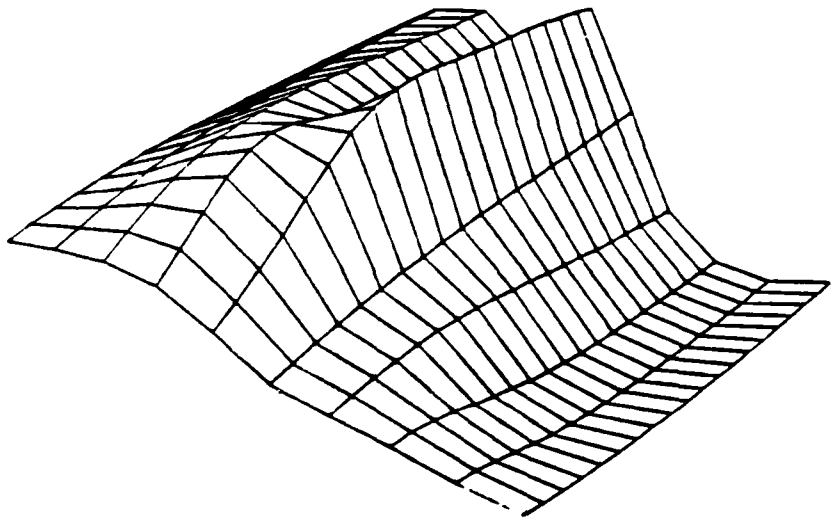


Figure 7

FIRST 300 SAMPLES ARE ACCESSIBLE

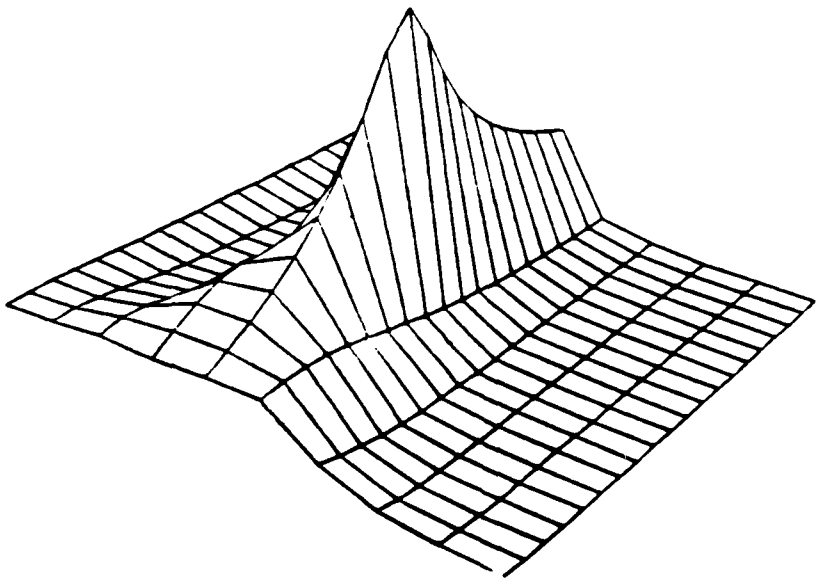


Figure 8

12/08/79
15:51:47

FIRST 400 SAMPLES ARE ACCESSIBLE

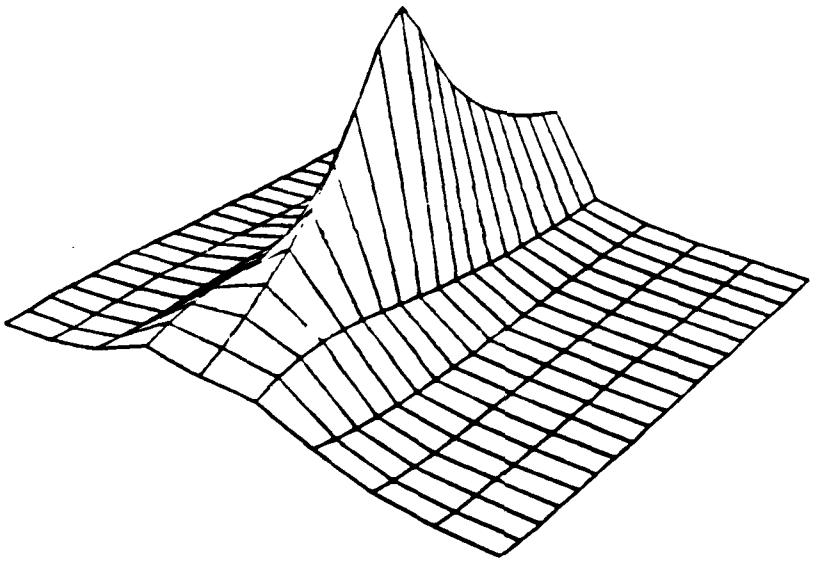


Figure 9

PEAKJK
08/01/79

The problem with a sequential estimator that processes the data in its causal order is that during the early sections of the waveform there are many parameter values that provide locally almost equally good fits. Then, when the filter is processing data in the latter section of the waveform it again has no information available to it concerning the fit in the other (earlier) parts of the waveform.

Recursive least square estimation of only linear parameter models does not suffer from this problem as the estimate at any time, t_k, is the least square estimate for <u>all</u> measurements received up till t_k. Nonlinear estimators generally do not enjoy this property.

An extended Kalman filter, for instance, maintains an error covariance matrix, $\underline{P}_k$, whose trace is non-increasing - (in the absence of system noise). Practically this means that a good fit during the early part of the waveform leads to unrealistically low error-covariance estimates. These represent such a high degree of assumed confidence in the current estimate that when more recent measurements show a lack of fit, through the growth of the residuals (innovations), the gain is set small enough to ignore them and the filter ultimately diverges.

Before continuing it should be noted that the unsuitability of sequential estimation has been discussed only for a particular type of model and only within the limits of our earlier definition of 'sequential estimation'. Tenney et.al.[2], for instance, obtained good results for a somewhat related model through a linearized version of the model equations that depended only on a single nonlinear parameter, thus allowing the use of parallel filters. Also, a real time estimator is possible if it stored all or some representative

portion of the data received up till time t_k and its estimate depended only on this data, not the previous estimate. The difficulty here, though, is computational.

While a simple sequential estimator may be inappropriate for real time processing, the next section will show that it is possible to effectively utilize it in an off-line manner.

Chapter 3

Off-Line Use of a Sequential Estimator

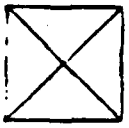
3.1 Introduction

The main impediment to the on-line implementation of a sequential estimator, in the context of the model (1), is that for the most recent estimate of the parameters, the error is only evaluated over a local section of the waveform. While one would like to evaluate the error over the entire waveform, this can only be done once the data has been completely received, that is, off-line. Since off-line processing may be justifiable in some applications, it will now be considered.

Given samples from the entire waveform, how can a sequential estimator, such as the extended Kalman filter, perform off-line processing? The simplest way would be to process the measurements not in their causal order $(y_1, y_2, y_3 \ldots)$ but in some random order $(y_{37}, y_{205}, y_{80} \ldots)$. An obvious possibility is equiprobable sampling without replacement. The advantage of such a random sampling estimator is that over a number of iterations the filter obtains a measure of the error over the entire waveform.

Again, the three dimensional error plots of Chapter 2 can be used for illustrative purposes. Figs. 10-13 were generated with $f(\theta_k) = e^{-\theta_k^2}$ and the parameters of (3). However, the measurements were randomly sampled (without replacement). Thus Fig. 10 is based on the first 10 randomly selected measurements. Fig. 11 uses an additional 30 measurements for a total of 40 randomly selected measurements and so on.

What these plots indicate is that in contrast to Fig. 2, even with 10 random measurements, there is a clearly global optimum.



10 RANDOMLY SELECTED SAMPLES

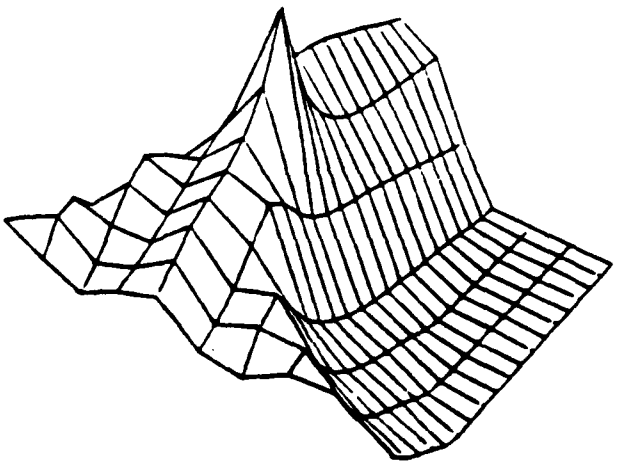


Figure 10

40 RANDOMLY SELECTED SAMPLES

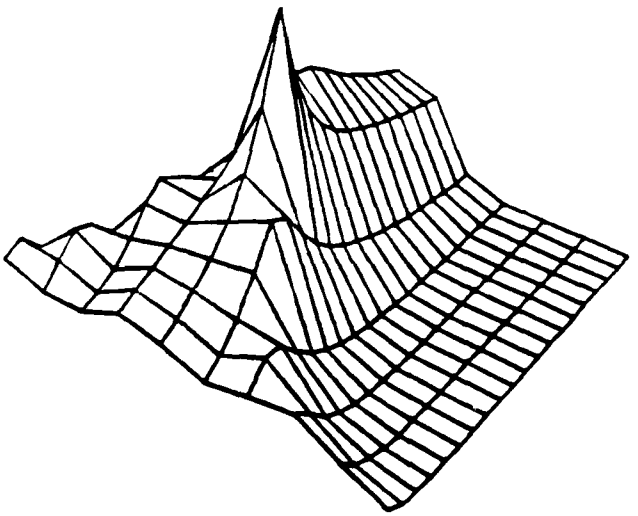


Figure 11

200 RANDOMLY SELECTED SAMPLES

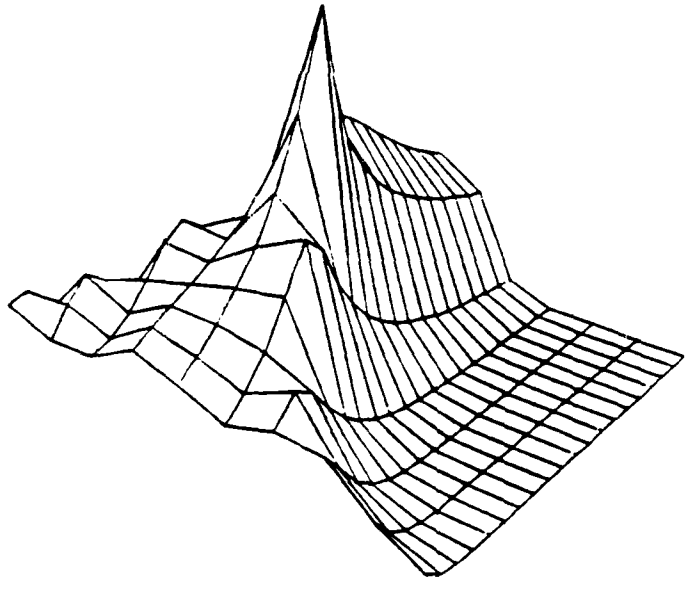


Figure 12

400 RANDOMLY SELECTED SAMPLES

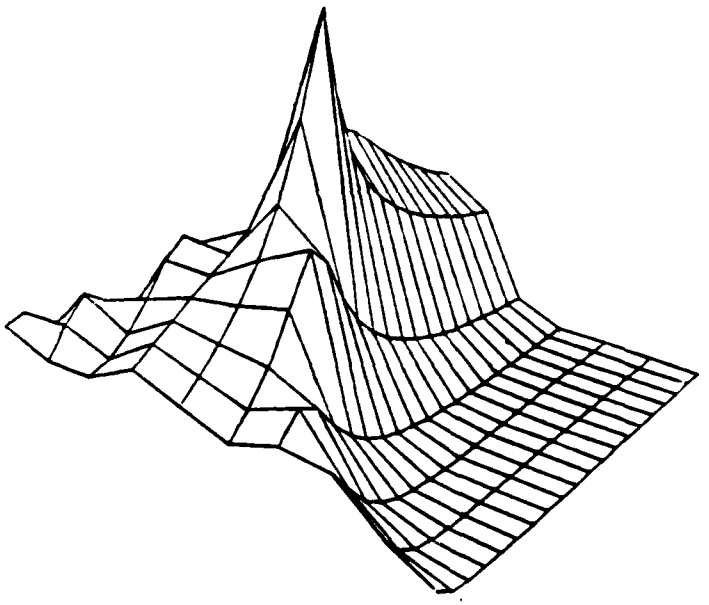


Figure 13

Additional measurements, of course, improve its "visibility".

Henceforth, an extended Kalman filter which processes pre-recorded measurements in a random order will be referred to as a random sampling Kalman filter (RSKF).

3.2 Relation to the Gauss-Newton Technique

There is a close relationship between the extended Kalman filter for (2) and the iterative and off-line (batch) Gauss-Newton optimization technique. Both express the measurement nonlinearity as a first order Taylor series expansion so as to obtain equations that are linear in the state/parameter deviation ($\underline{dx}_k = \underline{x}_{k+1} - \underline{x}_k$). Linear solution methods then can be applied.

It is well known that if an extended Kalman filter does not relinearize (continue using the original $\underline{x}$ estimate for generating basis functions and partial derivatives), after each measurement is processed, it can be made to produce results identical to those of a single Gauss-Newton iteration [8]. That is, without relinearization, an extended Kalman filter is a sequential version of the Gauss-Newton technique. It follows that the RSKF described previously can be thought of as a sequential version of a Gauss-Newton iteration where relinearization is performed after each measurement is processed.

While such relinearization offers the potential for faster convergence [8], there are often divergence problems associated with extended Kalman filters. Furthermore, for certain models the filter may converge to points other than the convergence points of the off-line (batch) technique [8,12]. Steps to mitigate against the former difficulty are discussed in sec. 3.4. The remainder of this section is devoted to a more detailed exposition of the previous points.

In the Gauss-Newton approach one starts with N measurement equations modeled as

$$\underline{Y} = \underline{\bar{g}}\ (\underline{x}) + \underline{\varepsilon} \tag{4}$$

where each term is a Nx1 column vector and the statistical properties of $\underline{\varepsilon}$ are assumed to justify a least squares fit. Expanding the nonlinearity one has:

$$\underline{Y} = \underline{\bar{g}}\ (\underline{\hat{x}}_k) + \left.\frac{\partial \underline{\bar{g}}(\underline{\hat{x}}_k)}{\partial \underline{x}}\right|_{\underline{x} = \underline{\hat{x}}_k} \underline{\partial x} + \underline{\varepsilon} \tag{5}$$

The least squares estimate of the deviation, $dx_k = \underline{\hat{x}}_{k+1} - \underline{\hat{x}}_k$ follows from:

$$\min_{\underline{\partial x}} \left|\left| \underline{Y} - \underline{\bar{g}}(\underline{\hat{x}}_k) - \left.\frac{\partial \underline{\bar{g}}(\underline{\hat{x}}_k)}{\partial \underline{x}}\right|_{\underline{x} = \underline{\hat{x}}_k} \underline{\partial x} \right|\right|^2 \tag{6}$$

$$\underline{\hat{x}}_{k+1} - \underline{\hat{x}}_k = \underline{dx}_k$$

$$= \left(\frac{\partial \underline{\bar{g}}^T(\underline{x})}{\partial \underline{x}} \frac{\partial \underline{\bar{g}}(\underline{x})}{\partial \underline{x}}\right)^{-1} \frac{\partial \underline{\bar{g}}^T(\underline{x})}{\partial \underline{x}} \left(\underline{Y} - \underline{\bar{g}}(\underline{x})\right)\Big|_{\underline{x} = \underline{\hat{x}}_k} \tag{7}$$

It should be noted that in practical use the correction term, $\underline{\hat{dx}}_k$, is multiplied by a scalar which is optimized during each iteration [3]. We are not considering this.

As was mentioned above, if the $\underline{x}$ used for generating the measurement function matrix, $\underline{\bar{g}}(\underline{x})$, and its partial derivatives is held fixed (no relinearization), then $\underline{\hat{x}}_{k+1}$ above can be arrived at through a set of sequential estimator equations. These are derived in the same way that the recursive least squares algorithm

can be derived from the off-line (batch) least squares solution [13]. One obtains:

$$\hat{\underline{x}}_{k+1} = \hat{\underline{x}}_k + \underline{K}_{k,i}[y_i - h_i(\hat{\underline{x}}_k) - \underline{H}_i'(\hat{\underline{x}}_k)(\hat{\underline{x}}_k - \hat{\underline{x}}_0)]$$

$$\underline{K}_{k,i} = \underline{P}_k \underline{H}_i'^T(\hat{\underline{x}}_k)[\underline{H}_i'(\hat{\underline{x}}_k)\,\underline{P}_k \underline{H}_i'^T(\hat{\underline{x}}_k) + \underline{R}_i]^{-1} \qquad (8)$$

$$\underline{P}_{k+1} = (\underline{I} - \underline{K}_{k,i}\underline{H}_i'(\hat{\underline{x}}_k))\underline{P}_k$$

where $h_i(\underline{x})$ is the measurement function from (2) and $\underline{H}_i(x)$ is the nx1 vector of partial derivatives of $h_i(\underline{x})$. The subscripts on the gain expression, k and i, represent the iteration and randomly chosen measurement, respectively. $\underline{R}_i$ is the measurement noise covariance matrix and $\underline{P}_k$ is the error covariance matrix.

