Order Number 8801708

Simultaneous state and parameter estimation in linear systems

Eliazov, Teymuraz, Ph.D.

City University of New York, 1987

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __√__.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print ______
3. Photographs with dark background _____
4. Illustrations are poor copy ______
5. Pages with black marks, not original copy __√__
6. Print shows through as there is text on both sides of page ______
7. Indistinct, broken or small print on several pages __√__
8. Print exceeds margin requirements ______
9. Tightly bound copy with print lost in spine _______
10. Computer printout pages with indistinct print ______
11. Page(s) __________ lacking when material received, and not available from school or author.
12. Page(s) __________ seem to be missing in numbering only as text follows.
13. Two pages numbered ______. Text follows.
14. Curling and wrinkled pages ______
15. Dissertation contains pages with print at a slant, filmed as received ________
16. Other ____________________

SIMULTANEOUS STATE AND PARAMETER ESTIMATION IN LINEAR SYSTEMS

by

TEYMURAZ ELIAZOV

A dissertation submitted to the Graduate Faculty in Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1987

This manuscript has been read and accepted for the Graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

8/3/87
Date

Frederick E. Than
Chair of Examining Committee

8/4/87
Date

Executive Officer

J. M. Kranc

L. Roytman

Supervisory Committee

The City University of New York

Abstract

SIMULTANEOUS STATE AND PARAMETER ESTIMATION
IN LINEAR SYSTEMS

by

Teymuraz Eliazov

Adviser: Professor Frederick Thau

In this work we consider the simultaneous state and parameter estimation problem of a linear discrete-time system subject to an arbitrary known input, and random input (driving noise), and with noisy output observations. A simultaneous state and parameter estimation problem is formulated as a least squares minimization problem. The special structure of the least squares minimization problem, allows separation of the original problem into two subproblems that are explicitly coupled: a nonlinear functional minimization to obtain parameter estimates and a linear least-squares problem to obtain state estimates. A new, computationally economical and robust algorithms are derived based on the proposed method. A theoretical and experimental study of the asymptotic properties of the parameter and state estimates are presented.

An adaptive control system for fine pointing of a flexible spacecraft is designed. Derived algorithms are used to simultaneosly identify unknown states and parameters of a discrete-time lumped-parameter model of the flexible structure. The identified states and parameters form the input to a bang-off-bang control law that, in conjuction with the identification algorithm, results in an adaptive system whose response closely approximates that of a system with known parameters. Simulation studies demonstrate the response achievable with the proposed approach.

ACKNOWLEDGMENTS

I am most grateful to Professor Frederick E. Thau, my thesis advisor, for his guidance, encouragement and support throughout the whole doctoral process. In particular my true understanding of adaptive estimation problem must be attributed to his inspiration and intellectual stimulation.

I would like to thank my wife for her constant support and encouragement during the preparation of this dissertation.

NOTATIONAL CONVENTIONS

A,B,C	matrices for state-space model
d	dimension of the parameter vector
$\bar{E}$	$\bar{E}f(k) = \lim_{k\to\infty} \frac{1}{k} \sum_{l=1}^{k} Ef(l)$ where E = expectation operator
k	time variable (interger valued)
m	dimension of the output vector y(k)
n	model order
R_1, R_2, R_{12}	covariance matrices
r	dimension of input vector u(k)
T	transpose
x(k)	state vector (of dimension n)
$\hat{x}_k(k)$	estimate of state vector x(k) based on data up to time k
y(k)	output vector (of dimension m)
θ	parameter vector (of dimension d)
θ_0	true value of the parameter vector θ
$\hat{\theta}_k$	recursive estimate of θ based on data up to time k
λ	forgetting factor
$\psi(k)$	gradient of predictions for fixed model parameter
$\lVert \cdot \rVert$	quadratic norm
A^+	Moore-Penrose pseudoinverse of matrix A
u(k)	input vector (of dimension r)
mw	size of the moving window

LIST OF ILLUSTRATIONS

CONTENTS

Simultaneous State and Parameter Estimation in Linear Systems

by

Teymuraz Eliazov

1. Introduction

In recent years the problem of simultaneously estimating the state vector and parameters of a dynamic process has received a great deal of attention [6,12,18,22,38,39,45,46]. The adaptive scheme (Figure 2) using the identifier and controller has wide application in design of adaptive control based on linear state space model. The applications reported most recently include process control [47], industrial robots [39] and satellite attitude control [48,53]. In practice most of the existing on-line algorithms for simultaneous state and parameter estimation suffer computational complexity and robustness problems or are limited to certain canonical state-space models.

Thus, a need for robust and computationally economical algorithms for a general state-space model clearly exists.

The combined parameter and state estimation problem was originally posed as a nonlinear estimation problem [15] by augmenting the state vector with the unknown parameters of the state space equations. Kopp and Orford [15], and Farison et al. [23] proposed the extended Kalman filter (EKF) to solve the resulting nonlinear filtering problem; however, the extended Kalman filter, while being computationally feasible, is prone to divergence. The convergence analysis given in [13] shows that if the noise characteristics of the model (2.1) are chosen ad hoc (as, apparently, is usually done) then the system parameter estimates will in general be biased. The reason for divergence is traced to the fact that the effect on the Kalman gain K of a change in a parameter vector θ is not taken care of in the usual EKF implementation. This lack of coupling between K and θ in the algorithm may, as demonstrated in [13], lead to divergence of the estimates, even for simple cases, where only the system parameters are unknown.

In practical applications of the EKF, "manual" adjustments of the noise covariances are often used to make the algorithm work. This is called "tuning of the filter". Several modifications to the algorithm have been suggested to improve convergence properties. Schmidt [49], Neal [50] and Ljung [13,45] proposed the addition of terms to the gain K to prevent divergence, but the approach follows from a heuristic discussion. One disadvantage with such a modification of the EKF is that it will require an amount of computation that may be forbidding for higher order systems.

Thus the EKF method either requires extensive "tuning" or adds to the computational complexity, either of which is a disadvantage in practical on-line applications.

The simultaneous state and parameter estimation problem for linear systems is inherently a nonlinear filtering problem. A straightforward use of the best known non-linear filters often leads to computational problems and to poor estimation. In [19,39] a nonlinear filter is introduced that suffers less from these problems. It is based on the Bayesian maximum a posteriori estimation technique.

A relationship with EKF-type filters is shown. It is further shown that the filter can be applied to simultaneous state and parameter estimation problems. The filter has been used by the authors in a real application to a heavy hydraulic manipulator which is digitally controlled by a two-level microcomputer system. However, the method assumes that distrubances are zero mean Gaussian stochastic processes with known covariance matrices. In addition, method is limited to systems with known nominal values for parameters, states and inputs. The convergence rate of parameter estimates is slow and depends on the choice of nominal values.

Bar-Shalom [7] obtained the optimal maximum likelihood solution to the problem; however, he only obtained the result for the single-input single-output case. Furthermore, his method requires calculating and storing all of the smoothed state estimates at each step. In [18] an algorithm is derived as a recursive solution to a maximum likelihood estimation problem. The algorithm has a partitioned structure into three distinct components: A Kalman filter state estimator, a recursive least squares parameter identifier, and a sensitivity matrix estimator. However, this algorithm assumes the system equations are jointly linearizable in the state and parameter vectors about known reference values at each stage, and may not be appropriate for identification of system dynamics with poor apriori initial estimates.

A frequently used and very reasonable sub-optimal estimation scheme for simultaneous estimation of parameter vector θ and state vector $x(k)$ is to implement two coupled estimation algorithms as now described. In estimating θ, x(k) is replaced by an estimate of x(k) from the state estimator, and in estimating x(k), θ is replaced by an estimate of θ from the parameter estimator. Many schemes for adaptive estimation in the engineering literature including "equation-error" or "series-parallel" schemes and "output-error" or "parallel" schemes have the general structure of the above sub-optimal arrangement. The suboptimal estimators as described above have been limited to a certain classes of state space models. In [6] a narrow a class of models is introduced for which suboptimal estimators can be implemented. The associated adaptive estimators specialize to a known and novel adaptive algorithms. The convergence analysis is given via martingale convergence theorems. The passivity condition and persistently exciting conditions on the noise and state estimates are shown to guarantee almost sure convergence results. In [16] a suboptimal parameter and state estimator is presented for a particular canonical form for the state-space equations of a linear system which allows parameters and states to be estimated separately using two linear estimators. The system is assumed to be completely controllable and observable. It is also assumed that the random processes are Gaussian and statistics are known. The Luenberger canonical form has been utilized in [20] to separate the non-linear state and parameter estimation problem into two linear problems. However the approach is limited to only deterministic systems and to systems with equal state and measurement disturbances. Simulation experiments demonstrated that the state estimator is prone to divergence.

The simultaneous state and parameter estimation has often been approached through adaptive observer techniques [47,51,52]. In these approaches, the state variables are linear combinations of delayed inputs and outputs, with the unknown system parameters as coefficients. This however introduces the problem of choosing a gain matrix to stabilize a non-linear estimation loop, due to products of estimated variables in the real time system model for observer. In [46] an approach is presented that solves this problem, however it is derived for deterministic systems and may have singularity problems.

A recursive parameter estimation algorithm for general state-space models is derived in [45] as special case of a prediction error algorithm. For the algorithm developed in [45] the adaptive state estimate is obtained as an automatic by-product of a recursive identification algorithm. The algorithm is robust and convergence to a local minimum of a performance criterion which is a quadratic function of the prediction error has been established. However, computational complexity considerations make the algorithm forbidding for higher order systems. It is also assumed that the statistics of the random processes are known.

In this work we have derived a simultaneous state and parameter estimation method for general state-space models of multivariable discrete-time linear systems.

A simultaneous state and parameter estimation problem is formulated as a least squares minimization problem. The special structure of the least squares minimization problem, allows separation of the original problem into two subproblems that are explicitly coupled: a nonlinear functional minimization to obtain parameter estimates and a linear least-squares problem to obtain state estimates. A robust and efficient adaptive scheme that follows from such separation is shown in Figure 1. It is obtained for a general state-space model, while in all previous reports [6,11,16,45] such adaptive schemes were derived only for certain canonical forms of the state-space model. Another advantage of the least-squares approach is that the derived algorithm does not require knowledge of the statistics of the random disturbance processes. However, if available, this information can be easily incorporated.

We have obtained specific algorithms for on-line parameter and state estimation for general state space models of linear discrete-time multivariable systems subject to an arbitrary known input, a random input (driving noise), and with noisy measurements. A theoretical and experimental study of the asymptotic properties of the parameter and state estimates appears in Section 7.

The contribution of this research rests in the development of the least squares method for simultaneous state and parameters estimation of state space model. The method is developed for a general state space model and does not require knowledge of the disturbance characteristics, the necessary condition in all the previous works. A new, computationally economical and robust alogrithm is derived based on the proposed method. The derived algorithm has been applied to an adaptive control system for fine pointing of flexible spacecraft.

In section 2, we consider the simultaneous state and parameter estimation problem for a discrete-time dynamic system described by a general state-space model.

In section 3, we present a least squares formulation of the simultaneous state and parameter estimation problem. Then the original problem is separated into two explicitly coupled subproblems by using a variable projection algorithm, and the structure of the adaptive estimator is then defined.

In section 4, we derive algorithms for on-line simultaneous state and parameter estimation of a general state space model. Two different approaches have been applied to obtain parameter estimation algorithms.

In section 5, we consider the convergence properties of these parameter estimation algorithms. Local convergence for an iterative continuation algorithm is established. We also demonstrate that the general convergence results for prediction error algorithms [45] apply to the recursive parameter estimation algorithm derived in section 4.

In section 6, we give a description of computational details for the derived algorithms. The implementation of the variable projection method which makes extensive use of orthogonal transformations is presented.

Simulation studies to demonstrate properties of the derived algorithms and to compare their performance with other algorithms is given in section 7.

In section 8 an adaptive control system for fine pointing of a flexible spacecraft is designed. The identification algorithm derived in section is used to simultaneously identify unknown states and parameters of a discrete-time lumped-parameter model of the flexible structure. The identified states and parameters form the input to a bang-off-bang control law that, in conjunction with the identification algorithm, results in an adaptive control system. Simulation studies demonstrate the adaptive system response achievable with the proposed approach.

In sections 9 and 10 we give conclusions and recommendations for further research.

2. Problem Statement

Consider linear discrete-time dynamic systems modeled by

$$x(k+1) = A_k(\theta)x(k) + B_k(\theta)u(k) + \nu(k) \tag{2.1}$$

$$y(k) = C_k(\theta)x(k) + \omega(k) \tag{2.2}$$

where $x(k)$ is an $n \times 1$ state vector, $u(k)$ is an $r \times 1$ deterministic input vector, $y(k)$ is an $m \times 1$ measurement vector, θ is a $p \times 1$ parameter vector, $\nu(k)$ is an $n \times 1$ state disturbance vector, $\omega(k)$ is an $m \times 1$ measurement disturbance vector ($k=1, 2.....$).

The $A_k(\theta)$, $B_k(\theta)$ and $C_k(\theta)$ are matrices of appropriate dimension and depend on a parameter vector θ in an arbitrary way. It is assumed, though, that the matrix elements are continuously differentiable functions of parameter vector θ.

The problem is to determine the parameter vector estimate $\hat{\theta}$ and state estimates $\hat{x}(k)$, of the model (2.1) on-line based upon processing of the measurements of input-output data.

3. Least Squares Approach To The Joint Parameter/State Estimation

3.1 A Variable Projection Algorithm

A least squares problem is called separable if the fitting function can be written as a linear combination of the functions involving further parameters in a nonlinear manner. Golub and Pereyra [3] have recently proposed an algorithm for solving the separable nonlinear least-squares problem in which, for given data (y_i,t_i), $i = 1, 2,\ldots,m$ and given functions $\phi_j(\alpha,t)$, $j = 1, 2\ldots,n$, $(m>n)$ where $\alpha = (\alpha_1, \alpha_2,\ldots \alpha_k)$, the vectors α and $a = (a_1\, a_2 ,\ldots, a_n)$ are determined which minimize the nonlinear functional

$$r(\alpha,a) = ||\, y - \Phi(\alpha)a \,||^2 \tag{3.1.1}$$

where $\Phi(\alpha)_{i,j} = \phi_j(\alpha,t_i)$. The approach of [3] to the solution of this problem is to modify the functional $r(\alpha,a)$, in such a way that consideration of the linear parameters is deferred.

Define $\Phi^+(\alpha)$ as the Moore-Penrose pseudoinverse of matrix $\Phi(\alpha)$ [Appendix B], then for each α, the linear operator

$$P_{\Phi(\alpha)} = \Phi(\alpha)\Phi^+(\alpha) \tag{3.1.2}$$

is the orthogonal projector on the linear space spanned by the columns of the matrix $\Phi(\alpha)$. We denote the linear operator $(I - P_{\Phi(\alpha)})$ by $P_\perp$. The operator $P_\perp$ is the projector on the orthogonal complement of the column space of $\Phi(\alpha)$.

For any given α we have the minimal least squares solution

$$\hat{a}(\alpha) = \Phi^+(\alpha)y \tag{3.1.3}$$

Thus,

$$\min_{a} r(a,\alpha) = r(\hat{a},\alpha) = \| y - \Phi(\alpha)\Phi^{+}(\alpha)y \|^2$$

$$= \| P_{\perp} y \|^2 . \qquad (3.1.4)$$

The modified functional is called the variable projection functional and can be rewritten as

$$r_2(\alpha) = \|P_{\perp} y\|^2 . \qquad (3.1.5)$$

Once a critical point (or minimizer) $\hat{\alpha}$ is found for this functional, then $\hat{a}$ is obtained by replacing α by $\hat{\alpha}$ in (3.1.3). The justification for employing this procedure is given by the following Theorem [3].

<u>Theorem 2.1</u>

Let $r(a,\alpha)$ and $r_2(\alpha)$ be defined as above. We assume that in the open set $\Omega \epsilon R^k$, $\Phi(\alpha)$ has constant rank $r \leq \min(m, n)$.

(a) If $\hat{\alpha}$ is a critical point (or global minimizer in Ω) of $r_2(\alpha)$, and

$$\hat{a} = \Phi^{+}(\hat{\alpha})y , \qquad (3.1.6)$$

then $(\hat{a}, \hat{\alpha})$ is a critical point of $r(a, \alpha)$ (or a global minimizer for $\alpha \epsilon \Omega$) and $r(\hat{a}, \hat{\alpha}) = r_2(\alpha)$.

(b) If $(\hat{a}, \hat{\alpha})$ is a global minimizer of $r(a,\alpha)$ for $\alpha \epsilon \Omega$, then $\hat{\alpha}$ is a global minimizer of $r_2(\alpha)$ in Ω and $r_2(\hat{\alpha}) = r(\hat{a}, \hat{\alpha})$. Furthermore, if there is a unique $\hat{a}$ among the minimizing pairs of $r(a, \alpha)$, then $\hat{a}$ must satisfy (3.1.6).

Krogh [24] has extended this results to the more general models

$$\| y - \Psi(\alpha) - \Phi(\alpha)a \|^2 \tag{3.1.7}$$

In [4] Kaufman introduced modifications to variable projection approach based on the trapezoidal decomposition of a matrix. According to this decomposition (see [25] and [26]) there exists an orthogonal matrix Q and a permutation matrix P such that for any given $m \times n$ matrix Φ of rank r

$$Q\Phi P = \begin{pmatrix} R & S \\ 0 & 0 \end{pmatrix} \tag{3.1.8}$$

where R is an $r \times r$ nonsingular upper triangular matrix. Then if the Q matrix in (3.1.8) is partitioned into

$$Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \tag{3.1.9}$$

It is shown [4] that one can solve (3.1.1) by finding $\hat{\alpha}$ which minimizes

$$r_3(\alpha) = \| Q_2(\alpha)y \|_2 \tag{3.1.11}$$

and then determining $\hat{a}$ by (3.1.3). This modification leads to a computationally more efficient algorithm.