The $\hat{\underline{x}}_{k+1}$ of (8) can be made arbitrarily close to the $\hat{\underline{x}}$ resulting from a single iteration of (7) on the same k+1 measurements that the sequential estimator (8) has processed if the sequential estimator is properly initialized. This is usually accomplished by setting the initial covariance matrix, $\underline{P}_0$, to a diagonal matrix with arbitrarily large diagonal terms [13,14]. While in linear sequential estimators $\underline{x}_0$ is often set at zero, this is not appropriate in the nonlinear case where the measurement function partial derivatives are functions of $\underline{x}$. The initial estimate of $\underline{x}$ is instead used.

The above sequential estimator equations correspond to the "linearized" filter of Gelb [15] for the model (2). That is, $\hat{\underline{x}}_0$ is the reference estimate (trajectory) and the filter equations

are linearized about it. If the equations are relinearized after each measurement ($\hat{x}_0=\hat{x}_k$), the last term in the equation for $\hat{x}_{k+1}$ drops out and (8) represents the extended Kalman filter equations for (2).

3.3. The Correction Vectors $\underline{dx}_k$

How do the correction vectors, $\underline{dx}_k$, generated by the RSKF compare to those produced by one of the standard iterative descent techniques?

In an iterative descent technique, the vector of all parameters, at the k+1 iteration, is:

$$\hat{\underline{x}}_{k+1} = \hat{\underline{x}}_k - a\underline{\Lambda}(\hat{\underline{x}}_k)\underline{\nabla E}(\hat{\underline{x}}_k) \tag{9}$$

where $\underline{\nabla E}(\hat{\underline{x}}_k)$ is the gradient of the error function, $\underline{\Lambda}(\hat{\underline{x}}_k)$ is a matrix depending on the particular descent technique employed and the scalar a is chosen small enough to assure convergence. For a square error criterion.

$$\begin{aligned}\underline{\nabla E}(\hat{\underline{x}}_k) &= \underline{\nabla} \sum_{j=1}^{N} [y_j - h_j(\hat{\underline{x}}_k)]^2 \\ &= -2 \sum_{j=1}^{N} \underline{H}_j'^{T}(\hat{\underline{x}}_k)[y_j - h_j(\hat{\underline{x}}_k)]\end{aligned} \tag{10}$$

where $\underline{H}_j'(\hat{\underline{x}}_k)$ is the matrix of partial derivatives of $h_j(\hat{\underline{x}}_k)$ with respect to the parameters.

Thus the iterative descent technique is

$$\hat{\underline{x}}_{k+1} = \hat{\underline{x}}_k + 2a\underline{\Lambda}(\hat{\underline{x}}_k) \sum_{j=1}^{N} \underline{H}_j'^{T}(\hat{\underline{x}}_k)[y_j - h_j(\hat{\underline{x}}_k)] \tag{11}$$

while the RSKF is

$$\begin{aligned}\hat{\underline{x}}_{k+1} &= \hat{\underline{x}}_k + \underline{P}_k[\underline{H}_i'(\hat{\underline{x}}_k)\underline{P}_k\underline{H}_i'^{T}(\hat{\underline{x}}_k) + \underline{R}_k]^{-1}\underline{H}_i'^{T}(\hat{\underline{x}}_k)[y_i - h_i(\hat{\underline{x}}_k)] \\ \underline{P}_{k+1} &= (\underline{I} - \underline{K}_{k,i}\underline{H}_i'(\hat{\underline{x}}_k))\underline{P}_k\end{aligned} \tag{12}$$

In the RSKF estimation equation, (12), the magnitude of the correction vector, $\underline{\hat{x}}_{k+1}-\underline{\hat{x}}_k$, is inversely proportional to a measure of the uncertainty due to both the measurement noise and the uncertainty in the current estimate. The matrix $\underline{P}_k$ weights the correction vector components so that directions corresponding to greater uncertainty are favored. Most importantly, the vector summation operation of 'improvement' vectors, $\underline{H}_i'(\underline{\hat{x}}_k)$, weighted by the pointwise errors, $(y_i-h_i(\underline{\hat{x}}_k))$, has been replaced in the RSKF by only a single such term.

What can be said about this approximation as it affects the RSKF? By itself, minimization of the error at a single point (given more than one linear parameter) is an undetermined problem. The intention is that <u>over a number of iterations</u>, that is on average, the direction of the correction vectors will be towards decreasing error, with respect to the error surface of all the available measurements.

At each iteration of the RSKF the single measurement that is processed leads to a correction vector. Depending on which measurement is randomly chosen for processing, different correction vectors will result. It might initially be believed that they would all be roughly of the same direction and magnitude but this is not necessarily true. Individual iterations of the RSKF may produce estimates worse than the previous estimates in spite of a trend of reduced error over a number of iterations.

One might also suppose that at least close to the error minimum, the possible correction vectors would have similar characteris-

tics. While this is true to some extent, very close to the error minimum it becomes important to appreciate the effect of the measurement noise in the RSKF equations.

There are two sources accounting for the pointwise errors:

$$y_1-h_1(\underset{\sim}{\hat{x}}_k)\ ,\ y_2-h_2(\underset{\sim}{\hat{x}}_k),\ldots\ldots\ y_N-h_N(\underset{\sim}{\hat{x}}_k)$$

One is the measurement noise. Even for the optimal $\underset{\sim}{\hat{x}}_k$, the differences are non-zero. This can be thought of as a random error. The second is due to using an estimate of the parameters that is not the optimal estimate. This can be thought of as a systematic error. Such concepts are discussed in the regression and filtering literature [11].

If ones' estimates are far from the optimal parameter estimates this second source of error dominates. But close to the optimal parameter estimates the measurement noise is more appreciable. As the optimum is approached, the pointwise errors above take on the statistical characteristics of the measurement noise. That is, they are approximately zero mean Gaussian random variables of variance σ^2.

Since the correction vectors are proportional to the pointwise errors, close to the optimum they will be pointing 180° away from the direction they would have, had there been no measurement noise, almost 50% of the time (See Appendix A). Also the correction vectors' magnitude becomes more dependent on the measurement noise.

This in itself is not fatal to the RSKF's operation. The same sort of statement could be made concerning clearly optimal linear estimation procedures. It is mentioned to show the effect of the measurement noise on the RSKF's correction vectors.

Finally, with a qualification, the RSKF is as likely to converge upon a local minimum as any iterative descent technique. The qualification is that because the RSKF does not move strictly in the direction of decreasing error, it conceivably could leave the region of a local minimum. Conceivably too, an unfortunate choice of correction vector could place the estimates close to a local minimum.

3.4 the RSKF with Ficticious Measurements Noise

In Chapter 2 it was pointed out that if the measurements are causally processed through on extended Kalman filter, divergence will occur simply because many values of parameter estimates provide almost equally good fits. While random sampling eliminates this difficulty, divergence can still occur for other reasons.

Divergence in extended Kalman filters occurs when error sources that are unaccounted for do, in reality, exist so that the error covariance matrix elements take on unrealistically low values. One very simple modification of the filter equations, described below, has been used with some success in practice to overcome this problem.

A first order approximation of the effect of uncertainties in the parameter estimates upon the measurement equation is:

$$y_i = h_i(\hat{\alpha}_k) + \frac{dh_i(\hat{\alpha}_k)}{dx_1} dx_1 + \frac{dh_i(\hat{\alpha}_k)}{dx_2} dx_2 + \ldots . v_i \qquad (13)$$

This is the measurement model used in both Gauss-Newton optimization and the extended Kalman filter where

$$\underline{x}_{k+1} - \underline{x}_k = \underline{dx}_k \tag{14}$$

and one can linearly solve for $\underline{dx}_k$ and hence $\underline{x}_{k+1}$ [3].

However, proceeding with the philosophy used in Tenney [2], $\underline{dx}_k$ can be considered to be a zero mean Gaussian random vector whose covariance matrix is the already available parameter estimated covariance matrix, $\underline{P}_k$. Assuming that the ficticious measurement noise is independent of the actual measurement noise. the new measurement noise covariance matrix is:

$$\underline{R}_{k,i}^{new} = \underline{R}_i + \underline{H}_i'(\underline{\hat{x}}_k)\underline{P}_k\underline{H}_i'^{T}(\underline{\hat{x}}_k) \tag{15}$$

But the second term in the sum is already available in the Kalman denominator so that one has:

$$\underline{K}_{k,i} = \underline{P}_k\underline{H}_i'^{T}(\underline{\hat{x}}_k)\,[2\underline{H}_i'(\underline{\hat{x}}_k)\underline{P}_k\underline{H}_i'^{T}(\underline{\hat{x}}_k) + \underline{R}_i]^{-1} \tag{16}$$

This modification can be generalized. Consider the gain denominator to consist of two measures of uncertainty, one due to the measurement noise and one due to the inaccuracy in the current parameter estimates. The nominal weighting of these two quantities is 1:1. A weighting of 2:1 as indicated in (16) corresponds to the statistical model of (13). But certainly other weights are possible. That is, we might use

$$\underline{K}_{k,i} = \underline{P}_k\underline{H}_i'^{T}(\underline{\hat{x}}_k)\,[a_k\underline{H}_i'(\underline{\hat{x}}_k)\underline{P}_k\underline{H}_i'^{T}(\underline{\hat{x}}_k) + \underline{R}_i]^{-1} \tag{17}$$

where a_k is suitably chosen.

As a_k is made greater than one, the rate of filter convergence decreases as the filter behaves more cautiously, believing an increasing amount of uncertainty is present in the measurements. Specifically, when the RSKF's estimates are far from the optimal estimates, $\underline{H}_i'(\hat{\underline{x}}_k)\underline{P}_k\underline{H}_i'(\hat{\underline{x}}_k)$ should be much greater than $\underline{R}_i$ so that the magnitude of the correction vector is $1/a_k$ of what it would be had $a_k=1$.

Examples of the filters' performance using different values of a_k are presented in Chapter 4.

Although a weight of $a_k=2$ corresponds to the statistical model of (13), one should not conclude that this is the best possible a_k setting. Rather the model of (13) is a heuristic device for adding an (not necessarily minimal or even sufficient) amount of ficticious measurement noise in order to prevent divergence.

The statistically unsatisfactory nature of (13) is two fold. First, even though the ficticious measurement noise random variables, $\underline{dx}_k$ and the parameter uncertainty, $\hat{\underline{x}}_k-\underline{x}_k$, are the same random

variables, in going from (15) to (16) we implicitly assume that they are uncorrelated. Secondly, the effect of parameter uncertainties in the partial derivative evaluations has not been included.

The former point can be demonstrated explicitly using Schmidt's [16] model for including the effect of parameter uncertainties in the measurement equation:

$$Y = \underline{M}(\underline{X},\underline{V},t) + q(t)$$

$$y = \left[\frac{\partial \underline{M}}{\partial \underline{X}}\right]_{\underline{X}=\hat{\underline{X}}} \underline{x} + \left[\frac{\partial \underline{M}}{\partial \underline{V}}\right]_{\underline{V}=\underline{V}_0} \underline{v} + q(t)$$

$$y = \underline{H}'(t)\underline{x} + \underline{G}'(t)\underline{v} = q(t)$$

$$E(\underline{v}) = \underline{0} \quad E(q) = 0$$

$$E(\underline{v}\underline{v}^t) = \underline{W} \quad E(qq^t) = Q$$

$$\underline{C} = E[(\underline{x}-\hat{\underline{x}})\underline{v}^t] \quad P = E[(\underline{x}-\hat{\underline{x}})(\underline{x}-\hat{\underline{x}})^t]$$

Here y,x and v are the deviations from the nominal estimates of Y,X and V. These are the measurements, states (parameters) and uncertain parameters, respectively.

The resulting gain and covariance updating expressions are:

$$\underline{K} = [\underline{P}\,\underline{H}'^T + \underline{C}\,\underline{G}'^T][\underline{H}'\underline{P}\,\underline{H}'^T + \underline{H}'\underline{C}\,\underline{G}'^T + \underline{G}'\underline{C}\,\underline{H}'^T + \underline{G}'\underline{W}\,\underline{G}'^T + \underline{R}]^{-1}$$

$$\underline{P} = [\underline{I}-\underline{K}\,\underline{H}']\underline{P} - \underline{K}\,\underline{G}'\underline{C}^T$$

These reduce to the expressions for the gain in (16) and the covariance in (8) if $\underline{C} = 0$, $\underline{W} = \underline{P}$ and $\underline{G}' = \underline{H}'$. These are the assumptions under which Schmidt's model is equivalent to (13).

We also mentioned that the effect of parameter uncertainties in the partial derivative evaluations has not been included.

Accounting for it using a measurement noise formulation is not directly possible. If one expands the partial derivatives in the linearized measurement equation (13) in a Taylor series about the current estimate, the first order terms in the series (which are second order partial derivatives) are multiplied by $\underline{dx}_k$. However, as discussed in sec. 3.2, the extended Kalman filter implicitly assumes this term to be zero. To be more specific, it assumes $\underline{dx}(k+1/k)$, that is $\underline{dx}$ at k+1 given k measurements, to be zero and linearly solves for $\underline{dx}(k+1/k+1)$.

Considering $\underline{dx}_k$ to be a random variable, as in sec. 3.4, becomes statistically involved. The following quadratic term must then be added to the linearized measurement equation

$$\underline{dx}_k^T \frac{\partial \underline{H}'(\underline{\hat{x}}_k)}{\partial \underline{x}_k} \underline{dx}_k$$

where $\underline{dx}^T$ comes from the partial derivative expansion and $\underline{dx}_k$ from the measurement function expansion.

Whether one considers these to be the same random vector ($\eta(0,\underline{P}_k)$) or uncorrelated random vectors with identical distributions, utilizing the resulting non-Gaussian distribution is a problem.

3.5 The Matrix $\underline{P}_k$

In the previous discussion of the RSKF little has been said about the matrix $\underline{P}_k$ except that it favorably weights correction vector components towards the region of greatest uncertainty. It has no strict statistical meaning in the RSKF. It arises naturally in linear Kalman filtering since it and $\underline{\hat{x}}_k$ describe the statistical distribution of the current parameter estimate as being $\underline{N}(\underline{\hat{x}}_k, \underline{P}_k)$, when the system and obser ation noise is additive and Gaussian.

For nonlinear filtering using an excellent reference trajectory, one might hope that $\underline{P}_k$ is still meaningful. The often-encounted Kalman filter divergence indicates that this is not necessarily so. One is attempting to represent the non-Gaussian conditional distribution function (the estimate distribution conditioned on the data received up till the present iteration) with only two moments. In the problem described by (1), a good reference trajectory is not assumed to be available. This makes matters worse. One's only hope is that after an initial rapid convergence to a neighborhood near the optimum, the conditional density is asymptotically Gaussian through a central limit type operation. The empirical success of the 'ficticious' measurement noise suggests that considering $\underline{dx}_k$ as as $N(0,\underline{P}_k)$ may indeed be meaningful. The random sampling will also tend to eliminate correlations in the error(due to inaccuracy in the current estimate)between samples that are adjacent in real time. However, this should be considered as vague speculation. At this time we can say nothing definite.

3.6 Computational Considerations

Why would one employ an estimator such as the RSKF instead of one of the well known iterative minimization techniques? One reason might be computational.

This involves the difference between a RSKF "iteration" and an iterative minimization technique "iteration". During the former only a single pointwise error evaluation and partial derivative matrix calculation is made. In the latter type of "iteration", such quantities are calculated for all of the available data.

Even if the latter algorithm converges in many fewer of its "iterations", each one involves many more nonlinear function evaluations. This computational cost could be significant.

Furthermore, the RSKF may be stopped when sufficiently accurate estimates are obtained. Although the waveform measurements need be stored off-line, not all of them need be processed.

To examine the question of computation time in some detail, let:

N = the number of measurements.

IMT = iterative technique (I.M.T.) of the form (8).

D = the number of computational units necessary to evaluate $h_i(\hat{\underline{x}}_k)$ and $\underline{H}_i^T(\hat{\underline{x}}_k)$. Henceforth, a computational unit is assumed to be a single multiplication.

M_{IMT}, M_{RSKF} = the number of computational units necessary to perform one iteration of an IMT or RSKF, respectively, disregarding $M_{F/D}$ and logic costs (program loops, etc...).

I_{IMT}, I_{RSKF} = the number of iterations for an IMT or RSKF, respectively, to sufficiently converge.

t_{IMT}, t_{RSKF} = the time necessary for each of the two techniques, to converge, expressed in computational units.