3.2 Least Squares Formulation of The Identification Problem

A least squares approach to joint state and parameter identification is obtained by writing (2.1) as

$$0 = x(k+1) - A_k(\theta)x(k) - B_k(\theta)u(k) - \nu(k) \tag{3.2.1}$$

Consider system (2.1) at a sequence of time instances $k+1, k, \ldots, k-M$. Then for any particular time moment $k+1$, combine Eqs. (3.2.1, 2.2) in way suggested by Duncan and Horn [1] to form one large overdetermined system.

$$\begin{bmatrix} y(k+1) \\ 0 \\ y(k) \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ y(k-M) \end{bmatrix} = \begin{bmatrix} C_{k+1}(\theta) & & & \\ -I & A_k(\theta) & & \\ & C_k(\theta) & & \\ & -I & A_{k-1}(\theta) & \\ & & \cdot & \\ & & \cdot & \\ & & \cdots & \\ & & & C_{k-M}(\theta) \end{bmatrix} \begin{bmatrix} x(k+1) \\ x(k) \\ \cdot \\ \cdot \\ \cdot \\ x(k-M) \end{bmatrix} +$$

$$+ \begin{bmatrix} 0 & \cdot & \cdots & 0 \\ B_k(\theta) & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & B_{k-1}(\theta) & 0 \cdots & \cdot\cdot \\ \cdot & 0 & 0 \cdots & \cdot\cdot \\ \cdot & \cdot & \cdots & \cdot\cdot \\ \cdot & \cdot & \cdots & \cdot\cdot \\ \cdot & \cdot & \cdots & B_{k-M}(\theta) \\ \cdot & \cdot & \cdots & 0 \end{bmatrix} \begin{bmatrix} u(k) \\ u(k-1) \\ \cdot \\ \cdot \\ \cdot \\ u(k-M) \end{bmatrix} + \begin{bmatrix} \omega(k+1) \\ \nu(k) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \omega(k-M) \end{bmatrix}$$

(3.2.3)

Equation (3.2.3) can be written more briefly as

$$Y(k+1) = F_{k+1}(\theta)X(k+1) + G_{k+1}(\theta)U(k+1) + N(k+1) \qquad (3.2.4)$$

though in the future we will usually drop the subscript.

Then based on Eq. (3.2.4) the joint state and parameter identification problem is formulated as a least-squares minimization problem.

$$\min_{X(k),\theta} \| Y(k) - F(\theta)X(k) - G(\theta)U(k) \|^2 \tag{3.2.5}$$

Let

$$r(X(k), \theta) = \| Y(k) - F(\theta)X(k) - G(\theta)U(k) \|^2 \tag{3.2.6}$$

Then at any particular time k, $\hat{X}(k)$ the optimum estimate of state sequence vector $X(k)$, and $\hat{\theta}_k$ the optimum estimate of parameter vector θ, are given by

$$(\hat{X}(k), \hat{\theta}_k) = \arg \min_{X(k),\theta} r(X(k), \theta) \tag{3.2.7}$$

3.3 Joint State and Parameter Estimator

The estimator for simultaneous state and parameter estimation is derived based on formulation (3.2.7) and the variable projection method [3]. According to [25, 26] there exists orthogonal matrix Q such that

$$QF(\theta) = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} F(\theta) = \begin{bmatrix} R \\ 0 \end{bmatrix} \tag{3.3.1}$$

where $R = R(\theta)$ is upper-triangular and $Q = Q(\theta)$. Since the second norm is unaffected by orthogonal transformations,

$$r(X(k), \theta) = \| Q_1(Y(k) - G(\theta)U(k)) - R(\theta)X(k) \|^2$$

$$+ \| Q_2(Y(k) - G(\theta)U(k)) \|^2 \tag{3.3.2}$$

Hence, functional (3.2.6) has been split into a sum of quadratic forms. For any given θ we have the minimal least-squares solution

$$\hat{X}(k, \theta) = R^{-1}Q_1(Y(k) - G(\theta)U(k)) \tag{3.3.3}$$

Thus,

$$\begin{aligned}\min_{X(k)} r(X(k), \theta) &= r(\hat{X}(k), \theta) \\ &= \| Q_2(\theta)(Y(k) - G(\theta)U(k)) \|^2\end{aligned} \tag{3.3.4}$$

Then, the minimization problem (3.2.5) is reduced to the minimization of modified functional that depends on parameter vector θ only. The modified functional may be rewritten as

$$r_2(\theta) = \| Q_2(\theta)(Y(k) - G(\theta)U(k)) \|^2 \tag{3.3.5}$$

Once the estimate $\hat{\theta}$ is obtained by minimization of (3.3.5), the estimate $\hat{X}(k)$ is obtained by replacing θ by $\hat{\theta}$ in (3.3.3). Thus the original minimization problem is separated into two subproblems that are explicitly coupled:

1) The minimization of nonlinear functional $r_2(\theta)$ to obtain estimates $\hat{\theta}$ of parameter vector θ, and

2) a linear least-squares problem to obtain state vector $X(k)$ estimates.

This separation defines structure of joint state parameter estimator as two estimators coupled together. The proposed estimator is shown in Figure 1.

4. Algorithms For Simultaneous State And Parameter Estimation

4.1 Recursive State Estimation Algorithm

The generalized state X(k+1) estimation problem is a linear least squares minimization problem given by (3.3.4). Let us denote by $\hat{\theta}_{k+1}$ the parameter vector estimate available at the k+1 time instant. Then for $\theta = \hat{\theta}_{k+1}$ the generalized state vector estimate $\hat{X}(k+1)$ is from (3.3.3) given by

$$\hat{X}(k+1) = \hat{R}_{k+1}^{-1} \cdot \hat{Q}_1^{k+1} [Y(k+1) - \hat{G}_{k+1} \cdot U(k+1)] \tag{4.1.1}$$

where

$$\hat{Q}_1^{k+1} = Q_1(\hat{\theta}_{k+1}), \quad \hat{G}_{k+1} = \hat{G}(\hat{\theta}_{k+1}) \tag{4.1.2}$$

$$\hat{R}_{k+1} = \left. \begin{bmatrix} R_{k+1,k+1}(\theta) & R_{k+1,k}(\theta) & & & & \\ & R_{k,k}(\theta) & R_{k,k-1}(\theta) & & & \\ & & \ddots & & & \\ & & & \cdot & R_{k-M-1,k-M-1}(\theta) & R_{k-M-1,k-M} \\ & & & & & \bar{R}_{k-M,k-M}(\theta) \end{bmatrix} \right|_{\theta = \hat{\theta}_{k+1}} \tag{4.1.3}$$

where $R_{ii}(\theta)$ - is $\{r \times r\}$ nonsingular, upper triangular matrix.

and

$$\hat{X}(k+1) = \begin{bmatrix} \hat{x}(k+1) \\ \hat{x}(k) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \hat{x}(k-M) \end{bmatrix}, \hat{Q}_1^{k+1}[Y(k+1) - \hat{G}_{k+1}U(k+1)] = \begin{bmatrix} b(k) \\ b(k-1) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b(k-M) \end{bmatrix} \quad (4.1.4)$$

Based on (4.1.1), (4.1.3) and (4.1.4) the recursive state estimation algorithm is given by

$$\hat{x}(k+1) = (R_{k+1,k+1}(\theta))^{-1}[b(k) - R_{k+1,k}(\theta) \cdot \hat{x}(k)]_{\theta=\hat{\theta}_{k+1}} \quad (4.1.5)$$

4.2 Smoothed State Estimation

The least squares formulation of the identification problem, together with a moving data-window approach can be used to obtain smoothing algorithms for state estimates of the system (2.1). Let us define initial state estimate as $\hat{u}(1)$

$$\hat{u}(1) = x(1) + \nu(1) \quad (4.2.1)$$

Then for any particular time $\underline{k}$ combine Eqs. (4.2.1), (3.2.1) and (2.2) to describe one large overdetermined system.

$$Y(k) = F_k(\theta)X(k) + G_k(\theta)U(k) + N(k) \quad (4.2.2)$$

where

$$Y(k) = \begin{bmatrix} \hat{u}(1) \\ y(1) \\ 0 \\ y(2) \\ . \\ . \\ . \\ 0 \\ y(k) \end{bmatrix} U(k) = \begin{bmatrix} u(1) \\ (u(2) \\ . \\ . \\ . \\ u(k-1) \end{bmatrix} X(k) = \begin{bmatrix} x(1) \\ x(2) \\ . \\ . \\ . \\ x(k) \end{bmatrix} N(k) = \begin{bmatrix} \nu(1) \\ \omega(1) \\ \nu(2) \\ \omega(2) \\ . \\ . \\ \nu(k-1) \\ \omega(k) \end{bmatrix} \tag{4.2.3}$$

$$F_k(\theta) = \begin{bmatrix} -I & & & & \\ C_1(\theta) & & & & \\ A_1(\theta) & -I & & & \\ & C_2(\theta) & & & \\ & A_2(\theta) & -I & & \\ & & . & & \\ & & . & & \\ & & & A_{k-1}(\theta) & -I \\ & & & & C_k(\theta) \end{bmatrix} \tag{4.2.4}$$

$$G_k(\theta) = \begin{bmatrix} 0 & . & . \; . \; 0 \\ 0 & 0 & . \; . \; . \\ B_1(\theta) & 0 & . \; . \; . \\ 0 & 0 & . \; . \; . \\ . & B_2(\theta) & 0 \; . \; . \\ . & 0 & . \; . \; . \\ . & . & . \; . \; . \\ . & . & . \; . \; B_{k-1}(\theta) \\ 0 & 0 & 0 \; . \; 0 \end{bmatrix} \tag{4.2.5}$$

Based on Eq. (4.2.2), the linear least squares estimate of state vector X(k), obtained by the variable projection approach is

$$\hat{X}(k) = \hat{R}_k^{-1} \cdot \hat{Q}_1^k[Y(k) - \hat{G}_k U(k)] \tag{4.2.6}$$

where Y(k) and U(k) are defined by (4.2.3), $\hat{G}_k = G_k(\hat{\theta}_k)$ and $G_k(\theta)$ is defined by (4.2.5).

$$\hat{R}_k = \left.\begin{bmatrix} R_{11}(\theta) & R_{12}(\theta) & & & & \\ & R_{22}(\theta) & R_{23}(\theta) & & & \\ & & \cdot & & & \\ & & \cdot & & & \\ & & \cdot & & & \\ & & \cdot & \cdots & R_{k-1,k-1}(\theta) & R_{k-1,k}(\theta) \\ & & & & & \overline{R}_{k,k}(\theta) \end{bmatrix}\right|_{\theta = \hat{\theta}_k} \tag{4.2.7}$$

$$\hat{X}(k) = \begin{bmatrix} \hat{x}_k(1) \\ \hat{x}_k(2) \\ \cdot \\ \cdot \\ \cdot \\ \hat{x}_k(k-1) \\ \hat{x}_k(k) \end{bmatrix} \qquad \hat{Q}_1^k[Y(k) - \hat{G}_k U(k)] = \begin{bmatrix} b(1) \\ b(2) \\ \cdot \\ \cdot \\ \cdot \\ b(k-1) \\ \overline{b}(k) \end{bmatrix} \tag{4.2.8}$$

where $\hat{x}_k(j)$ means the linear least squares estimate of state vector x(j) at time j, given y(1), y(2),... y(k), the observations up to time k. Thus, $\hat{x}_k(j)$, j = 1 ,... k−1 are smoothed estimates, $\hat{x}_k(k)$ is a filtered estimate. The filtered value, and as many smoothed values as are needed, can be found from (4.2.6) by back substitution using the fact that R (4.2.7) is an upper triangular bidiagonal matrix. For parameter estimates $\hat{\theta}(k)$, the algorithms (4.1.1) and (4.2.6) will give optimal estimates of the generalized state vector $\hat{X}(k)$.

4.3 Recursive Parameter Estimation Algorithm

At any particular time moment k ($k \geq M$), the parameter estimates can be obtained by minimizing $r_2(\theta)$ in (3.3.5),

$$\min_{\theta} r_2(\theta, h(k)) = \min_{\theta} \|Q_2(\theta)(Y(k) - G(\theta)U(k))\|^2 \tag{4.3.1}$$

where

$$h(k) = \begin{bmatrix} Y(k) \\ U(k) \end{bmatrix} \tag{4.3.2}$$

Since the value of the function $r_2(\theta, h(k))$ is random, it is natural to select θ, so that criterion

$$\frac{1}{2} E r_2(\theta, h(k)) \tag{4.3.3}$$

is minimized. Where E-denotes expectation over all h(k), for fixed values of model parameter vector θ.

Criteria such as (4.3.3) can be minimized recursively from observations using a stochastic approximation approach. The function $r_2(\theta, h(k))$ is quadratic in the prediction error vector $f(\theta, h(k))$:

$$r_2(\theta, h(k)) = f^T(\theta, k)f(\theta, k) \tag{4.3.4}$$

where

$$f(\theta,k) = Y(k) - \hat{Y}(k/\theta) \tag{4.3.5}$$

$\hat{Y}(k|\theta)$ - a prediction of $Y(k)$ using fixed model parameter θ is computed directly by

$$f(\theta,k) = Q_2(\theta)(Y(k) - G(\theta)U(k)) \tag{4.3.6}$$

The gradient of the $f(\theta,k)$ is the matrix $\psi\,(\theta, k)$,

$$\psi(\theta,k) = -\left[\frac{d}{d\theta}\, f(\theta,k)\right]^T \tag{4.3.7}$$

can be computed as

$$\psi^T(\theta,k) = -\; Q_2\left\{D[F(\theta)]\hat{X}(k) + D[G(\theta)]U(k)\right\} \tag{4.3.8}$$

where $D[F(\theta)]$, $D[G(\theta)]$ are Frechet derivatives of $F(\theta)$, $G(\theta)$, respectively (see section 6, below for computational details).

Since $r_2(\theta, h(k))$ is quadratic in $f(\theta, k)$ the stochastic approximation approach can be applied to minimize criteria (4.3.3). Using (4.3.6) and (4.3.8) will result in a recursive Gauss-Newton type prediction error algorithm of the type described in [45]:

$$\begin{cases} f(\hat\theta(k-1), k) = Q_2(\hat\theta(k-1))[Y(k) - G(\hat\theta(k-1))U(k)] \\ R(k) = R(k-1) + \gamma(k)[\psi(\hat\theta(k-1), k)\psi^T(\hat\theta(k-1), k) - R(k-1)] \\ \hat\theta(k) = \hat\theta(k-1) + \gamma(k)R^{-1}(k)\psi(\hat\theta(k-1), k) \cdot f(\hat\theta(k-1), k) \end{cases} \quad (4.3.9)$$

where the gain sequence $\{\gamma(k)\}$ satisfies

$$\gamma(k) \geq 0, \qquad \sum_{k=1}^{\infty} \gamma(k) = \infty \;, \quad \sum_{k=1}^{\infty} \gamma^2(k) < \infty.$$

4.4 Iterative Parameter Estimation Algorithm

In this section we describe iterative continuation process for parameter estimation. We consider deterministic nonlinear least squares minimization problem (4.3.1) at each stage of iterative continuation process. Then various algorithms, developed for such problems can be applied for parameter estimation.

At any particular time moment k, parameter estimation problem is stated as minimization problem given by (3.3.5) or (4.3.1). To solve problem one usually takes advantage of the fact that objective function is a sum of squares of nonlinear function. Doing this gives the Gauss-Newton method or modifications thereof. Recent theoretical and practical investigations [28,29,31] have shown the viability of the Gauss-Newton algorithm for nonlinear least squares problems. It is also the algorithm of choice in standard treatises on parameter estimation, for example, the work of Bard [30].

The Gauss-Newton algorithm applied to problem (4.3.1) with k a fixed observation time is summarized as follows:

$$\left.\begin{aligned} \hat{\theta}_k^0 &= \theta^0 \\ \hat{\theta}_k^{j+1} &= \hat{\theta}^j + \alpha_j d_j(k) \end{aligned}\right\} \tag{4.4.1}$$

where

$$d_j(k) = -K_j^+ \cdot f(\hat{\theta}_k^j, k) \tag{4.4.2}$$

and K_j^+ is the Moore-Penrose pseudo-inverse (defined in Appendix B) of the matrix

$$K_j = \psi(\hat{\theta}_k^j, k) \tag{4.4.3}$$

where $\psi(\hat{\theta}_k^j, k)$ is defined by (4.3.8). Let us write the iterative process (4.4.1) in the form

$$\hat{\theta}_k^{j+1} = J(\hat{\theta}_k^j, k) \tag{4.4.4}$$

where

$$J(\hat{\theta}_k^j, k) = \hat{\theta}_k^j + \alpha_j d_j(k) \tag{4.4.5}$$

In general under conditions described in section 5.2 below, every sequence $\{\hat{\theta}_k^j\}$, k fixed, j = 1,2,..., obtained by process (4.4.5), will converge to a critical point of $r_2(\theta, k)$ which by the definition of $r_2(\theta, k)$ is independent of k, and is the true constant parameter vector θ^*. To obtain parameter estimates, on-line, we consider following iterative continuation process: for a sequence of the iterative processes $\{J(\theta, k)\}$ $k = 0, 1, \ldots$ specified by (4.4.5) choose corresponding sequence of integers $\{j_k\}$ $k = 1, 2, ..$ such that points $\hat{\theta}_k^j$ satisfying

$$\left.\begin{aligned} \hat{\theta}_k^{j+1} &= J(\hat{\theta}_k^j, k) \\ j &= 0,1,2,..j_k \\ \hat{\theta}_{k+1}^0 &= \hat{\theta}_k^{j_k} \quad \theta_1^0 = \theta_0 \end{aligned}\right\} \tag{4.4.6}$$

are well defined.

Clearly (4.4.6) may be implemented on-line, by setting all $j_k = 1$. It will be shown in the section 5 that under local conditions described there, the following convergence result holds,

$$\lim_{k \to \infty} \hat{\theta}_k^{j_k} = \theta^* \tag{4.4.7}$$

5. Analysis of Parameter Estimation Algorithms

5.1 Convergence of Recursive Parameter Estimation Algorithm

In this section, we shall analyze asymptotic properties of the recursive parameter estimation algorithm (4.3.9). The algorithm minimizes the quadratic prediction error criterion and is a special case of the Recursive Prediction Error algorithm [45] described in Chapter 3. The convergence result to be established below is an application of theorem 4.3 in [45] and implies that estimates $\hat{\theta}(k)$ will converge to local minimum of criterion

$$\bar{E}r_2(\theta,\ h(k)) \tag{5.1.1}$$

where

$$\bar{E}r_2(\theta,\ h(k)) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} Er_2(\theta,\ h(k)) \tag{5.1.2}$$

To demonstrate the convergence result we shall impose a number of conditions on the algorithm. Nothing is assumed about the true system, other than data generation be stable and "asymptotically mean stationary."