It will be assumed that it is sufficient to compare only the number of multiplications and function/derivative evaluations. In a first analysis logic costs will also be neglected. We have:

$$t_{IMT} = I_{IMT}[N \cdot M_{F/D} + M_{IMT}]$$
$$t_{RSKF} = I_{RSKF} \cdot (M_{F/D} + M_{RSKF}) \tag{18}$$

How large must I_{IMT} be before the RSKF becomes computationally more economical? This can be found by letting $t_{IMT} = t_{RSKF}$, for

$$\frac{(M_{F/D} + M_{RSKF})}{(N \cdot M_{F/D} + M_{IMT})} I_{RSKF} \tag{19}$$

The simplest iterative technique of the form (8), is steepest descent ($\underline{\psi}(\hat{\underline{x}}_k)=I$). Here

$$N \cdot M_{F/D} >> M_{IMT} \tag{20}$$

and we will assume

$$M_{RSKF} >> M_{F/D}$$
$$I_{RSKF} = N \tag{21}$$

Then (19) reduces to

$$\frac{M_{RSKF}}{M_{F/D}} \tag{22}$$

The inclusion of logic costs will increase M_{IMT} and M_{RSKF} though the increase in the latter will be most significant both because of the more complex computational structure of the RSKF and because M_{RSKF} is a larger fraction of $(M_{F/D} + M_{RSKF})$ than M_{IMT} is of $(N \cdot M_{F/D} + M_{IMT})$, assuming that the inequalities of (20) and (21) are true. In Chapter 4 the inclusion of logic costs in a 5 parameter problem will be seen to increase M_{RSKF} by a factor of three.

Asymptotically, as N and/or $M_{F/D}$ increases, the RSKF is favored. The asymptotic effect of an increase in the number of parameters is more involved, not only because the convergence behavior may be affected, but also because the increase in $M_{F/D}$ and M_{IMT} will be favorable to the RSKF while that in M_{RSKF} will not be.

There are a number of other computationally related issues that involve the rate of convergence and accuracy of the estimate one obtains.

Numerically, the classical Kalman filter equations are known to be unreliable [4]. Our simulations have been obtained using double precision on an IBM 370 machine. There is no conceptual difficulty with utilizing the more recently developed Kalman filter formulations with improved numerical properties [4] though additional computation is involved.

Specifically, the term that is inverted in (17) does not appear in the inverse form of the Kalman filter equations, the square root ariance filter or the square root information filter. Because of this the use of the scalar a_k, as in (17), is not possible. However $\underline{R}$ can be replaced during each iteration with the increased $\underline{R}^{new}$ of (15). The evaluation of $\underline{R}^{new}$ requires about n^2 multiplications, where n is the number of parameters.

Since a random sequence determines the processing order, one might ask if there is an optimal order. Could the randomly selected samples be concentrated in certain sections of the waveform that contained the most information about the parameters so as to achieve a faster convergence? Alternately, given the waveform, what are the sampling locations which provide minimum variance estimates?

Unfortunately, expressions for the trace of the error covariance matrix, while independent of the parameters for a linear estimation problem, do indeed depend on the parameters to be estimated for nonlinear models.

Therefore, this approach could only be viable if very good a priori estimates of the nonlinear parameters are already available. It could be employed once the filter had sufficiently converged, but there is a real question concerning the cost of the extra computation needed to minimize the trace expression.

Another point is that if it is possible to process only one measurement per iteration, it is also possible to process several simultaneously during each iteration. This could reduce the number of iterations required for convergence but the additional measurement equations increase the Kalman filter's computational complexity.

Conceivably one could also process measurements that had already been used. In the simulations presented in the following chapter the RSKF made only a single pass through the data. While this seemed adequate, additional passes through the data could have been made in order to refine the estimate.

Chapter 4

An Application

4.1 Introduction

Sensors are available that are capable of measuring a ship's magnetic field intensity. Sunahara [5] has considered sensors placed under a harbor entrance for traffic control purposes. Aircraft, engaged in anti-submarine activity, may also carry such sensors. In this case, the submarine is assumed stationary and the sensor moves in a straight line relative to it.

The magnetic field of the ship can be modeled as originating from a magnetic dipole [6]. The field intensity measured by the sensor can in turn be approximated by some "suitable" deterministic function. Sunahara, who was interested primarily in detecting this "signal" in the presence of substanial noise, modeled the field intensity waveform as a single cycle of a sine function of unknown amplitude and phase. Another representation for the field intensity, in a simplified form is:

$$h(\theta_k) = (1+\theta_k^2)^{-5/2}(c_1+c_2\theta_k+c_3\theta_k^2)$$
$$\theta_k = S(t_k-T_0) \qquad (23)$$

This is somewhat representative of other tracking models in that the location parameter, T_0, represents the time at which the closest point of approach (CPA) between ship and sensor is made and the scale parameter, S, is proportional to the velocity of the moving object and inversely proportional to the distance between ship and sensor at CPA.

One artifice employed to insure a satisfactory filter performance was to replace S with e^{β} and estimate β. This replaces the physically constrained S (S>0) with the unconstrained β (otherwise, the filter may converge to a local minimum with a negative scale estimate). The variance in (25) is that of β.

4.2 Computational Comparison

Based on expressions for the computational requirements of Kalman filters appearing in Mendel's work [7], one finds that for a five parameter, single measurement equation problem, approximately 600 computational units are necessary for multiplication and 1500 if logic costs are included. (A logic operation is assumed to be ten times faster than a multiplication.) The minimum number of computational units necessary to evaluate (23) and the associated partial derivatives is about 64 (see Appendix C).

Assuming that $M_{F/D} >> M_{IMT}$ and $I_{IMT} = N = 400$ the number of IMT iterations at which the RSKF becomes computationally competitive is:

$$\frac{M_{F/D} + M_{RSKF}}{M_{F/D}} = \frac{64+1500}{64} \approx 25 \qquad (24)$$

if logic costs are included.

4.3 An Example

Several simulations were performed using an IBM 370 with double precision. The classical extended Kalman filter equations were used with the parameters:

	Actual	Initial Estimate	Initial Variance
S =	.016	.030	2.4
T_0 =	200.	250.	1600.
c_1 =	.1	4.0	160.
c_2 =	1.0	1.9	1.0
c_3 =	.15	-4.0	16.0
σ^2 =	10^{-4}	known	
N =	400	known	

(25)

$$-2.8 \leq \theta_k \leq 2.8$$

4.3.1 The Weight $a_k = 2.0$

Figs. 14-20 illustrate the operation of the RSKF with the weight in the Kalman gain denominator set equal to two.

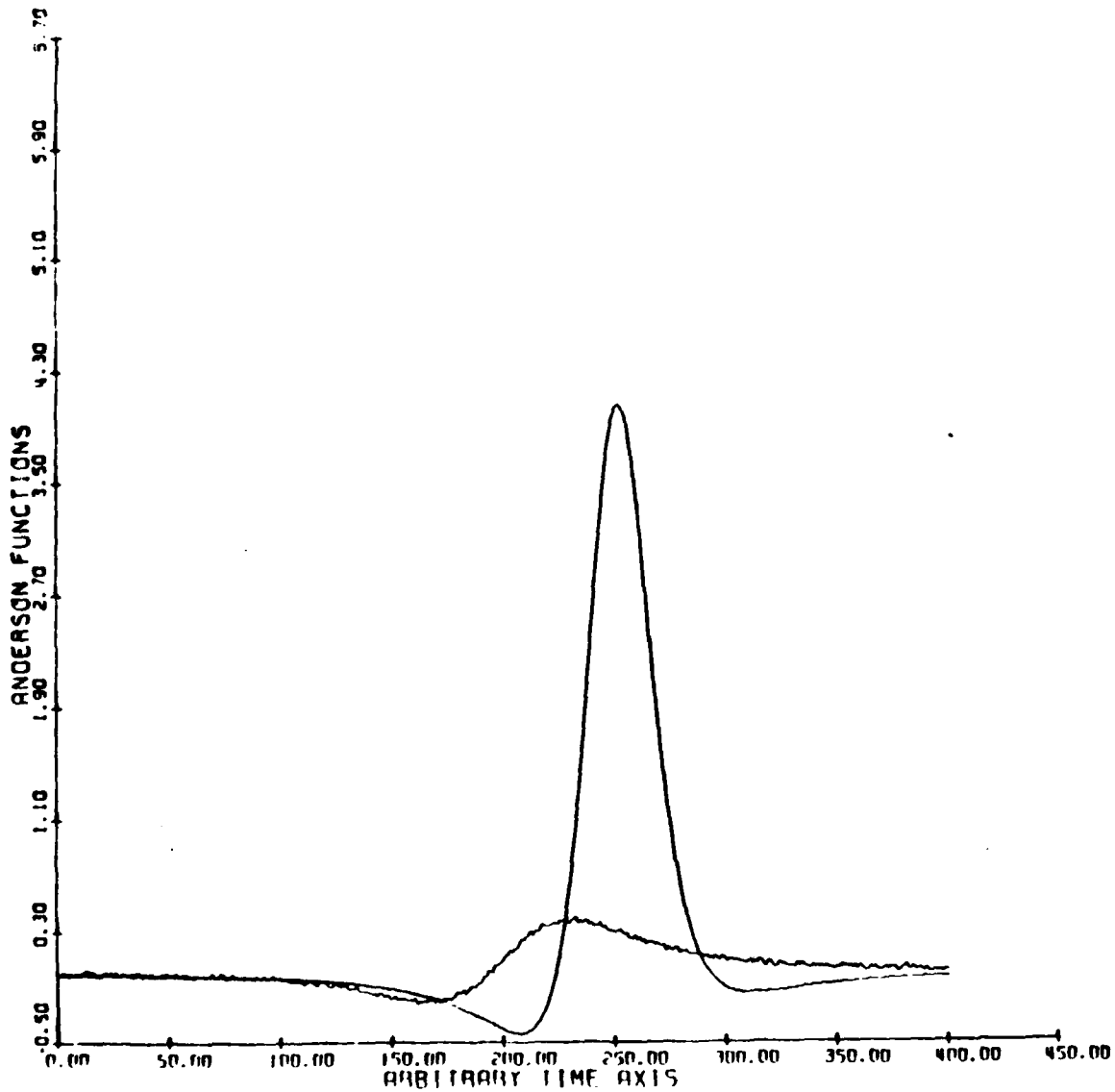
Fig. 14 illustrates the received waveform and the waveform corresponding to the initial parameter estimates.

Fig. 15 is a plot of a measure of the error of the parameter estimates. Specifically, the following quantity is plotted for each iteration.

$$\frac{1}{5}\left[\frac{|c_1-\hat{c}_1|}{c_1} + \frac{|c_2-\hat{c}_2|}{c_2} + \frac{|c_3-\hat{c}_3|}{c_3} + \frac{|S-\hat{S}|}{S} + \frac{|T_0-\hat{T}_0|}{T_0}\right] \qquad (26)$$

Fig. 16 is a plot in the nonlinear state space of the convergence of $\hat{S}$ and $\hat{T}_0$. It can be seen that the correction vectors generally decrease in magnitude as the optimum is approached. Furthermore, although the general trend of the trajectory is to approach the optimal parameter estimate, it can be seen that individual correction vectors often do not point toward it. This leads to a number of "knots" and "loops" in the trajectory.

Fig. 17 is a plot of the normalized covariance trace:



RECEIVED SIGNAL AND INITIAL ESTIMATE

Figure 14

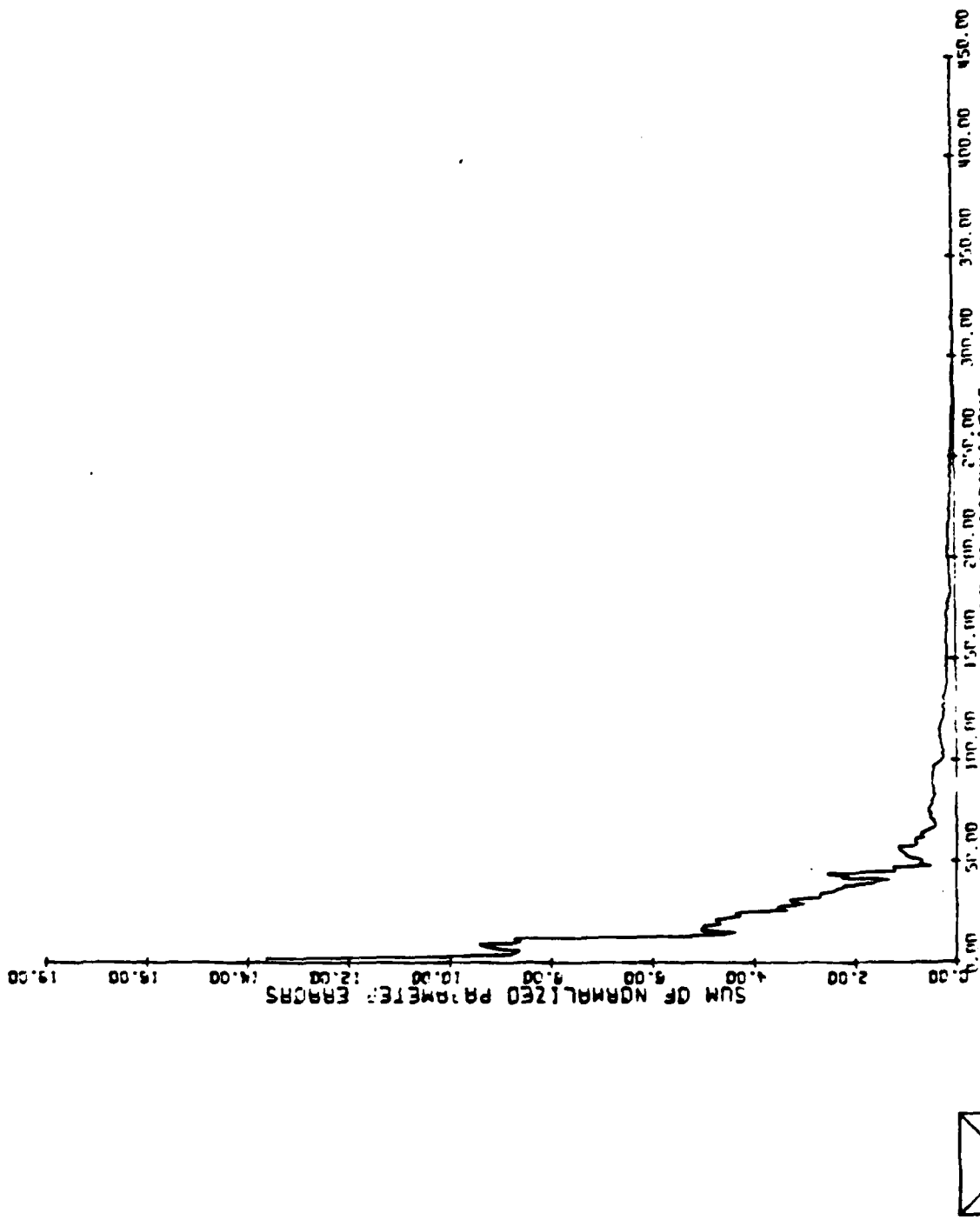

SUM OF NORMALIZED PARAMETER ERRORS
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
NUMBER OF ITERATIONS
400.00
450.00

Figure 15

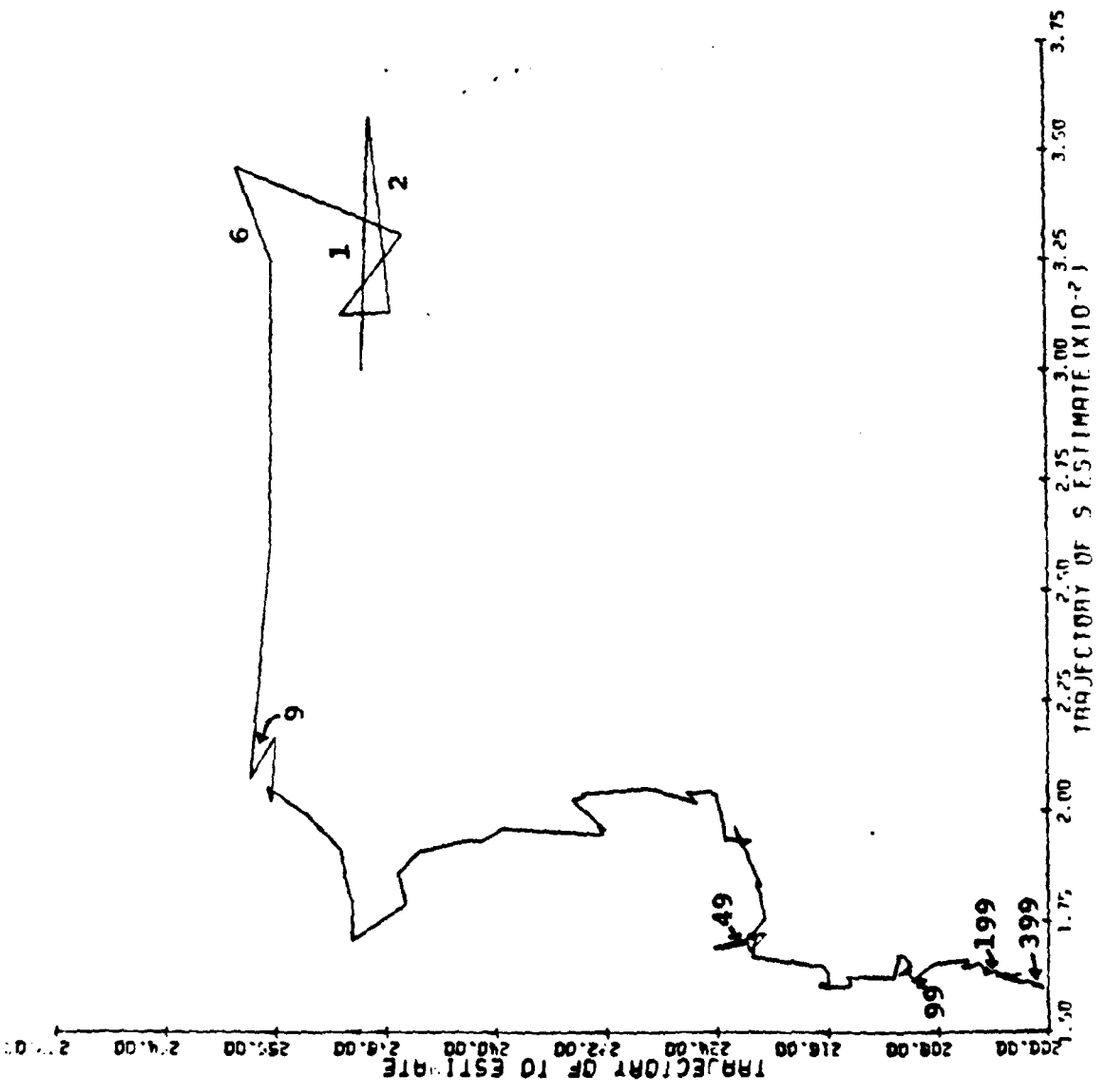

1
2
6
9
49
99
199
399
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.00
208.00
216.00
224.00
232.00
240.00
248.00