We shall also make the following assumption on the state space model (2.1), (2.2);

M2: This matrices $A(\theta), B(\theta)$ and $C(\theta)$ are continuously differentiable with respect to the model parameters θ.

When we consider algorithm (4.3.9) we must introduce two smoothness conditions for the function $l(k,\theta,f) = r_2(\theta,\ k)$ and $H(k,R,\theta,f,\psi) = R(k) - \psi(\theta,\ k)$

$\psi^T(\theta, k)$. These are as follows.

Cr2: $H(k,R,\theta,f,\psi)$ is differentiable w.r.t. R, θ, f and ψ such that for some $C < \infty$,

$$|H(k,R,\theta,f,\psi)| \le C(1 + |f|^2 + |\psi|^2 + |R|) \tag{5.1.3}$$

$$|H_R(k,R,\theta,f,\psi)| + |H_\theta(k,R,\theta,f,\psi)| \le C(1 + |f|^2 + |\psi|^2)$$

and

$$|H_f(k,R,\theta,f,\psi)| + |H_\psi(k,R,\theta,f,\psi)| \le C(1 + |f| + |\psi|)$$

Cr3: The function $l(k,\theta,f)$ is twice-continuously differentiable w.r.t. θ and f and

$$|l_\theta(k,\theta,f)| + |l_{\theta\theta}(k,\theta,f)| \le C(1 + |f|)^2, \tag{5.1.4}$$

$$|l_f(k,\theta,f)| + |l_{f\theta}(k,\theta,f)| \le C(1 + |f|),$$

$$|l_{ff}(k,\theta,f)| \le C$$

In the algorithm (4.3.9) the inverse of the matrix $R(k)$ is used. To ensure that no problems arise here, we introduce the following condition.

R1: The generation of the matrix $R(k)$ by (4.3.9) is such that $R(k)$ is symmetric and $R(k) \ge \delta I \; \forall k$ for some $\delta > 0$.

In practice any implementation of Gauss-Newton should always include the following modification

$$\bar{R}(k) = R(k-1) + \gamma(k)[\psi(\theta, k)\psi^T(\theta, k) - R(k-1)] ,$$

$$R(k) = \begin{cases} \bar{R}(k) \text{ if } \bar{R}(k) \geq \delta I \\ R(k) + M_\delta(k) \text{ otherwise} \end{cases} \tag{5.1.5}$$

where $M_\delta(k)$ is chosen so that $R(k) \geq \delta I$. A modification of this sort is required to ensure good numerical behavior.

Regarding the gain sequence $\{\gamma(k)\}$, we shall impose the condition:

G1: $\lim_{k\to\infty} k \cdot \gamma(k) = \mu > 0.$

This condition restricts the choice of $\gamma(k)$ to asymptotically behave like μ/k.

We shall introduce conditions on the observed data $\{h(k)\}$. At first, we require the data generation to be exponentially stable. Let us denote h^k - data set made up of $h(1), h(2), \cdots h(k)$. Then this condition is expressed formally as follows.

S1: For each k, s, $k \geq s$, there exists a random vector $h_s^0(k)$ that belongs to the σ - algebra generated by h^k but is independent of h^s (for $s = k$ take $h_s^0(k) = 0$), such that

$$E\,|h(k) - h_s^0(k)|^4 < C \cdot \lambda^{k-s}, C > \infty , \lambda < 1 .$$

That is to say that what happened before time s has very small influence on what is going on at time k, $k \geq s$.

This appears to be a reasonable condition for most data sequences. The only situation where it is not realistic to assume S1 a priori is when generation of $\{h(k)\}$ may depend on past estimates.

We introduce the following condition

A3: The limits $\bar{V}(\theta)$, $\phi(\theta)$, $\Phi(\theta)$ define by

$$\left.\begin{aligned} \bar{V}(\theta) &= \bar{E} f^T(\theta, k) f(\theta, k) \\ \phi(\theta) &= \bar{E} \psi(\theta, k) \cdot f(\theta, k) \qquad \phi(\theta) \quad \left(= - [\bar{V}'(\theta)]^T\right) , \\ \Phi(\theta) &= \bar{E} \psi(\theta, k) \psi^T(\theta, k) \end{aligned}\right\} \tag{5.1.6}$$

are assumed to exist.

We have introduced a new condition on the criterion function. This condition will hold when the data sequence is asymptotically mean stationary.

We can now formulate the following theorem.

Theorem. Consider the algorithm (4.3.9). Assume that condition M2 holds, that the data generation process is exponentially stable (condition S1) and asymptotically mean stationary (condition A3). Then $\{\theta(k)\}$ will converge to a local minimum of criterion.

$$\bar{E} r_2(\theta,k) \tag{5.1.7}$$

Proof. For this case, it is easy to demonstrate that Cr2 and Cr3 are satisfied.

Cr2: From $H(k,R,\theta,f,\psi) = R(k) - \psi(\theta,k)\psi^T(\theta, k)$ we find

$$|H| = |R - \psi\psi^T| \le C(1 + |f|^2 + |\psi|^2 + |R|)$$

$$|H_R| + |H_\theta| = |1| + |0| \le C(1 + |f|^2 + |\psi|^2)$$

and

$$|H_\theta| + |H_\psi| = |0| + |2\psi| \le C(1 + |f| + |\psi|)$$

Here H_θ denotes the partial derivative w.z.t. θ, etc.

Condition Cr3: From $l(k,\theta,f) = r_2(\theta, k)$ we find

$$|f| \le C(1 + |f|)$$

$$|1| \le C$$

Condition R1: We would use modification (5.1.5) to ensure this condition.

Condition G1: This is a condition on choice of $\gamma(k)$, that can be easily satisfied.

Then, theorem 4.3 [45] can be applied, with the result that $\hat{\theta}(k)$ will converge to a local minimum of (5.1.7).

5.2 Convergence of Iterative Parameter Estimation Algorithm

Consider nonlinear least squares parameter estimation problem

$$\min_{\theta} r_2(\theta, k) \tag{5.2.1}$$

where k is fixed. The algorithm to solve (5.2.1) is the Gauss-Newton algorithm defined by (4.4.1-4.4.2). We shall analyze the convergence behavior of this algorithm.

By the definitions of $F(\theta)$ and $G(\theta)$ and assumptions on $A(\theta)$, $B(\theta)$ and $C(\theta)$ the $f(\theta, k) = Q_2^T(Y(k) - G(\theta)U(k))$ is continuously differentiable.

Let us denote $L^0 = L^0(r_2(\theta^0, k))$ the path connected component of the level set $L(r_2(\theta^0,k))$ which contains θ^0 itself.

$$L(r_2(\theta^0, k)) = \{\theta \epsilon N \mid r_2(\theta, k) \leq r_2(\theta^0, k)\} \tag{5.2.2}$$

Suppose that there is θ^0 so that L^0 is compact. Assume, further that

$$\operatorname{rank} D[f(\theta, k)] = p \tag{5.2.3}$$

(p-dimension of parameter vector θ) for all $\theta \epsilon L^0$, and that the functional $r_2(\theta,k)$ has unique critical point $\theta^* \epsilon L^0$.

Then by theorem 14.4.4 [36] there exists α_j, $j = 0, 1, 2, \ldots$ such that sequence $\{\theta^j\}$ obtained by (4.4.1) is well defined and remains in L^0, and $\lim_{j \to \infty} \hat{\theta}^j = \theta^*$. We choose some admissible steplength algorithm for α_j, say Curry-Altman.

Hence, the convergence of the iterative process (4.4.1) (k = const) has been established.

5.3 Local Convergence Analysis For Iterative Contunation Process

In this section, we analyze convergence of the iterative continuation process from the local or asymptotic viewpoint. The resulting local theorem is important because it characterizes the theoretical behavior of the process in the neighborhood of a solution. We first analyze asymptotic properties of the (4.4.4), where k is fixed. Then the behavior of the iterative continuation process (4.4.6) is considered.

Lemma 1.

Let $J : N \times N_h \subset R^n \times R^m \rightarrow R^n$ and assume

a) $J_\theta'(\theta, h)$ is strong at $\theta^* \epsilon$ int (N)

b) $\rho(J_\theta'(\theta^*, h)) < 1$

for all $h \epsilon N_h$

Then for any sequence $\{h(k)\} \subset N_h$ such that for all $k = 1, 2....$ $||h(k+1) - h(k)|| < \delta$, $\delta > 0$, there exists a norm such that

$$||J_\theta'(\theta^*, h(k))|| < 1 \tag{5.3.1}$$

for $k = 0, 1, 2...$

Proof:

For h(0) arbitrary, $h(0) \epsilon N_h$, by the assumption $\rho(J_\theta'(\theta^*, h(0)) < 1$.