Figure 16

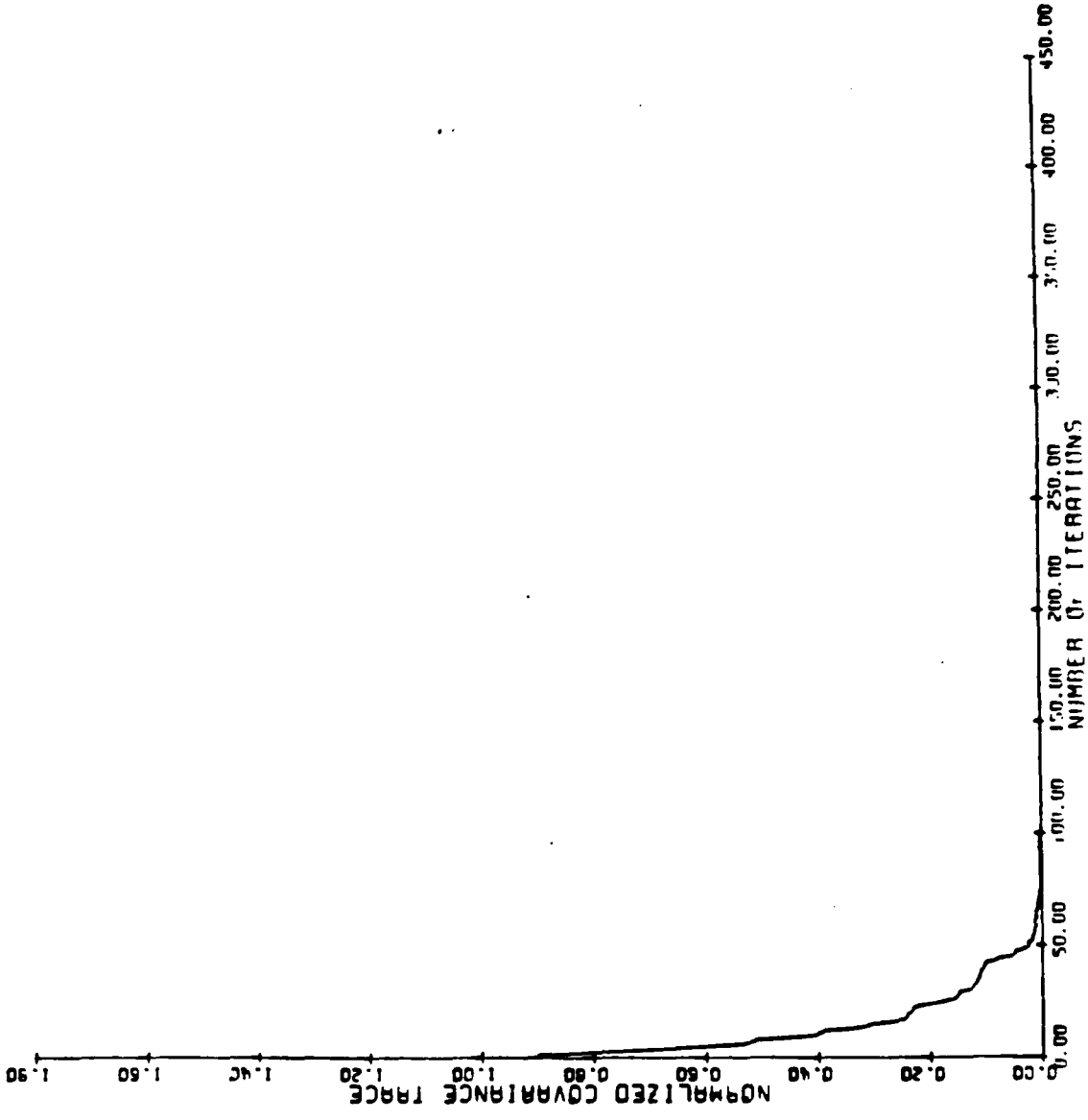

NORMALIZED COVARIANCE TRACE
0.00
0.20
0.40
0.60
0.80
1.00
1.20
1.40
1.60
1.90
0.00
50.00
100.00
150.00
200.00
250.00
300.00
350.00
400.00
450.00
NUMBER OF ITERATIONS

Figure 17

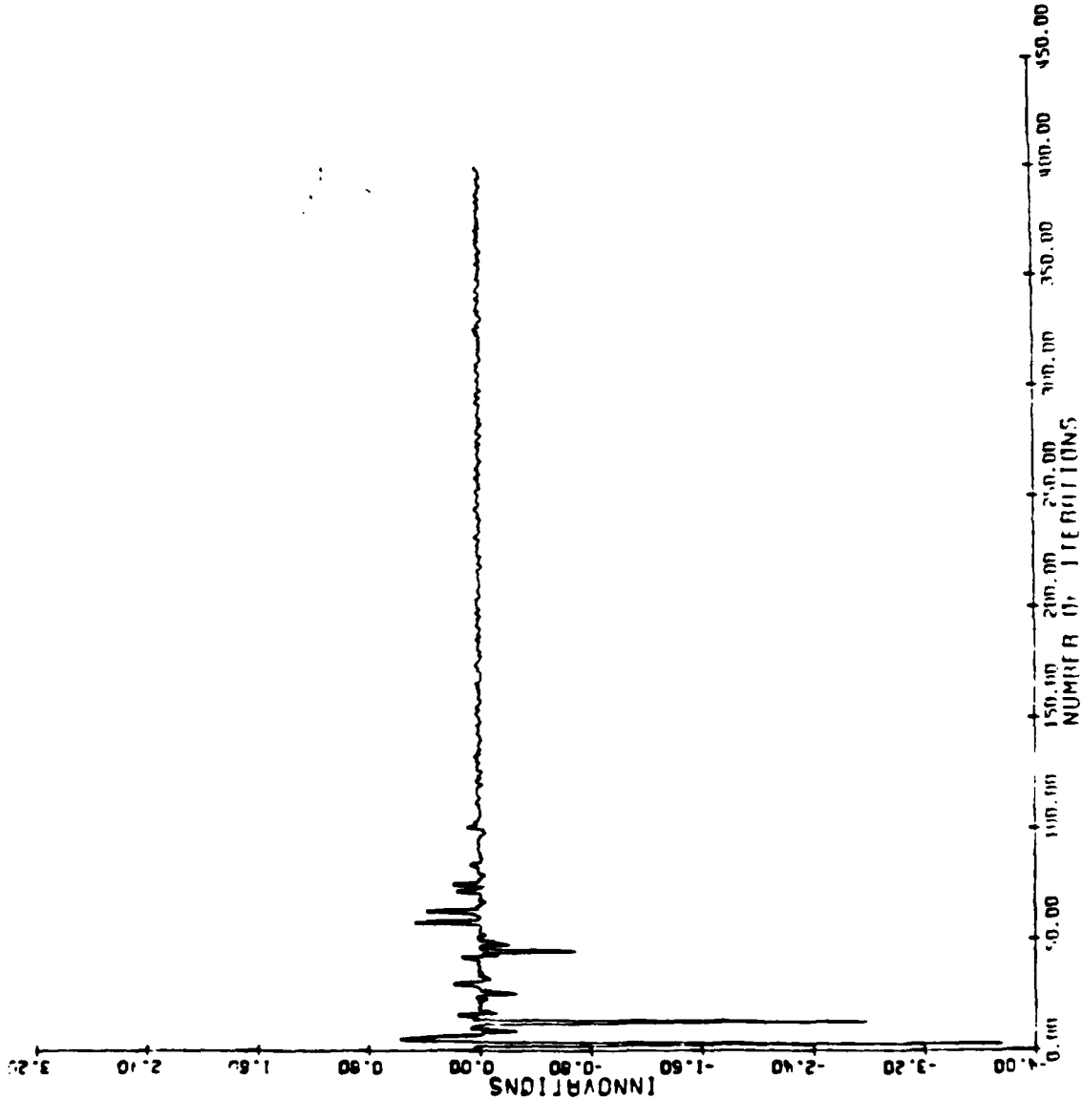


Figure 18

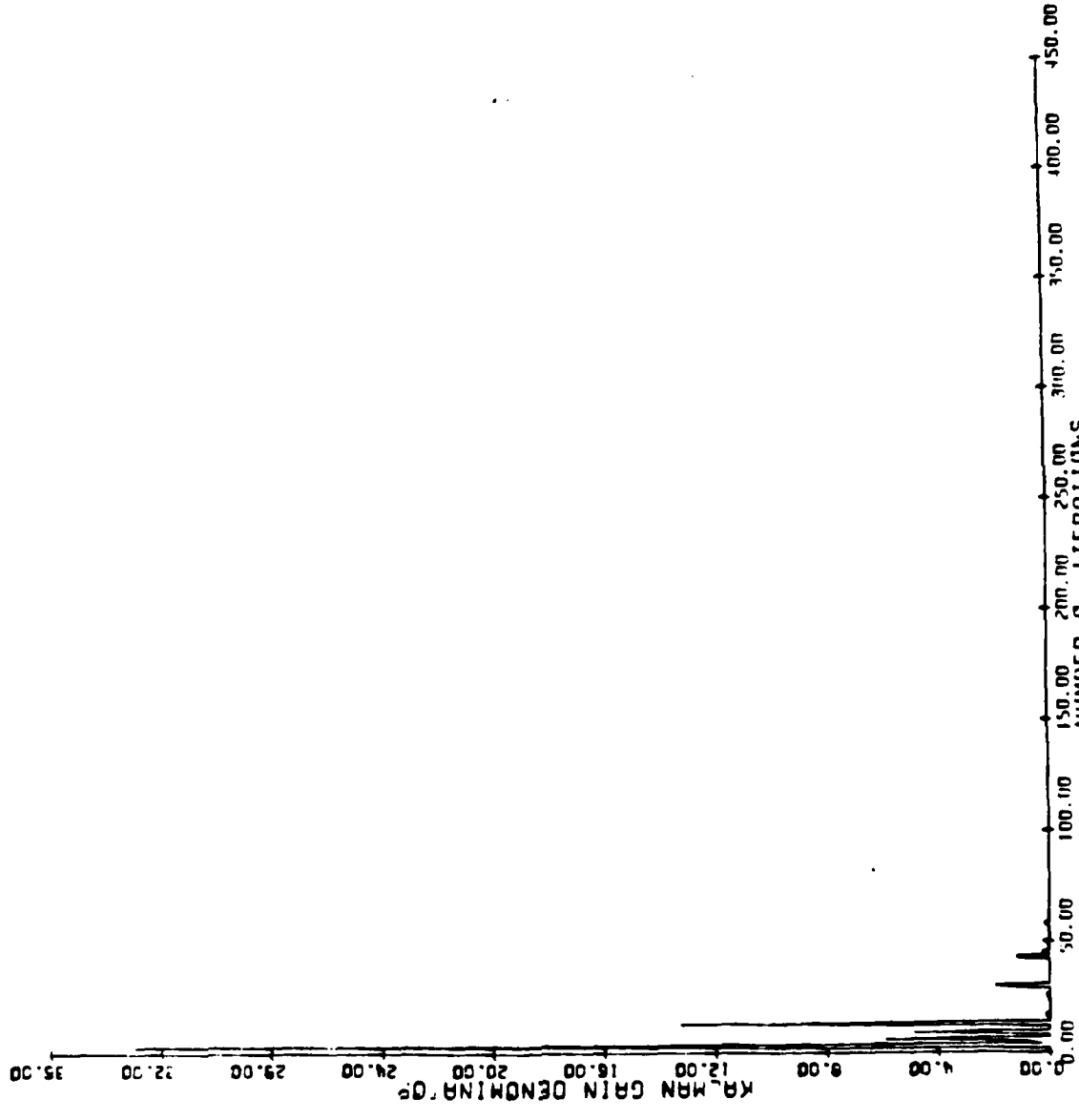

KALMAN GAIN DENOMINATOR
0.00
4.00
8.00
12.00
16.00
20.00
24.00
28.00
32.00
36.00
0.00
50.00
100.00
150.00
200.00
250.00
300.00
350.00
400.00
450.00
NUMBER OF ITERATIONS

Figure 19

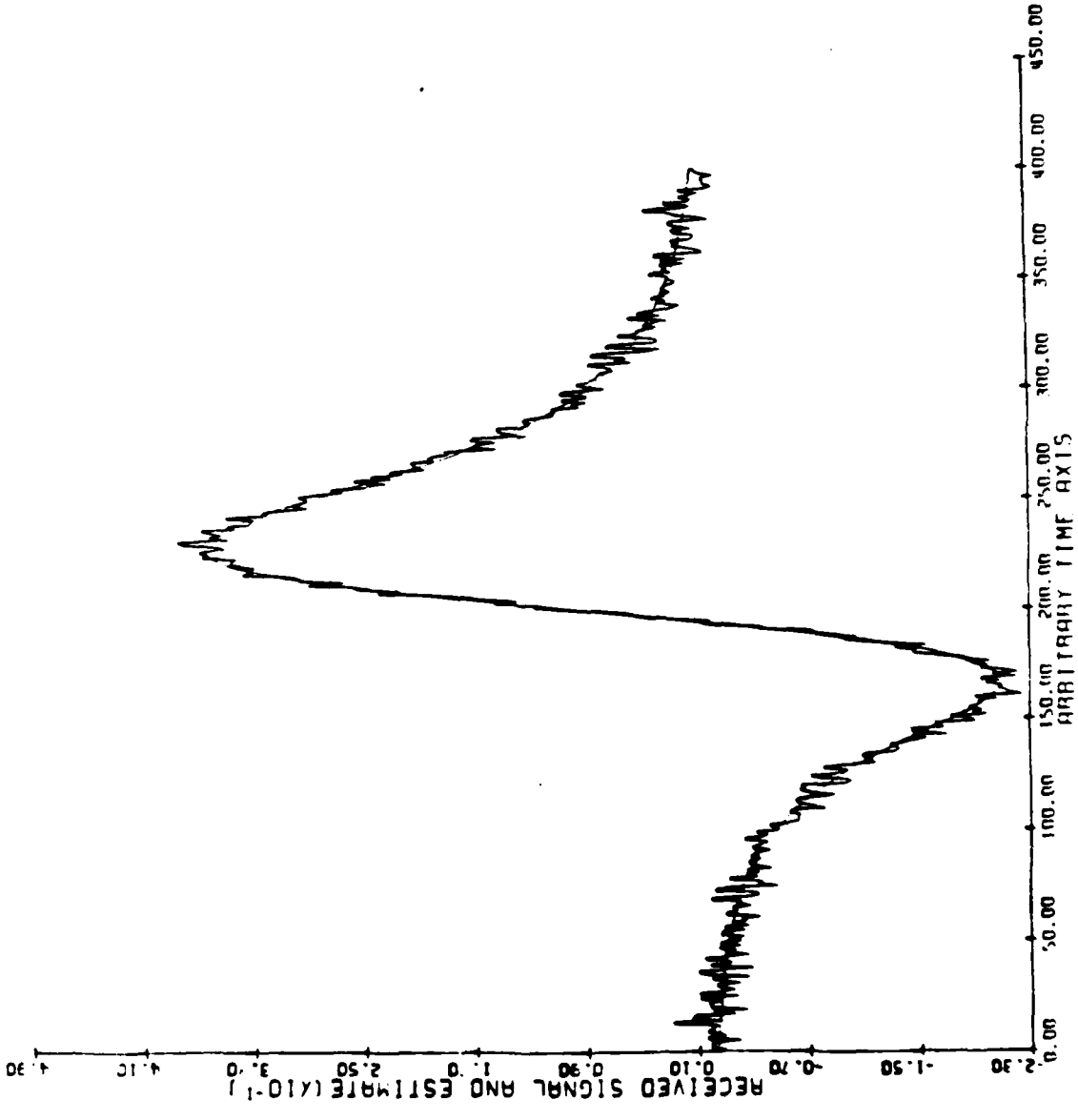

RECEIVED SIGNAL AND ESTIMATE (X10⁻¹)
-2.30
-1.50
-0.70
0.10
0.90
1.70
2.50
3.30
4.10
4.90
0.00
50.00
100.00
150.00
200.00
250.00
300.00
350.00
400.00
450.00
ARBITRARY TIME AXIS

Figure 20

$$\frac{1}{5}\left[\frac{P_k(1,1)}{P_0(1,1)} + \frac{P_k(2,2)}{P_0(2,2)} + \frac{P_k(3,3)}{P_0(3,3)} + \frac{P_k(4,4)}{P_0(4,4)} + \frac{P_k(5,5)}{P_0(5,5)}\right] \qquad (27)$$

Fig. 18 displays the filter innovations $(y_i - h_i(\hat{\alpha}_k))$. They seem to settle out at about 100-150 iterations. This is consistent with Fig. 17.

Fig. 19 displays the Kalman gain denominator. Since the measurement noise covariance component of the denominator is at about zero on the scale of the graph, one is seeing only the other component, the uncertainty due to the uncertainty in the curren parameter estimate. This term appears to have a 'spiky' nature. It generally decreases in amplitude.

Finally, Fig. 20 is the received waveform along with the waveform corresponding to the final parameter estimates:

		mean	variance
S	=	.0160	$.20 \cdot 10^{-4}$
T_0	=	201.	.16
c_1	=	.102	$.30 \cdot 10^{-4}$
c_2	=	.999	$.30 \cdot 10^{-4}$
c_3	=	.145	$.23 \cdot 10^{-3}$

(28)

These may be compared to the actual parameter values in (25).

The next eight plots (Figs. 21-28) illustrate the population of correction vectors that the RSKF randomly samples from. Each plot corresponds to a particular iteration of the previous example (a_k=2.0). The common point shared by all the vectors is the current estimate in the nonlinear state space. Each vector corresponds to the particular correction vector that would result if one of the

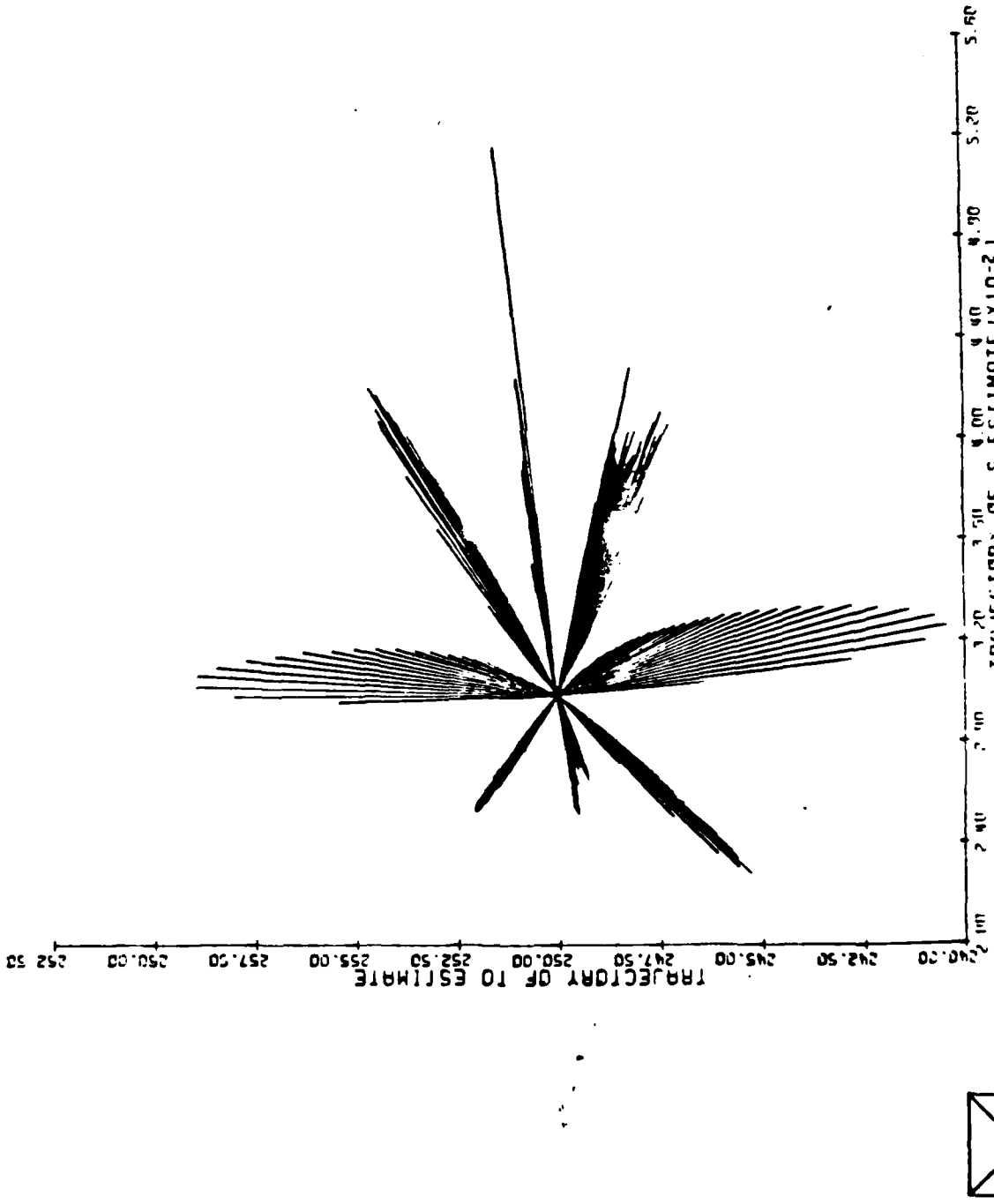

RSKF ITERATION NUMBER 1
TRAJECTORY OF TO ESTIMATE
242.50
245.00
247.50
250.00
252.50
255.00
257.50
260.00
262.50

Figure 21

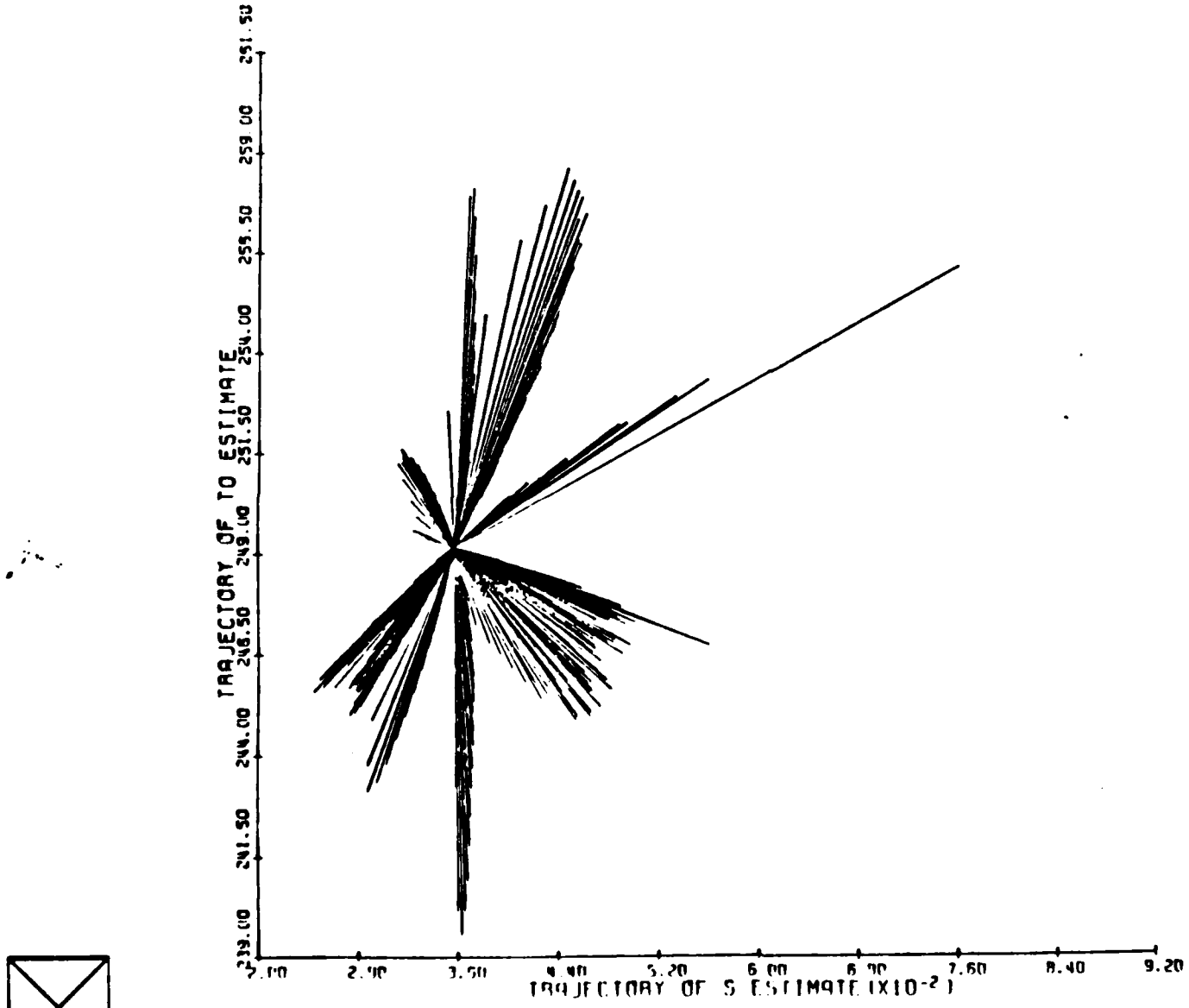
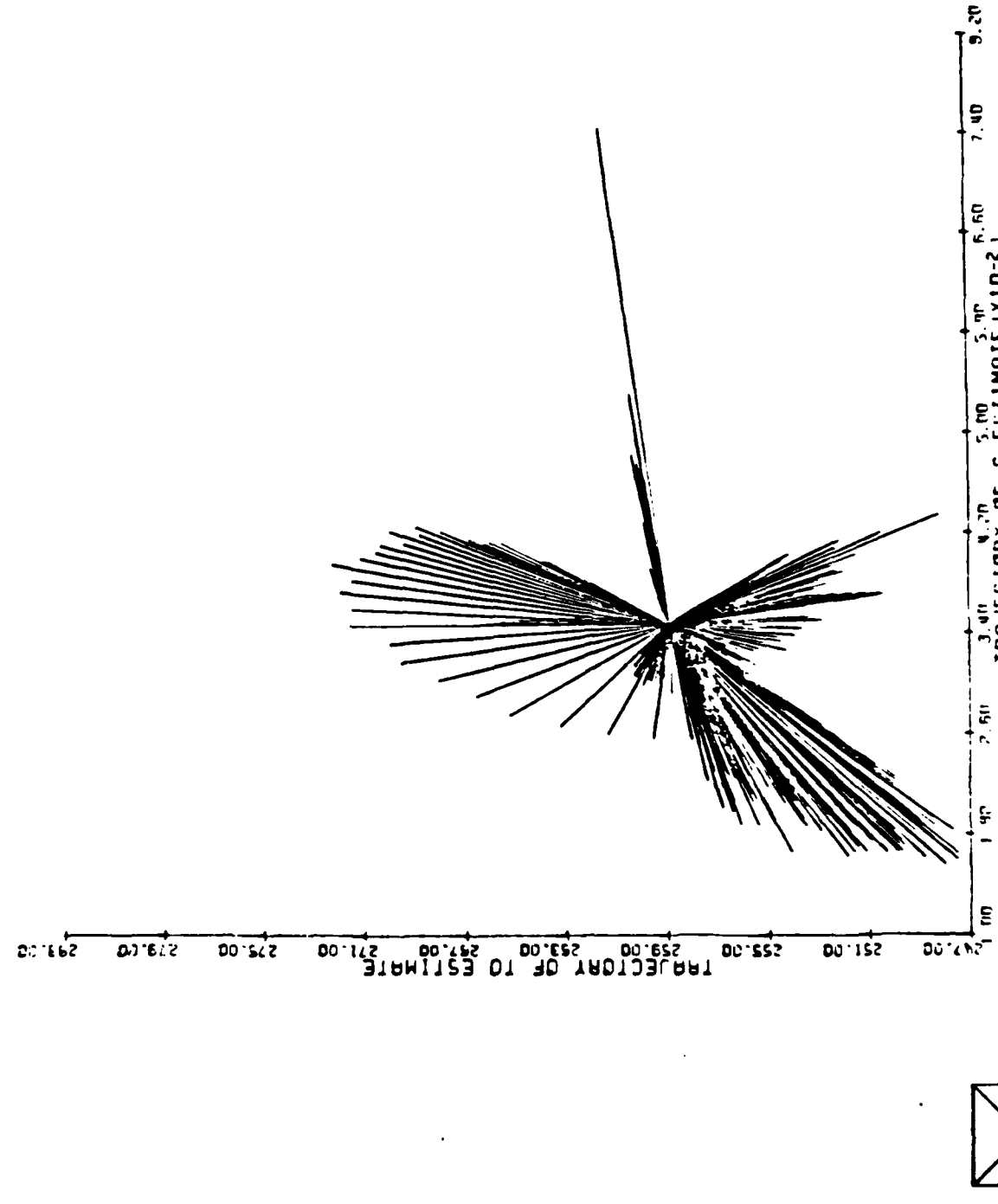
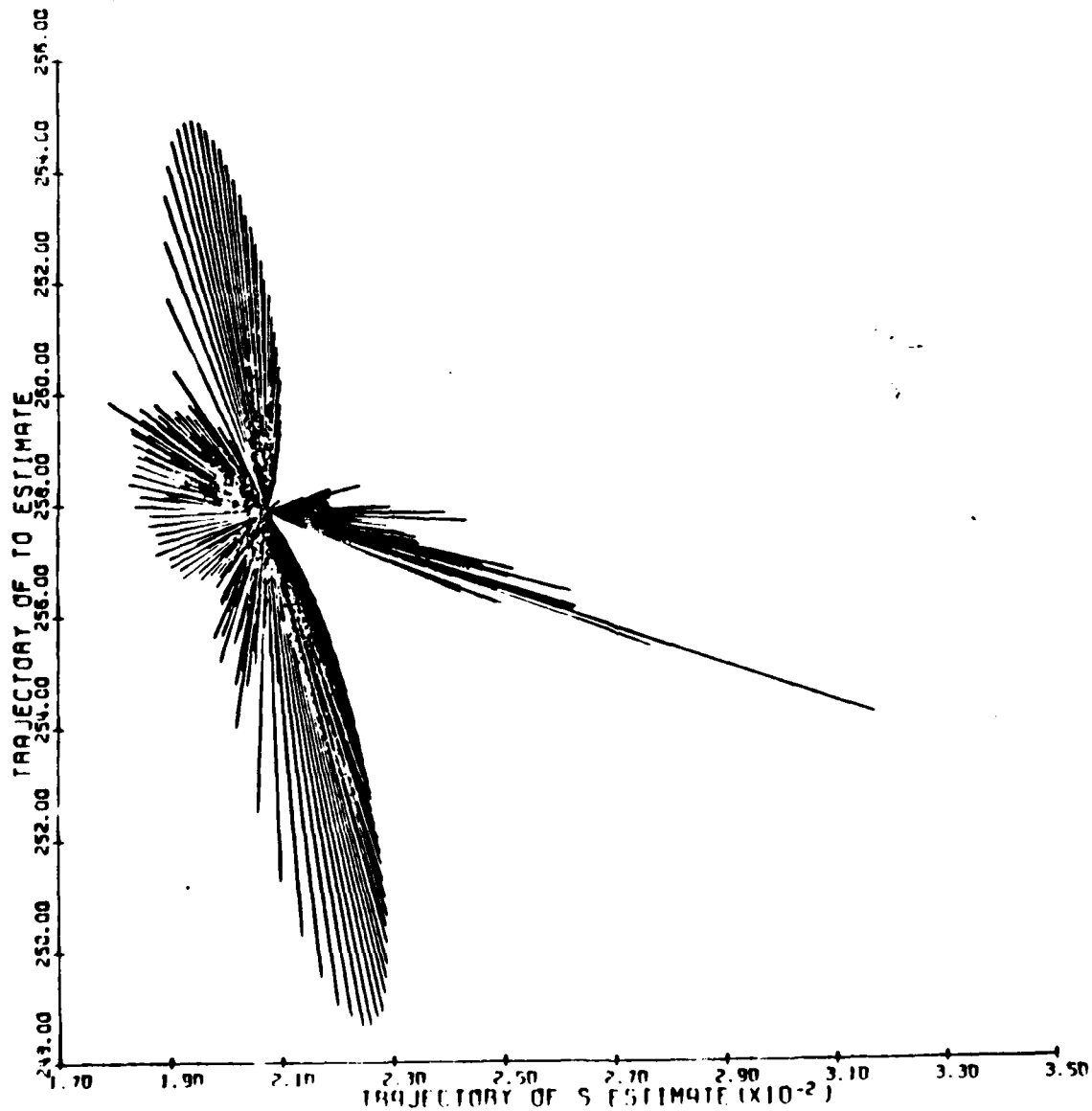