Then for given $\epsilon > 0$, Lemma 2.2.8 [36] ensures the existence of norm on R^n for which

$$||J_\theta'(\theta^*, h(0))|| \leq \sigma + \epsilon \tag{5.3.2}$$

where

$$\sigma = \rho(J_\theta'(\theta', h(0))) \tag{5.3.3}$$

Choose ϵ such that $\sigma + \epsilon < 1$, then from (5.3.2)

$$||J_\theta'(\theta^*, h(0))|| < 1 \tag{5.3.4}$$

Since $J_\theta'(\theta^*, h)$ is strong for $\forall h \epsilon N_h$, then for h(k) arbitrary $h(k) \epsilon N_h$, and given $\epsilon > 0$, there exists $S(h(k),\delta)$ $\delta > 0$, such that

$$||J_\theta'(\theta^*, h(k+1)) - J_\theta'(\theta^*, h(k))|| < \epsilon \tag{5.3.5}$$

whenever

$$h(k+1) \epsilon S(h(k), \delta)$$

It follows from (5.3.5) that

$$\left.\begin{array}{l} \big|\, ||J_\theta'(\theta^*, h(k+1))|| - ||J_\theta'(\theta^*, h(k))||\, \big| < \epsilon \\ \\ \text{whenever} \\ \\ ||h(k+1) - h(k)|| < \delta \end{array}\right\} \tag{5.3.6}$$

If we assume that $||J_\theta' (\theta^*, h(k))|| < 1$ it follows from (5.3.6) that

$$||J_\theta'(\theta^*, h(k+1))|| < 1 \tag{5.3.7}$$

Then based on (5.3.4) and (5.3.6) by induction Lemma 1 follows.

<u>Theorem 1</u>

For $J : N \times N_h \subset R^n \times R^m \rightarrow R^n$ and $\theta^* \epsilon\, \text{int}(N)$ such that $\theta^* = J(\theta^*, h)$ for all $h \epsilon N_h$, assume that the collection of mappings $J_h : N \subset R^n \to R^n \quad \theta \in N,\ h \epsilon N_h$ is such that

a) $J_\theta'(\theta^*, h)$ is strong.

b) $\rho(J_\theta'(\theta^*, h)) < 1$ for all $h \epsilon N_h$

(ρ-denotes the spectral radius of $J_\theta'(\theta, h)$)

Then there are sets $S_1(\theta^*, \delta)$ and $S_2(h(0), \delta)$ $\forall h(0) \epsilon N_h$, such that for any $\theta^0 \in S_1$ and any sequence $\{h(k)\} \in S_2$ the interates $\{\theta^k\}$ given by

$$\theta^{k+1} = J(\theta^k, h(k)) \tag{5.3.8}$$

are well defined and converge to θ^*.

$$\lim_{k \to \infty} \theta^k = \theta^* \tag{5.3.9}$$

Proof:

For $h(0)$ arbitrary, $h(0) \in N_h$, by the assumption

$$\rho(J_\theta'(\theta^*, h(0)) < 1 \tag{5.3.10}$$

Then given $\epsilon > 0$, Lemma 2.2.8 [36] ensures the existence of a norm on R^n, for which

$$\|J_\theta'(\theta^*, h(0))\| \le \sigma + \epsilon \tag{5.3.11}$$

$$\sigma = \rho(J_\theta'(\theta^*, h(0))) \tag{5.3.12}$$

In this norm by the definition of strong derivative for a given $\epsilon > 0$, there is a $\delta > 0$ so that

$$\|J(\theta^k, h(k)) - J(\theta^*, h(k)) - J_\theta'(\theta^*, h(0))(\theta^k - \theta^*)\| \le \epsilon\|\theta^k - \theta^*\| \tag{5.3.13}$$

whenever $\|h(k) - h(0)\| \le \delta$, $\|\theta^k - \theta^*\| \le \delta$.

It follows from (5.3.13) and triangle inequality:

$$\|J(\theta^k, h(k)) - \theta^*\| \le \|J(\theta^k, h(k)) - J(\theta^*, h(k)) - J_\theta'(\theta^*, h(0))(\theta^k - \theta^*)\| + \|J_\theta'(\theta^*, h(0))(\theta^k - \theta^*)\| \tag{5.3.14}$$

and

$$\|J(\theta^k, h(k)) - \theta^*\| \le \epsilon\|\theta^k - \theta^*\| + \|J_\theta'(\theta^*, h(0))\|\,\|\theta^k - \theta^*\| \tag{5.3.15}$$

or

$$\|J(\theta^k, h(k)) - \theta^*\| \le (\epsilon + \|J_\theta'(\theta^*, h(0))\|)\|\theta^k - \theta^*\| \tag{5.3.16}$$

for all $h \in S_2$

We may assume that ϵ satisfies

$$\sigma + 2\epsilon < 1 \tag{5.3.17}$$

Then (5.3.11) implies $\epsilon + \|J'_\theta(\theta^*, h(0))\| < 1$, and it follows from (5.3.16) and Lemma 11.1.2 [36] that $\lim_{k\to\infty} \theta^k = \theta^*$.

Theorem 2

For $J : N \times N_h \subset R^n \times R^m \to R^n$ and $\theta^* \in \text{int}(N)$ such that $\theta^* = J(\theta^*, h)$ for all $h \in N_h$. Consider the collection of mappings

$$J_h : N \subset R^n \to R^n \quad \theta \in N,\ h \in N_h$$

$$J_h = J(\theta, h)$$

assume:

a) $J'_\theta(\theta^*, h)$ is strong.

b) $\rho(J'_\theta(\theta^*, h)) < 1$

for all $h \in N_h$.

Then there is an open neighborhood $S(\theta^*, \delta)$ of θ^* such that for any $\theta^* \in S$, and any sequence $\{h(k)\} \in N_h$ such that $\|h(k) - h(k-1)\| < \delta$ $k = 1, 2\ldots$, the iterates $\{\theta^k\}$ obtained by

$$\theta^{k+1} = J(\theta^k, h(k)) \tag{5.3.18}$$

are well defined and converge to θ^*

$$\lim_{k\to\infty} \theta^k = \theta^* . \tag{5.3.19}$$

Proof:

By the definition of a strong derivative for a given $\epsilon > 0$, there is a $\delta > 0$ so that

$$||J(\theta^k, h(k)) - J(\theta^*, h(k)) - J_\theta'(\theta^*, h(k-1)) (\theta^k - \theta^*)|| \leq$$

$$\epsilon||\theta^k - \theta^*|| \tag{5.3.20}$$

whenever

$$||h(k) - h(k-1)|| < \delta \ , \quad ||\theta^k - \theta^*|| < \delta \tag{5.3.21}$$

for all $h(k) \subset D_h$ $k = 1, 2...$ Then based on (5.3.20), (5.3.21) and triangle inequality

$$||J(\theta^k, h(k)) - \theta^*|| \leq ||J(\theta^k, h(k)) - J(\theta^*, h(k)) - J_\theta'(\theta^*, h(k-1)) (\theta^k - \theta^*)||$$

$$+ ||J_\theta'(\theta^*, h(k-1)) (\theta^k - \theta^*)|| \tag{5.3.22}$$

or

$$||J(\theta^k, h(k)) - \theta^*|| \leq ||J_\theta'(\theta^*, h(k-1))|| \, ||\theta^k - \theta^*|| + \epsilon||\theta^k - \theta^*|| \tag{5.3.23}$$

$$||J(\theta^k, h(k)) - \theta^*|| \leq (\epsilon + ||J_\theta'(\theta^*, h(k-1))||)||\theta^k - \theta^*|| \tag{5.3.24}$$

for any $h(k) \epsilon N_h$ $k = 1, 2...$ whenever (5.3.21) holds.

It follows from Lemma 1 that for any sequence $\{h(k)\} \subset N_h$ satisfying (5.3.21) there exist a norm such that

$$||J_\theta'(\theta^*, h(k))|| < 1 \tag{5.3.25}$$

for all $h(k) \epsilon N_h$.

Since (5.3.25), we may assume that ϵ satisfies

$$\epsilon + ||J_\theta'(\theta^*, h(k))|| < 1 \tag{5.3.26}$$

for all $h(k) \epsilon N_h$, and it follows immediately from (5.3.24) and 11.1.2 [36] that

$$\lim_{k \to \infty} \theta^k = \theta^* \tag{5.3.27}$$

Assume that $F(\theta)$, $G(\theta)$ are twice continuously differentiable at θ^*. Then $f(\theta, h(k))$ will be two times Frechet differentiable in a neighborhood of θ^* and

$$D[f(\theta^*, h(k))]\, f(\theta^*, h(k)) = 0 \tag{5.3.28}$$

Assume further that rank $D[f(\theta^*, h(k))] = p$.

Then [36] there exists a ball $S(\theta^*, \delta)$, $\delta > 0$, such that the Gauss-Newton iteration function

$$J(\theta, h(k)) = \theta - K^+ f(\theta, h(k)) \tag{5.3.29}$$

is well defined; moreover $J(\theta, k)$ is Frechet differentiable at θ^*, and

$$J'(\theta^*, h(k)) = -\{D[f(\theta^*, h(k))]^T D[f(\theta^*, h(k))]\}^{-1}$$

$$\cdot f''(\theta^*, h(k)) \cdot f(\theta^*, h(k)) \tag{5.3.30}$$

If in addition

$$\rho(J'(\theta^*, h(k))) = \sigma < 1 \tag{5.3.31}$$

then θ^* is a point of attraction for the iterative process (4.4.4) or

$$\lim_{j \to \infty} \theta^j = \theta^* \tag{5.3.32}$$

where θ^j are obtained by (4.4.4), and $\rho(J'(\theta, h(k)))$ is the spectral radius. From assumptions on $F(\theta)$, $G(\theta)$ and (5.3.30) follows that $J'(\theta^*, h(k))$ is continuous at θ^*.