TRAJECTORY OF TO ESTIMATE
239.00
241.50
244.00
246.50
249.00
251.50
254.00
256.50
259.00
261.50
2.00
2.90
3.50
4.40
5.20
6.00
6.90
7.60
8.40
9.20
RSKF ITERATION NUMBER 2

Figure 22



RSKF ITERATION NUMBER 6



RSKF ITERATION NUMBER 9

Figure 24

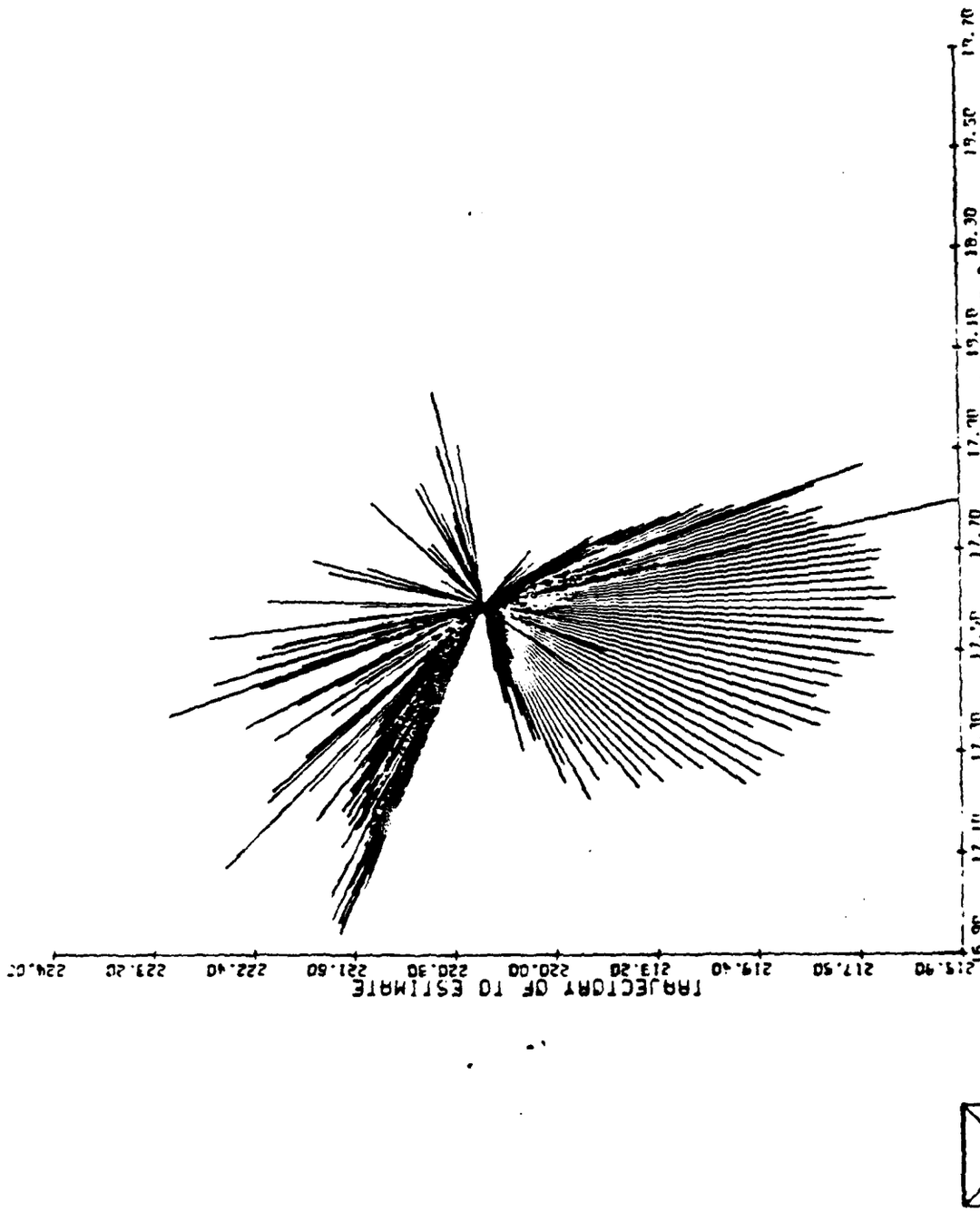

RSKF ITERATION NUMBER 49
TRAJECTORY OF S ESTIMATE (X10-3)
TRAJECTORY OF TO ESTIMATE
217.60
219.40
213.20
220.00
220.30
221.60
222.40
223.20

Figure 25

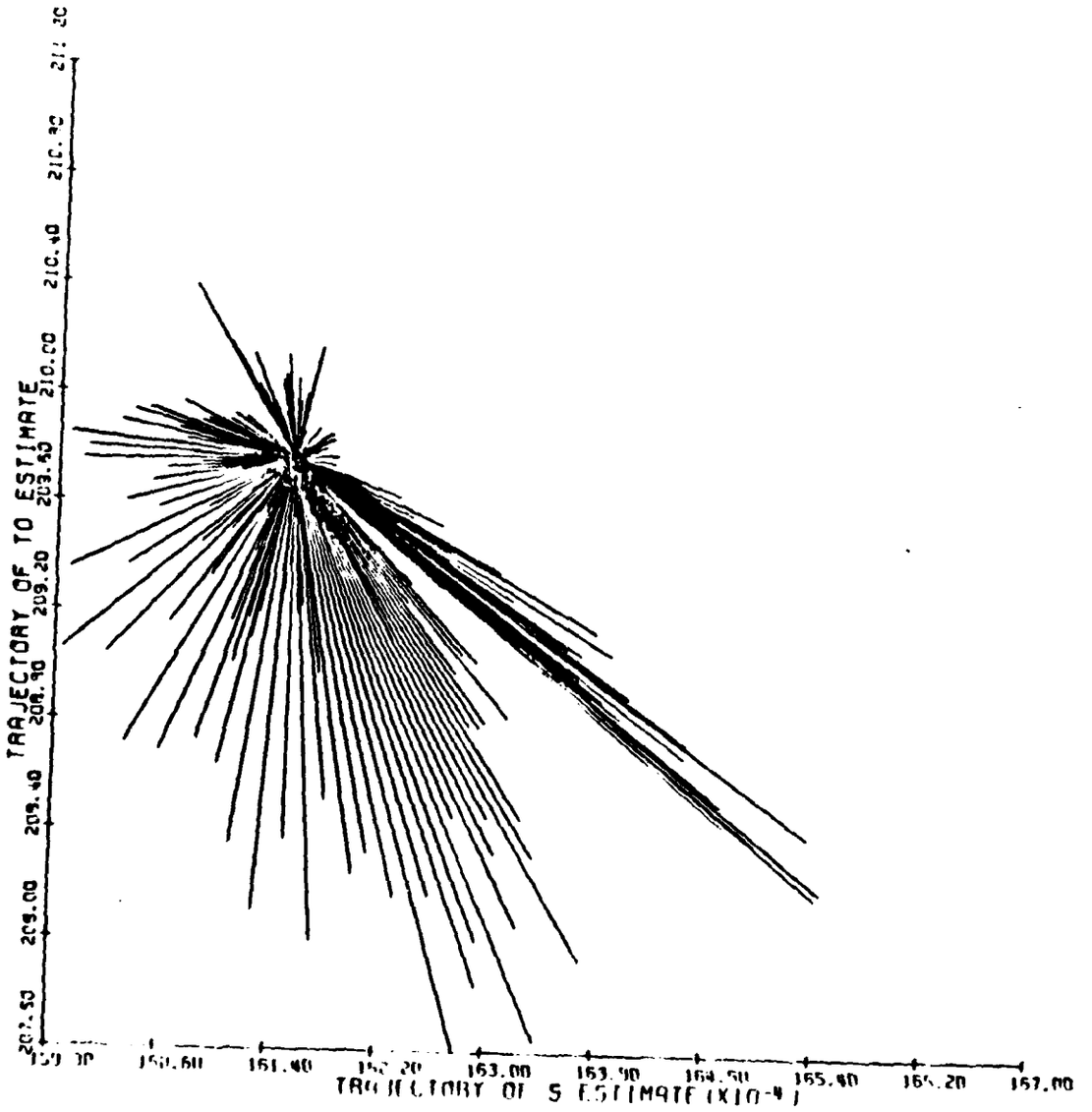

TRAJECTORY OF TO ESTIMATE
209.00
209.40
209.80
210.00
210.40
161.40
163.00
163.80
164.60
165.40
167.00
TRAJECTORY OF S ESTIMATE (X10-4)
RSKF ITERATION NUMBER 99

Figure 26

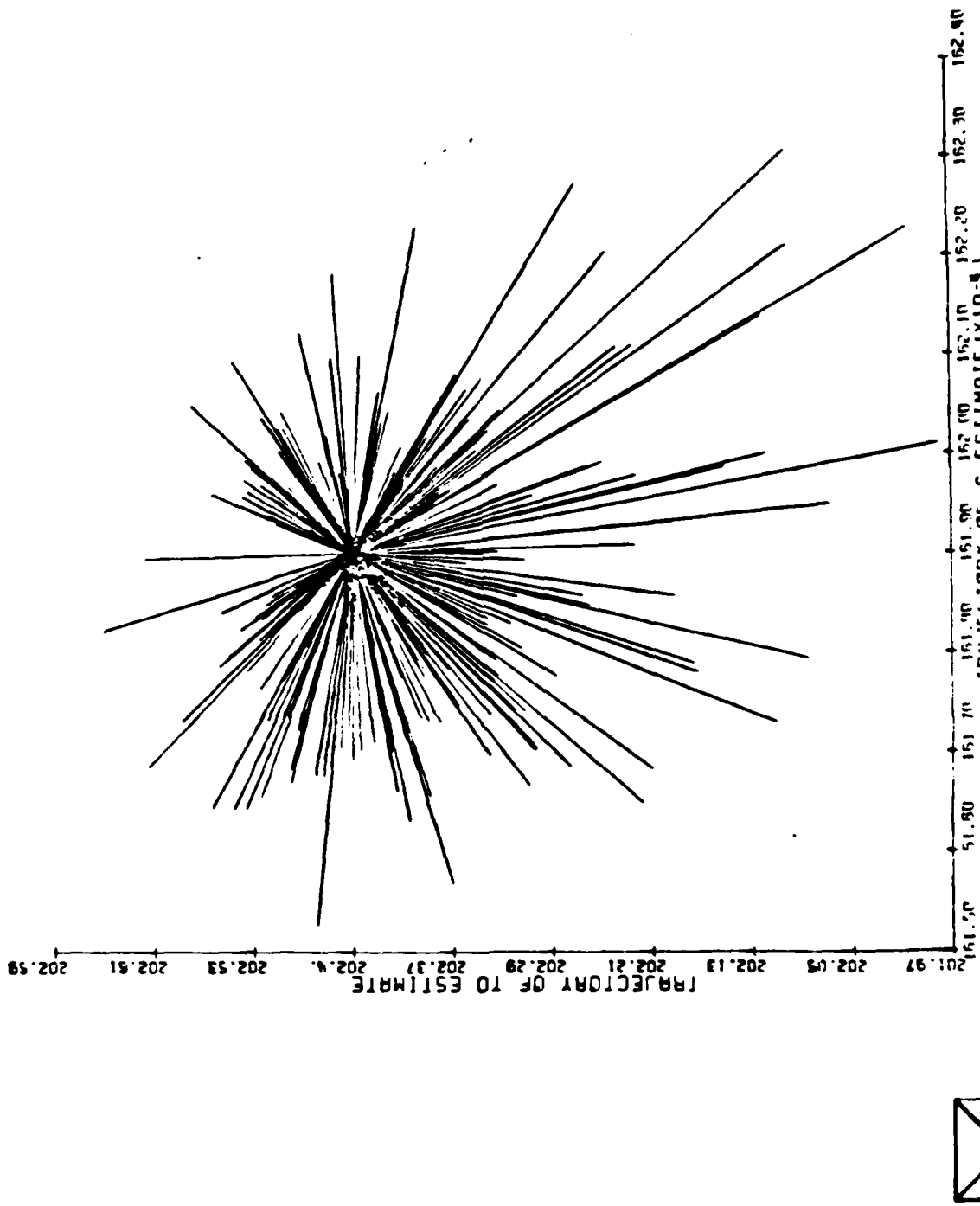

RSKF ITERATION NUMBER 199
TRAJECTORY OF TO ESTIMATE
202.59
202.61
202.53
202.45
202.37
202.29
202.21
202.13
202.05
201.97
162.40
162.30
162.20

Figure 27

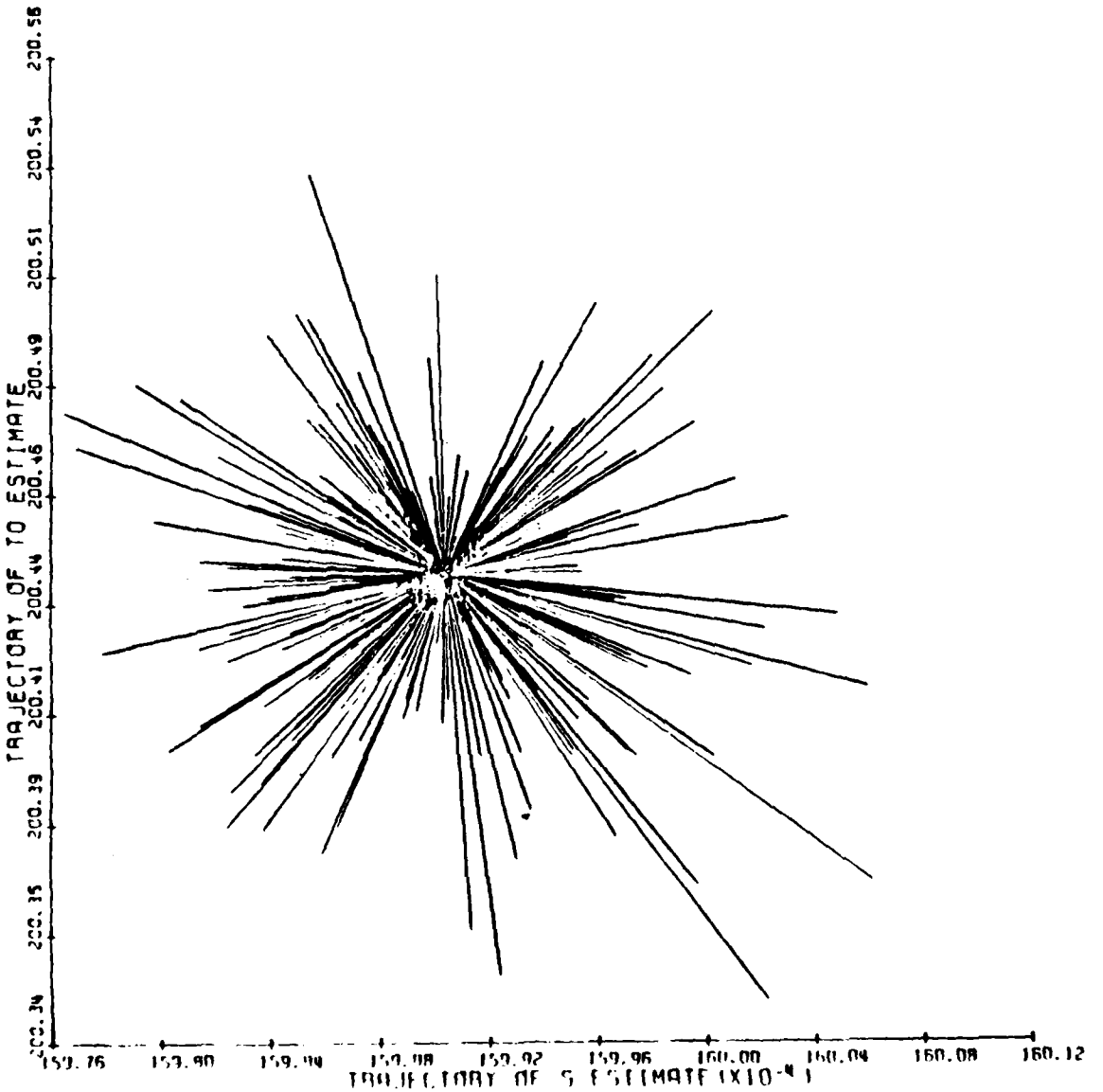

TRAJECTORY OF TO ESTIMATE
200.58
200.54
200.51
200.49
200.46
200.44
200.41
200.39
200.35
200.34
159.76
159.80
159.84
159.88
159.92
159.96
160.00
160.04
160.08
160.12
TRAJECTORY OF S ESTIMATE (X10^-4)
RSKF ITERATION NUMBER 309

Figure 28

400 measurements were processed during the iteration.

Each plot shows 400 possible correction vectors. However, this is merely for illustrative purposes. In actual operation the RSKF samples the measurements without replacement. This means that the size of the sample population decreases by one after each iteration.

The most important feature in these plots is that the possible correction vectors generally do not point in the same direction.

Also, as the processing continues, the fluctuation in the magnitudes of adjacent correction vectors tends to increase. This is due to the increased effect of the measurement noise on the correction vector magnitude (and direction if there is a sign change).

4.3.2 Suboptimal Weights

The next several plots will illustrate the effect of the use of different weights in the Kalman gain denominator.

Figs. 29 and 30 are plots of the normalized covariance trace when a_k=4.0 and a_k=8.0, respectively. These may be compared to Fig. 17 where a_k=2.0. The slower convergence that results from increasing the weight beyond 2:1 is evident in the longer time it takes for the trace to converge to zero.

The filter's performance with a_k=1.0 is illustrated in Figs. 31, 32 and 33. Fig. 31 shows that a reasonable fit has not been achieved. In Fig. 32 it can be seen that the trajectory of the nonlinear parameter estimates has failed to correctly converge. As previously mentioned, the problem is that of nonlinear filter divergence. The filter takes an overly optimistic view of the accuracy its estimates. The very rapid decrease in the normalized covariance trace of Fig. 33 in comparison to that of Fig. 17 confirms this.

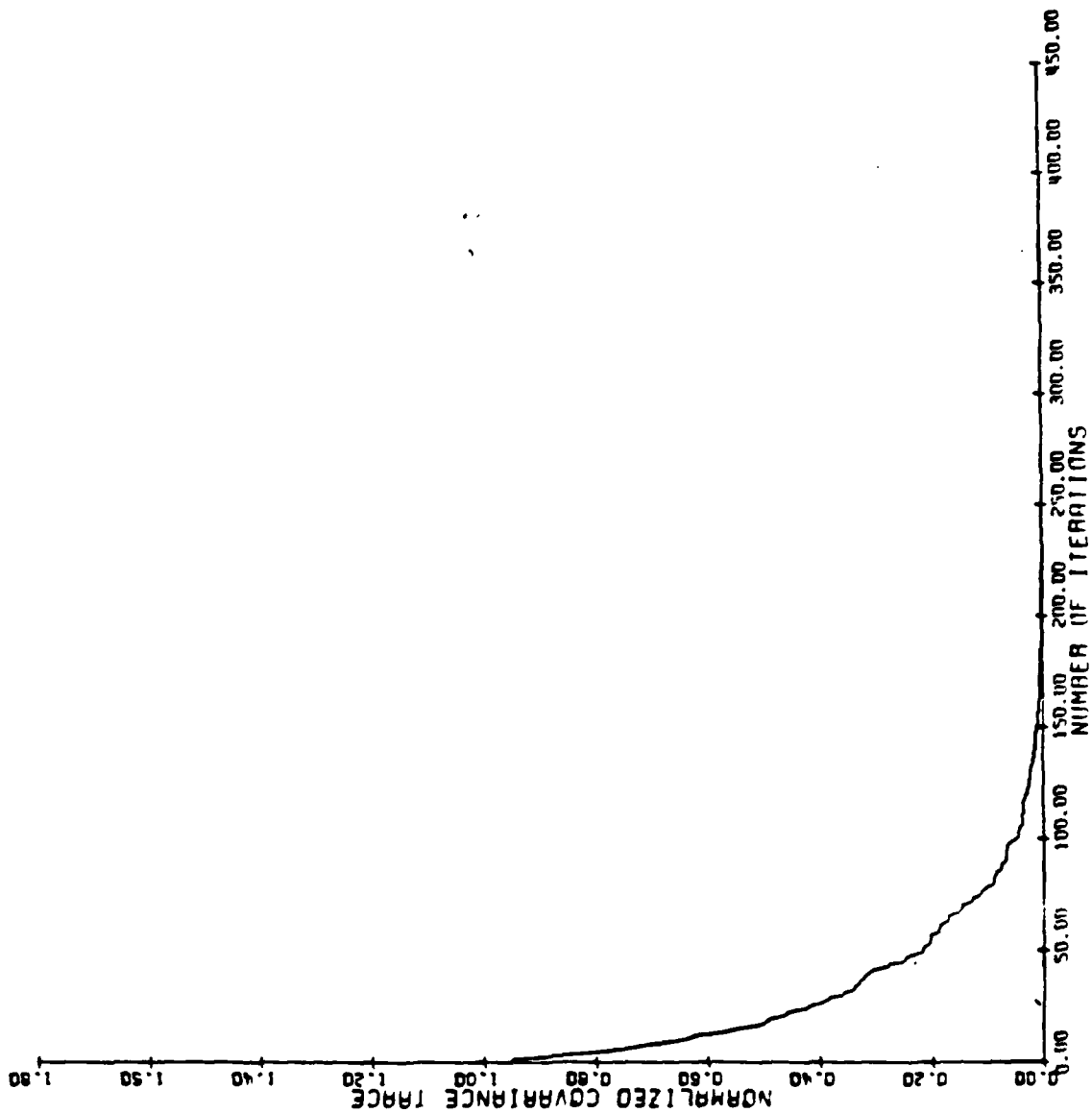

NORMALIZED COVARIANCE TRACE
0.00
0.20
0.40
0.60
0.80
1.00
1.20
1.40
1.60
1.80
0.00
50.00
100.00
150.00
200.00
250.00
300.00
350.00
400.00
450.00
NUMBER OF ITERATIONS

Figure 29

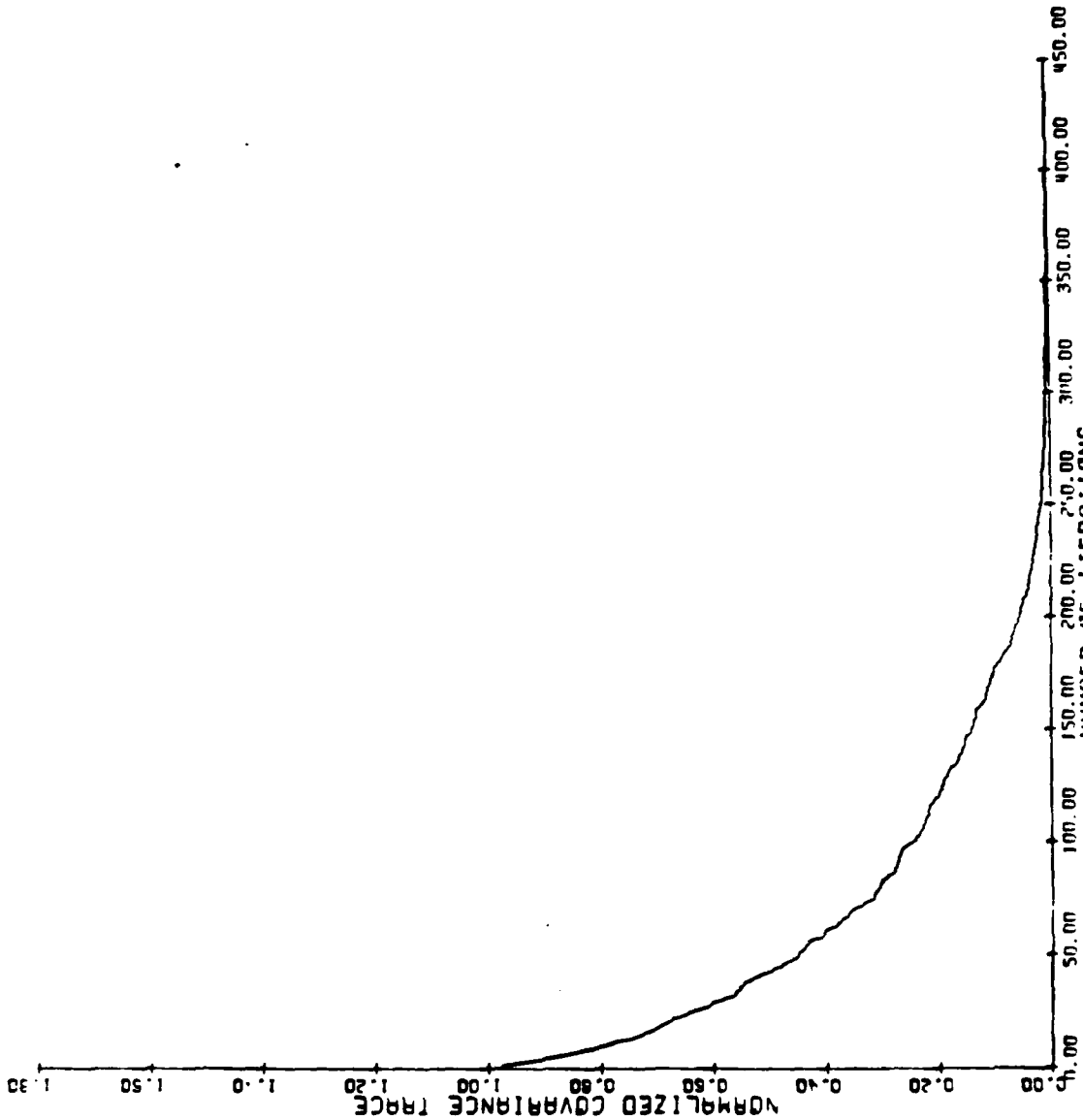

NORMALIZED COVARIANCE TRACE
0.00
0.20
0.40
0.60
0.80
1.00
1.20
1.40
1.50
1.30
0.00
50.00
100.00
150.00
200.00
250.00
300.00
350.00
400.00
450.00
NUMBER OF ITERATIONS

Figure 30

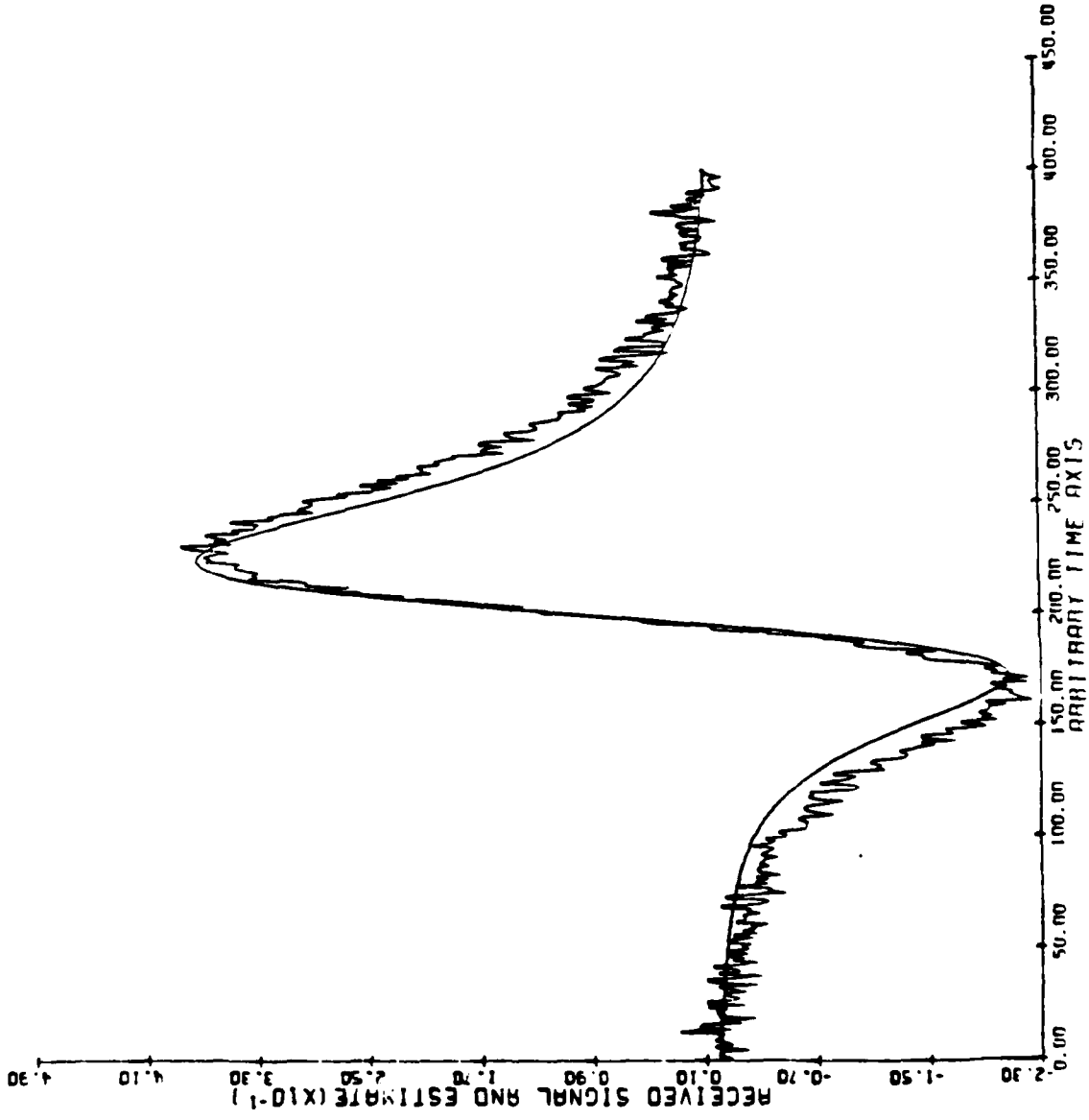

RECEIVED SIGNAL AND ESTIMATE (X10-1)
-2.30
-1.50
-0.70
0.10
0.90
1.70
2.50
3.30
4.10
0.00
50.00
100.00
150.00
200.00
250.00
300.00
350.00
400.00
450.00
ARRAY TIME AXIS

Figure 31

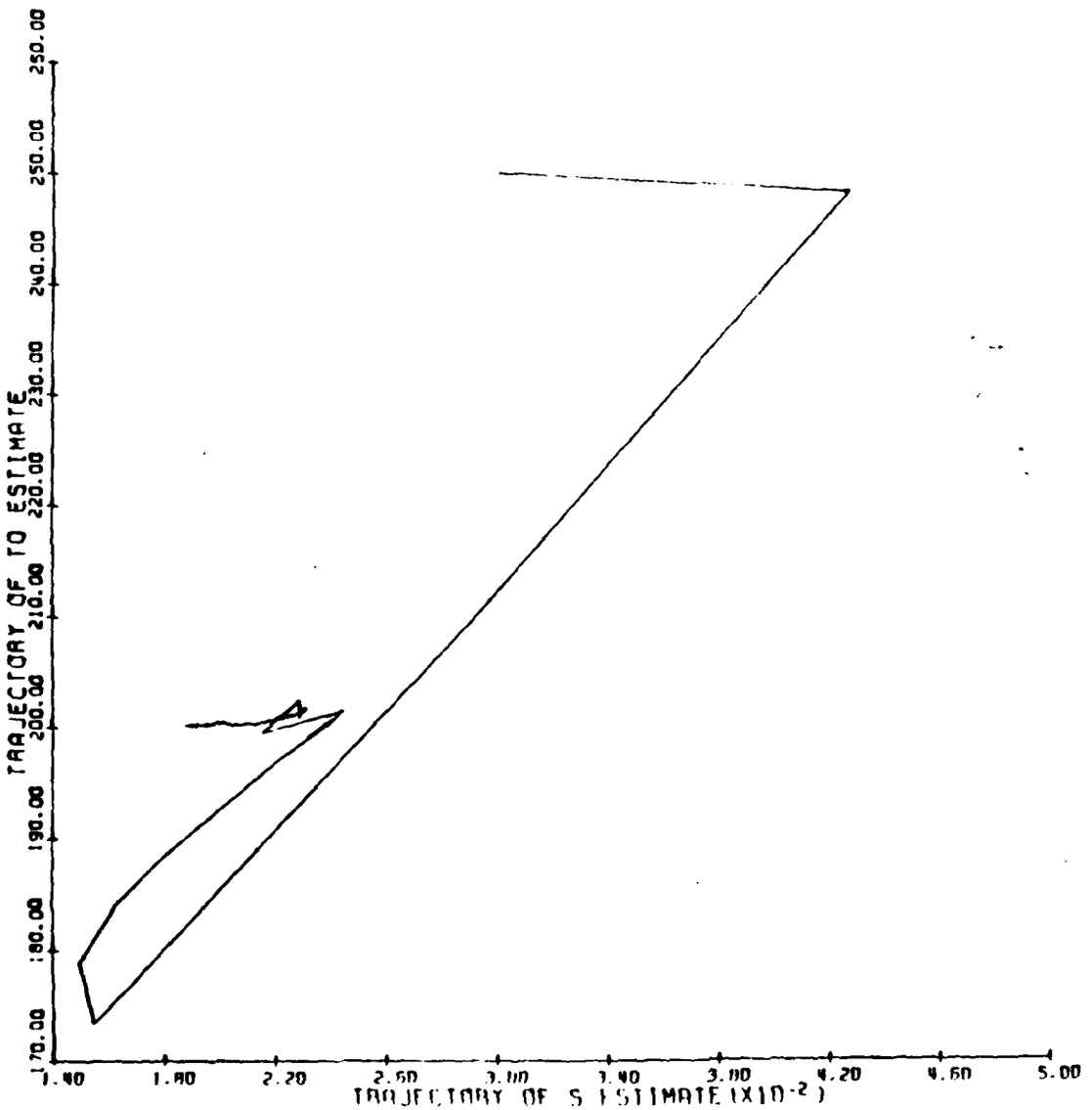

TRAJECTORY OF TO ESTIMATE
250.00
240.00
230.00
220.00
210.00
200.00
190.00
180.00
170.00
1.40
1.80
2.20
2.60
3.00
3.40
3.80
4.20
4.60
5.00
TRAJECTORY OF S ESTIMATE (X10-2)