From (5.3.28) and (5.3.29) follows that

$$J(\theta^*, h(k)) = \theta^* \tag{5.3.33}$$

We established sufficient conditions for convergence of iterative process (4.4.4) with h(k) fixed, Theorem 10.2.1 [36].

Next, we consider convergence of the iterative continuation process. Let us denote by N_h set of vectors h, such that for any $h \in N_h$, sufficient conditions for convergence of iterative process (4.4.4) are satisfied. Then we formulate the following theorem.

Theorem.

For iterative continuation process (4.4.6) there exists an open neighborhood $S(\theta^*,\delta)$ of θ^*, such that for any $\theta^0 \epsilon S$, and any sequence $\{h(k)\} \subset N_h$ such that $||h(k) - h(k-1)|| \leq \delta$, the iterative continuation process, $J(\theta^k, h(k))$, (4.4.6) is well defined and $\lim_{k \to \infty} \theta^k = \theta^*$.

Proof:

The iterative continuation process may be viewed as a collection of mappings $J(\theta, k) = J(\theta, h(k))$ characterized by vector $h(k)$. By the definition of N_h, for all $h \epsilon N_h$, $J(\theta^*, k)$ exists and is continuous at θ^*. Then by Theorem 3.2.10 [36] $J'(\theta^*, h(k))$ is strong for all $h \epsilon N_h$. Conditions (5.3.31) and (5.3.33) are also satisfied for all $h \epsilon N_h$ by assumption. Then proof follows from Theorem 2.

6. Computational Details

In this section, we give computational details of implementing algorithms described in section 4. The most significant part of the implementation is the computation of the derivative of the prediction error vector, $\psi(\theta, k)$. The derivative is derived based on the results of Kaufman [4] for derivatives of orthogonal matrices and (4.2.6). Each component i of tensor $D[f(\theta, k)]$ is given by

$$D_i[Q_2(Y(k) - G(\theta)U(k))] = \tag{6.1}$$

$$\frac{\partial}{\partial \theta_i}\left\{Q_2(Y(k) - G(\theta)U(k))\right\} =$$

$$\frac{\partial}{\partial \theta_i} Q_2(Y(k) - G(\theta)U(k)) + Q_2 \frac{\partial}{\partial \theta_i}(Y(k) - G(\theta)U(k)) =$$

$$= -Q_2 \frac{\partial F(\theta)}{\partial \theta_i} R^{-1} Q_1(Y(k) - G(\theta)U(k)) - Q_2 \frac{\partial G(\theta)}{\partial \theta_i} U(k) =$$

$$= -Q_2 \left[\frac{\partial F(\theta)}{\partial \theta_i} \hat{X}(k) + \frac{\partial G(\theta)}{\partial \theta_i} U(k)\right]$$

Thus

$$D[f_j(k)] = Q_2\left\{D[F(\hat{\theta}^j_k)]\hat{X}(k) + D[G(\hat{\theta}^j_k)]U(k)\right\} \tag{6.2}$$

where $D[F(\hat{\theta})]$, $D[G(\hat{\theta})]$ are the Frechet derivatives of $F(\hat{\theta})$, $G(\hat{\theta})$ respectively.

Our implementation, is similar to that proposed by Krogh [24], and takes advantage of the sparsity of $F(\theta)$, $G(\theta)$. An outline of the procedure for computing the gradient of prediction error is as follows:

1. Compute $F(\theta)$, $G(\theta)$ and $D[F(\theta)]$, $D[G(\theta)]$.

2. Factor $F(\theta)$ using Householder transformations

$$QF = \begin{bmatrix} R \\ 0 \end{bmatrix}$$

   where R is $(M \times n) \times (M \times n)$ nonsingular, upper triangular matrix.

3. Set $v = Q(Y - GU)$ with

$$v = \begin{pmatrix} v_1 \\ \hline v_2 \end{pmatrix} \begin{matrix} \} \; M \times n \\ \} \; M \times m \end{matrix}$$

   The $f(\theta, k) = v_2$

4. Solve $R\hat{X} = v_1$ for $\hat{X}$

   $\hat{X}$-vector of generalized state estimates.

5. Set

$$W = -Q(D[F(\theta)] \cdot \hat{X} + D[G(\theta)]U)$$

   See Golub and Pereyra [3] concerning the multiplication of a vector by a tensor.

6. Partition W into $W = \begin{pmatrix} W_1 \\ \hline W_2 \end{pmatrix} \begin{matrix} \} \; M \times n \\ \} \; M \times m \end{matrix}$

The matrix

$$\psi^T(\theta, k) = W_2$$

Note that matrix Q need never be formed in full nor stored.

For the recursive algorithm (4.3.9) the gain vector is given by

$$L(k) = \gamma(k)R^{-1}(k)\psi(\theta, k) \tag{6.3}$$

$$R(k) = R(k-1) + \gamma(k)[\psi(\theta,k)\psi^T(\theta,k) - R(k-1)] \tag{6.4}$$

To compute L(k) without inverting large matrices in each step, the matrix inversion lemma is used [45], which gives

$$P(k) = [P(k-1) - P(k-1)\,\psi\,(\theta, k)\,S^{-1}(k)\,\psi^T\,(\theta, k)P(k-1)]/\lambda(k) \tag{6.5}$$

$$S(k) = \psi^T(\theta, k)\,P(k-1)\,\psi(k) + \lambda(k)I \tag{6.6}$$

where the time-varying forgetting factor $\lambda(k)$ is related to the gain sequence $\{\gamma(k)\}$ by

$$\lambda(k) = \frac{\gamma(k-1)}{\gamma(k)}\,[1 - \gamma(k)] \tag{6.7}$$

Then the gain vector is easily found to be

$$L(k) = P(k)\,\psi(k) = P(k-1)\,\psi(k)\,S^{-1}(k) \tag{6.8}$$

In case of the algorithm (4.4.4) the increment $d_j(k)$ is computed by linear least squares algorithm based on Householder transformations.

7. Simulation Studies of the Algorithms

7.1 Properties of the Algorithms

This section presents the results of some numerical experiments which demonstrate some of the properties of the derived algorithms.

The following second-order linear system that represents a state space model of the first flexible mode in a modal model for a free-free beam,

$$x(k+1) = \begin{bmatrix} 0 & 1 \\ -1.4 & -0.45 \end{bmatrix} x(k) + \begin{bmatrix} 1 \\ 0.7 \end{bmatrix} u(k)$$

$$y(k) = [1 \;\; 0]\, x(k) + \nu(k)$$

The disturbance $\nu(k)$ was obtained as $\nu(k) = \epsilon \cdot \xi(k)$, where $\xi(k)$ was simulated using a random number generator with uniform distribution on [-1, 1] and $0 < \epsilon \leq 1$. The input u(k) is also random process with the following statistics:

$$E\,\{u(k)\} = 0 \quad E\,\{u^2(k)\} = 1.0$$

and is independent of $\xi(k)$.

We denote as Algorithm I the recursive Gauss-Newton algorithm given by (4.3.9) and as Algorithm II the iterative estimation algorithm given by (4.4.6), and mw-denotes the size of the moving data-window. The parameter ϵ was varied in the interval $(10^{-1} - 10^{-4})$.

In case of Algorithm I the forgetting factor $\lambda(k)$ was computed as

$$\lambda(k) = 0.99 \cdot \lambda(k-1) + 0.01, \quad \lambda(0) = 0.95$$

and

$$P(0) = \rho I \qquad , \quad \rho = 500.$$

The evolution of the estimates of the parameters for both algorithms are plotted in Figures 3a-11a, and residual norm for generalized state vector estimates $\hat{X}(k)$ are shown in Figures 3b-11b where the true parameters are $a(1) = 1.4$, $a(2) = 0.45$, $b(1) = 1$ and $b(2) = 0.7$. The performance of both algorithms is quite good. For all practical purposes the estimates converge after fewer than 10 iterations. The Algorithm I appears more robust with respect to disturbances. The CPU time for run of 500 samples was 82.94 sec for Algorithm I and 242.59 sec for Algorithm II.

All simulation experiments have been implemented on VAX 11-780 computer, under VMS operating system.

7.2 Comparison with other Algorithms

In this section we will compare an algorithm derived in section 4 with two other algorithms for simultaneous state and parameter estimation of the general state-space model: The Recursive Prediction Error algorithm given in Appendix B and the Extended Kalman Filter given in Appendix A. The algorithm of section 4 is the of recursive parameter estimation algorithm (4.3.9), and the state estimation algorithm (4.2.6) coupled together in a manner shown in Figure 1. For further convenience we shall refer to the later as the algorithm of section 4.

When comparing different recursive algorithms key issues are speed of convergence, computational complexity and robustness.