Figure 32

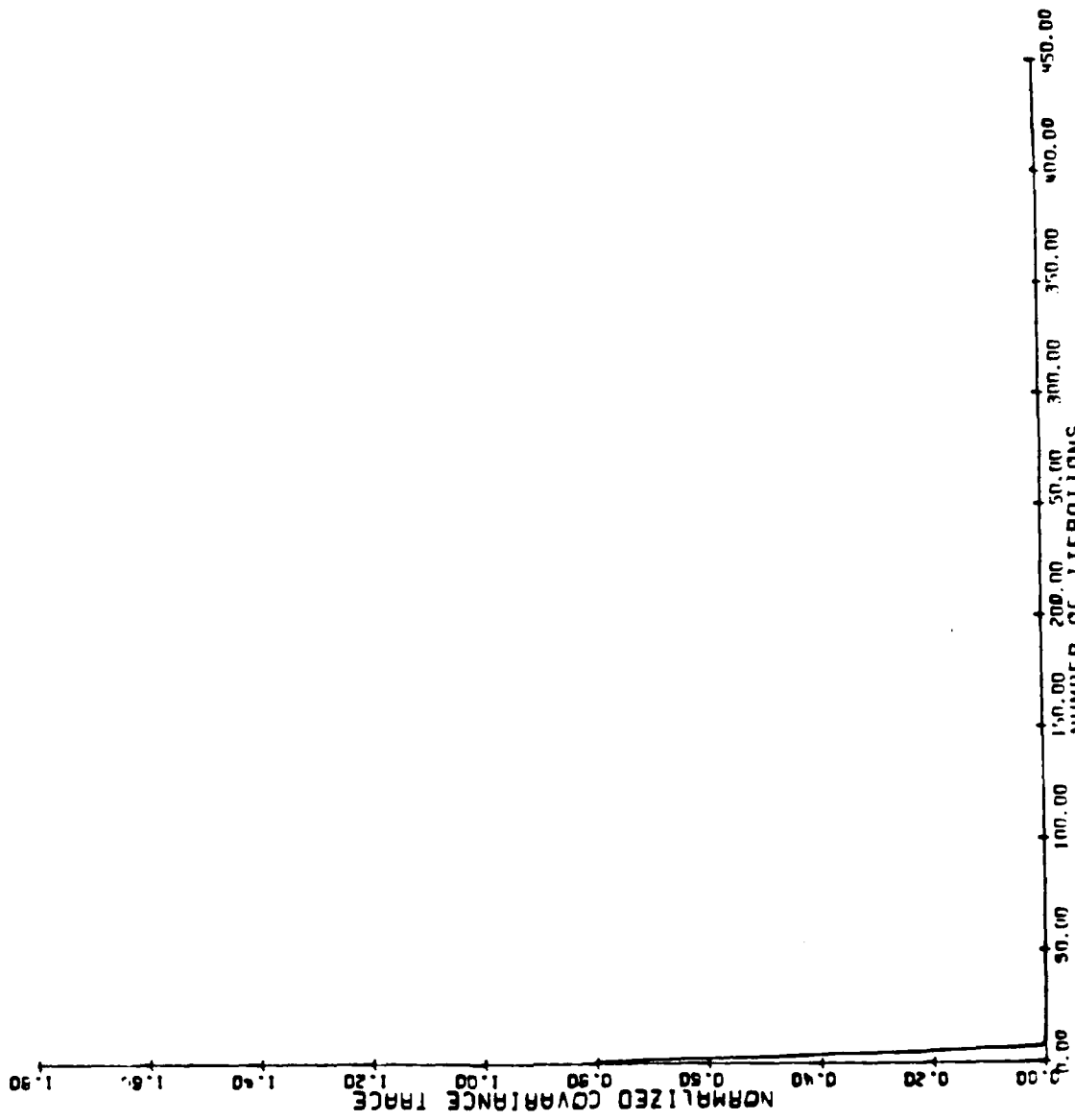

NORMALIZED COVARIANCE TRACE
0.00
0.20
0.40
0.60
0.80
1.00
1.20
1.40
1.60
1.80
50.00
100.00
150.00
200.00
250.00
300.00
350.00
400.00
450.00
NUMBER OF ITERATIONS

Figure 33

Chapter 5

Conclusion

An application of off-line processing by means of a sequential estimator has been presented. The technique is essentially a sequential version of Gauss-Newton optimization with ficticious measurement noise added to prevent filter divergence.

The RSKF described does not implicitly make use of the form of the model (1), its separability and most of the associated assumptions. Therefore, it would appear to be possible to use if for other non-linear square error regression problems and, in particular, for curve fitting. Such a method would seem to be most appropriate in situations requiring fast off-line processing.

However, computational savings if they are possible at all, will depend on the complexity of the model being employed. Also, the extent to which convergence can be guaranteed has not been sufficiently delineated. Furthermore, the numerical properties of the filter equations are crucial to the feasibility at their implementation in a limited precision computer.

Appendix A

At first glance it may appear that the well known optimization maxim, that one's estimates should proceed in a direction of non-increasing error, is being violated. The confusion apparently arises because a distinction must be made between the error surface based on all the data that will eventually be received and the error surface based on only the data received up to the point of the current iteration. There is no reason to think that ones' estimate should always be moving toward an optimal point on an error surface that depends on data not yet received.

To illustrate this, consider the sequential estimation of the stationary mean of a sequence of N i.i.d. Gaussian random variables, y_k, k=1,2...N.

$$\hat{x}_{k+1} = \frac{k-1}{k}\hat{x}_k + \frac{1}{k}y_k \qquad (1)$$

$$\hat{x}_{k+1} = \hat{x}_k + \frac{1}{k}[y_k - \hat{x}_k] \qquad (2)$$

At each iteration $\hat{x}_{k+1}$ is clearly the optimal (say square error) estimate on the error surface of data received so far. However, as $\hat{x}_k$ converges toward the actual mean, (2) is adding what is almost a zero mean (actually $\mu - \frac{1}{k}\sum_{j=1}^{k} y_j$ / k) Gaussian random variable to the current estimate in order to produce the updated estimate.

Therefore, with respect to the error surface of all the data to eventually be received, the estimate $\hat{x}_k$ moves aways from the optimum point of that surface a percentage of times that asymptotically approaches 50%. Put another way, the current estimate moves

away from the final estimate almost as many times as it moves towards it.

Appendix B

The rank condition can be violated in models such as

$$y(t_k) = c_1 e^{B_1 t_k} + c_2 e^{B_2 t_k} + \ldots + c_m e^{B_m t_k}$$

The problem is that columns in the matrix of basis functions may become linearly dependent, reducing the matrix rank. In the above model this will occur if $\hat{B}_i = \hat{B}_j$:

$$\Phi(B_1, \ldots, B_m) = \begin{bmatrix} e^{B_1 t_1} & e^{B_2 t_1} & e^{B_m t_1} \\ e^{B_1 t_2} & e^{B_2 t_2} & e^{B_m t_2} \\ & & \\ e^{B_1 t_n} & e^{B_2 t_n} & e^{B_m t_n} \end{bmatrix}$$

In the Anderson model the basis functions are linearly independent for any $\hat{S}$ and $\hat{T}_0$, and so, analytically, present no problem.

Under certain circumstances though, columns could 'numerically' resemble each other. For instance if the location parameter estimate and/or scale estimate are quite far off the columns in the matrix of basis functions will assume values close to zero.

Appendix C

The number of computational units (multiplications) necessary to evaluate the model (23), and associated partial derivatives, depends on how much extra programming complexity one is willing to accept.

To establish a lower bound, one should note that the function evaluation (and thus linear parameter partial derivatives) requires at least six multiplications. The evaluation of the nonlinear parameter partial derivatives each require about 18 multiplications.

In addition, a quantity must be raised to the 2.5 power. This is equivalent to one multiplication, one logarithm and one exponentiation. Nonlinear function evaluations are approximately equivalent to eight to twenty multiplications, depending on the accuracy desired [9],[10].

Assuming that ten multiplications per nonlinear evaluation is a good estimate, the time required to evaluate the function in question and its associated partial derivatives is equivalent to performing 64 or more multiplications.

ACKNOWLEDGEMENT

The authors would like to acknowledge a number of interesting and useful conversations with Dr. William Schmidt of the Naval Air Development Center, Warminster, Pa.

REFERENCES

[1] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," SIAM J. Numer. Anal., Vol. 10, No. 2, April 1973 pp. 413-432.

[2] R. R. Tenney, R. S. Hebbert and N. R. Sandell Jr., "A tracking filter for maneuvering sources," IEEE Trans. Automat. Contr., April 1977, pp. 246-251.

[3] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters", J. SIAM, Vol. 11, No. 2, June 1963, pp. 431-441.

[4] B.D.O. Anderson and J.B. Moore, "Optimal Filtering," Prentice-Hall, 1979, pp. 138-152.

[5] Y. Sunahara, "On a method of signal detection by using the nonlinear estimation theory," Proc. of the Third Symp. on Nonlinear Estimation Theory and Applications, San Diego, California, 1972, pp. 216-223.

[6] A.D. McAulay, "Computerized Model Demonstrating Magnetic Submarine Localization," IEEE Trans. on Aerospace and Elec. Sys., Vol AES-13, No. 3, May 1977, pp. 246-254.

[7] J. M. Mendel, "Computational requirements for a discrete Kalman filter," IEEE Trans. Automat. Contr., Vol. AC-16, No. 6 , Dec. '71, pp. 748-758.

[8] B.T. Fang, "A Nonlinear Counterexample for Batch and Extended Sequential Estimation Algorithms," IEEE Trans. Automat. Contr., Vol AC-21, No. 1, Feb. 1976, pp. 138-139.

[9] C. Hastings, "Approximations for Digital Computers," Princeton University Press, Princeton, N.J., 1955.

[10] M. Necula, "An Algorithm for Fast Evaluation of Time Functions," IEEE Trans. Comp., Vol. C-28, No. 7, July 1979, pp. 506-513.

[11] N. R. Draper & H. Smith, "Applied Regression Analysis," John Wiley & Sons, New York, 1967

[12] L. Ljung, "Asymptotic Behavior of the Extended Kalman Filter as a Parameter Estimator for Linear Systems," IEEE Trans. AC-24, No. 1, Feb. 1979, pp. 36-50.

[13] T.C. Hsia, "System Identification: Least Squares Methods," D.C. Heath and Company, Lexington, Mass., 1977, pp. 22-25.

[14] J.M. Mendel, "Discrete Techniques of Parameter Estimation: The Equation Error Formulation," Marrel Dekker, N.Y., 1973, pp. 91-107.

[15] "Applied Optimal Estimation", A. Gelb, ed., M.I.T. Press, Cambridge, Mass., 1974, pp. 188-189.

[16] S.F. Schmidt , "Application of State-Space Methods to Navigation Problems," in "Advances in Control," C. T. Leondes , ed., Vol. 3, 1966, Academic Press.

OFFICE OF NAVAL RESEARCH
STATISTICS AND PROBABILITY PROGRAM

BASIC DISTRIBUTION LIST
FOR
UNCLASSIFIED TECHNICAL REPORTS

JANUARY 1980

	Copies
Statistics and Probability Program (Code 436) Office of Naval Research Arlington, VA 22217	3
Defense Technical Information Center Cameron Station Alexandria, VA 22314	12
Office of Naval Research New York Area Office 715 Broadway - 5th Floor New York, New York 10003	1
Commanding Officer Office of Naval Research Branch Office Attn: Director for Science 666 Summer Street Boston, MA 02210	1
Commanding Officer Office of Naval Research Branch Office Attn: Director for Science 536 South Clark Street Chicago, Illinois 60605	1
Commanding Officer Office of Naval Research Branch Office Attn: Dr. Richard Lau 1030 East Green Street Pasadena, CA 91101	1
Office of Naval Research San Francisco Area Office One Hallidie Plaza - Suite 601 San Francisco, CA 94102	1
Office of Naval Research Scientific Liaison Group Attn: Scientific Director American Embassy - Tokyo APO San Francisco 96503	1
Applied Mathematics Laboratory David Taylor Naval Ship Research and Development Center Attn: Mr. G.H. Gleissner Bethesda, Maryland 20084	1
Commandant of the Marine Corps (Code AX) Attn: Dr. A.L. Slafkosky Scientific Advisor Washington, DC 20380	1
Director National Security Agency Attn: Mr. Stahly and Dr. Maar (R51) Fort Meade, MD 20755	2
Navy Library National Space Technology Laboratory Attn: Navy Librarian Bay St. Louis, MS 39522	1

	Copies
U.S. Army Research Office P.O. Box 12211 Attn: Dr. J. Chandra Research Triangle Park, NC 27706	1
OASD (I&L), Pentagon Attn: Mr. Charles S. Smith Washington, DC 20301	1
ARI Field Unit-USAREUR Attn: Library c/o ODCSPER HQ USAEREUR & 7th Army APO New York 09403	1
Naval Underwater Systems Center Attn: Dr. Derrill J. Bordelon Code 21 Newport, Rhode Island 02840	1
Library, Code 1424 Naval Postgraduate School Monterey, California 93940	1
Technical Information Division Naval Research Laboratory Washington, DC 20375	1
Dr. Barbara Bailar Associate Director, Statistical Standards Bureau of Census Washington, DC 20233	1
Director AMSAA Attn: DRXSY-MP, H. Cohen Aberdeen Proving Ground, MD 21005	1
Dr. Gerhard Heiche Naval Air Systems Command (NAIR 03) Jefferson Plaza No. 1 Arlington, Virginia 20360	1

	Copies
B. E. Clark RR #2, Box 647-B Graham, North Carolina 27253	1
ATAA-SL, Library U.S. Army TRADOC Systems Analysis Activity Department of the Army White Sands Missile Range, NM 88002	1

OFFICE OF NAVAL RESEARCH
STATISTICS AND PROBABILITY PROGRAM

SIGNAL PROCESSING DISTRIBUTION LIST
FOR
UNCLASSIFIED TECHNICAL REPORTS

JANUARY 1980

	Copies
Library Naval Ocean Systems Center San Diego, CA 92152	1
Professor G. S. Watson Department of Statistics Princeton University Princeton, NJ 08540	1
Professor T. W. Anderson Department of Statistics Stanford University Stanford, CA 94305	1
Professor M. R. Leadbetter Department of Statistics University of North Carolina Chapel Hill, NC 27514	1
Professor M. Rosenblatt Department of Mathematics University of California, San Diego La Jolla, CA 92093	1
Professor E. Parzen Department of Statistics Texas A&M University College Station, Texas 77840	1
Professor H. L. Gray Department of Statistics Southern Methodist University Dallas, Texas 75275	1
Professor P.A.W. Lewis Department of Operations Research Naval Postgraduate School Monterey, CA 93940	1
Professor E. Masry Department of Applied Physics and Information Science University of California La Jolla, CA 92093	1
Professor I. Rubin School of Engineering and Applied Science University of California Los Angeles, CA 90024	1
Professor L. L. Scharf, Jr. Department of Electrical Engineering Colorado State University Fort Collins, CO 80521	1
Professor M. J. Hinich Department of Economics Virginia Polytechnic Institute and State University Blacksburg, VA 24061	1
Naval Coastal Systems Center Code 741 Attn: Mr. C.M. Bennett Panama City, FL 32401	1

	Copies
J. S. Lee Associates, Inc. 2001 Jefferson Davis Highway Suite 802 Arlington, VA 22202	1
Naval Electronic Systems Command (NELEX 320) National Center No. 1 Arlington, Virginia 20360	1
Professor D. P. Gaver Department of Operations Research Naval Postgraduate School Monterey, CA 93940	1
Professor Bernard Widrow Stanford Electronics Laboratories Stanford University Stanford, CA 94305	1
Dr. M. J. Fischer Defense Communications Agency Defense Communications Engineering Center 1860 Wiehle Avenue Reston, VA 22090	1
Professor L. A. Aroian Institute of Administration and Management Union College Schenectady, New York 12308	1
Professor Grace Wahba Department of Statistics University of Wisconsin Madison, Wisconsin 53706	1
Professor S. C. Schwartz Department of Electrical Engineering and Computer Science Princeton University Princeton, NJ 08540	1
Professor Charles R. Baker Department of Statistics University of North Carolina Chapel Hill, NC 27514	1

	Copies
Mr. David Siegel Code 210T Office of Naval Research Arlington, VA 22217	1
Professor Balram S. Rajput Department of Mathematics University of Tennessee Knoxville, Tennessee 37916	1
Professor Harold A. Titus Department of Electrical Engineering Naval Postgraduate School Monterey, CA 93940	1
Mr. Dan Leonard Code 8105 Naval Ocean Systems Center San Diego, CA 92152	1
Professor R. de Figueiredo Department of Electrical Engineering Rice University Houston, Texas 77001	1
Professor C. H. Chen Department of Electrical Engineering Southeastern Massachusetts University North Dartmouth, MA 02747	1
Professor K. S. Fu School of Electrical Engineering Purdue University West Lafayette, Indiana 47907	1
Professor Melvin F. Janowtiz Dept. of Mathematics and Statistics University of Massachusetts Amherst, MA 01002	1
Professor M. M. Rao Department of Mathematics University of California Riverside, CA 92521	1

	Copies
Professor James A. Cadzow Department of Electrical Engineering Virginia Polytechnic Institute and State University Blacksburg, VA 24061	1
Professor Stanley L. Sclove Department of Mathematics University of Illinois at Chicago Circle Box 4348 Chicago, Illinois 60680	1