The RPE algorithm and algorithm of section 4, are both a special case of the general prediction error algorithm derived in [45]. Therefore, asymptotic properties of the recursive algorithms are the same. The computational complexity of this family of recursive algorithms is determined by the gradient of the prediction error. The RPE algorithm is clearly more complex and time consuming then the algorithm of section 4. For the RPE algorithm, the predictor is given by (B.4), where K(k) is the time-varying Kalman gain. The gradient of the prediction can be computed by equations (B.6-B.8). Since the Kalman gain K(k) is not a direct function of θ, the components of K(k) are obtained by differentiating the Riccati equation. That gives a straightforward but lengthly set of equations, the number of equations in (B.14) is $n^2 \cdot d$, and that may be forbidding for higher order problems. As for the algorithm of section 4, the least squares formulation allows computation of the prediction error directly and the variable projection algorithm [4] gives equations for the gradient of the prediction error. Thus the main difference between two algorithms is in computational complexity.

The Extended Kalman Filter is widely used in practice for recursive identification of parameters and states for a general state-space model. The extended Kalman filter can be seen as a recursive prediction error algorithm, where the term corresponding to the coupling between the parameters and the Kalman gain K(k) has been neglected. Hence the major computational burden of the RPE algorithm is eliminated in the EKF. The penalty, however, is in the loss of certain convergence properties. The lack of coupling between K(k) and in the algorithm may, as demonstrated in [13] lead to divergence of the estimates, even for simple cases. The interpretation is consistent with observations in practical applications, that the behavior of the EKF often is worse when residuals are large and/or the inputs are small. In such cases the coupling term obviously becomes more important.

Therefore, the simultaneous state and parameter estimation algorithm derived in section 4 has the computational complexity that is significantly less than that of RPE, and convergence properties superior to those of the EKF.

To compare the speed of convergence and robustness of the algorithm of section 4 with the EKF we have conducted a series of simulation experiments. The following second order system was considered

$$x(k+1) = \begin{bmatrix} 0 & 1 \\ -1.5 & -0.7 \end{bmatrix} x(k) + \begin{bmatrix} 1 \\ -1 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(k) + \nu(k)$$

The disturbance $\nu(k)$ and input u(k) were obtained similar to that of section 7.1.

The simulation experiments demonstrated that both algorithms perform quite well for a low level of measurement noise. The algorithm of section 4 converges in fewer iterations as shown in Figures 12a and 12b. But when we increase the level of measurement noise or when estimates of the covariance matrices are not perfectly matched, or the initial values for state and parameter estimates are far from true values, the EKF generates parameter and state estimates that are unstable and diverge, while algorithm of section 4 does not require knowledge of covariance matrices and is insensitive to initial values of parameters and states. A sample simulation run for the EKF for a high level of measurement noise, is given in Figure 13a. Note that the parameter estimates are biased and the speed of convergence is slow, while the algorithm of section 4, under the same experimental conditions, generates stable and consistent estimates as shown in Figure 13b.

8. An Adaptive Control System for Fine Pointing of Flexible Spacecraft

A flexible spacecraft positioning control system was proposed in [46]. A lumped-parameter dynamic model with control actuators mounted on a central rigid body was assumed for a single axis where the transfer function between the control acceleration u and the pointing error α was written as

$$\frac{\alpha}{u}(s) = \frac{Ao}{s^2} + \sum_{j=1}^{m} \frac{Aj}{s^2 + \omega^2} \tag{8.1}$$

where ω_j is the free natural frequency of jth mode and A_j is a constant parameter $j = 1,\ldots m$. No damping was included, since, in practice, damping factors are very low (of the order of 10^{-3}), precise values being difficult to predict. The control law was assumed to provide the necessary damping actively. A state space representation of the dynamic system given by eqn. (8.1) was utilized to derive control law a bang-off-bang for actively controlling the flexure modes together with the mean spacecraft angle and rate. Switching times were calculated using the parabolic time-optimal switching boundary of the rigid body part, the origin of this switching boundary being changed according to the motion of flexure modes. The resulting bang-off-bang control function compromised a sequence of double pulses with zero time integral, separated by periods of zero control. The observations were obtained by sampling a rate integrating gyro angle output.

The above control law assumes that the parameters and states of the spacecraft are perfectly known. Therefore practical implementation of the control system requires solution of the identification problem for the flexible spacecraft.

8.1 Identification of Flexible Spacecraft

The discrete time state space model of the dynamics of a single axis represented by eqn. (8.1) may be written as

$$x(k+1) = A(\omega,h)\,x(k) + B(\omega,h)\,u\,(k) \tag{8.2}$$

$$y(k) = Cx(k) + \nu(k)$$

where x(k) is the state vector of dimension 2(m+1) with components, $x_1(k)$ the mean spacecraft angle; $x_2(k)$, the mean spacecraft rate; $x_{2j+1}(k)$ the jth flexural mode displacement; $x_{2j+2}(k)$ the jth flexural mode rate (j=1,...m). y(k) is the measurement vector of dimension n, u(k) is the input vector of dimension r,ν(k) is measurement noise, and h is the sampling interval.

$$A = \begin{bmatrix} 1 & h & & & & \\ 0 & 1 & & & & \\ & & \cos\omega_1 h & \sin\omega_1 h/\omega_1 & & \\ & & -\omega_1 \sin\omega_1 h & \cos\omega_1 h & & \\ & & & \ddots & & \\ & & & & \cos\omega_m h & \sin\omega_n h/\omega_m \\ & & & & -\omega_m \sin\omega_m h & \cos\omega_m h \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 0 \\ (1-\cos\omega_1 h)/\omega_1^2 \\ \sin\omega_1 h/\omega_1 \\ \vdots \\ (1-\cos\omega_m h)/\omega_m^2 \\ \sin\omega_m h/\omega_m \end{bmatrix}$$

C is the output matrix whose elements are determined by the physical structure of the particular spacecraft.

The adaptive control system for fine pointing of a flexible spacecraft (Figure 2) requires the solution of the following identification problem: For a dynamic system described by eqn. (8.2) obtain on-line estimates for state vector $x(k)$ and parameter vector $\omega = (\omega_1, \omega_2, \cdots \omega_m)$ from measurements of the output vector $y(k)$ and input vector $u(k)$ ($k = 1,2,\cdots$).

8.2 Adaptive Control System for Flexible Spacecraft

An adaptive control system for fine pointing of a flexible spacecraft is represented in the block diagram of Figure 2. The identifier contains the algorithms (4.3.9) and (4.2.6) for on-line state and parameter estimation presented above. The controller utilizes a control law similar to that generated by Dodds and Williamson [46] for the purpose of correcting pointing error of a flexible spacecraft.

The control function is a bang-off-bang function comprising a sequence of double pulses, positive (negative) level followed by negative (positive) level, with zero time integral, separated by periods of zero control. A non-linear state-variable feedback control law provides pulse start and finish times as function of the estimates of the pointing error, mean spacecraft rate, modal frequences, modal amplitudes and rates.

The control is calculated using a switching boundary in the (x_1, x_2) plane which is similar to the double integrator time-optimal switching boundary but displaced along the x_1 -axis by an amount x_0, which is changed according to the motion of the flexure modes. The switching times τ are calculated by

$$\tau_k = \left| \left| \frac{2u_* (x_1 (k) - x_0) - x_2^2 (k)}{\bar{u}_*(u_* - \bar{u}_*)} \right|^{1/2} \bar{u}_* + x_2 (k) \right| \frac{1}{u_*} \tag{8.3}$$

where u_* is the constant control level to be applied and $\bar{u}_*$ is the constant control level occurring immediately after the switch.

The offset x_0 is updated by the following rule

$$x_0 = \frac{1}{2} \left(x_1(k) - S_1(k)\right) \qquad \text{whenever}$$

$$S_2 (k) \qquad \text{changes sign}$$

where

$$S_1 (k) = \frac{1}{m} \sum_{i=1}^{m} x_{2i+1}(k)$$

$$S_2 (k) = \frac{1}{m} \sum_{i=1}^{m} \frac{(x_{2i+2}(k) - x_2(k))}{\omega_i}$$

8.3 Simulation Experiments

To demonstrate the performance of the adaptive control system described above, we present results of computer simulations of the dynamic response of a flexible spacecraft which contains dominant vibration modes. The simulations show adaptive control loop responses with various modal frequences for the same initial conditions. Simulation runs with one and two dominant flexible modes are presented.

The spacecraft is initially at rest but with a pointing error of 100 arcsec. The magnitude of the control acceleration is 50 arcsec s^{-2} in all cases, a typical figure obtained with low-level cold gas thrusters. The modal frequences typical of those experienced with solar arrays on communications satellites has been utilized as in [46]. The sampling interval between measurements in all experiments was 0.25 sec.

Figure 14 and 15 show responses of the spacecraft with one dominant mode. The true values for modal frequencies are 0.1 rad s^{-1} for the experiment of Figure 14 and 0.3 rad s^{-1} for the experiment of Figure 15. Initial values for modal frequencies are 0.15 and 0.45 respectively. The state variables have been brought nearly to zero.

Figure 16a shows the dynamic response in the case of a spacecraft with two flexible modes with initial guesses for modal frequency of estimates of 0.15 rad s^{-1} and 0.25 rad s^{-1}. Figure 16b shows the performance of the parameter estimation process.

Figures 17a and 17b show results similar to those of figures 4a and 4b but for a different set of modal frequency. Initial values for modal frequencies are 0.15 rad s^{-1} and 0.45 rad s^{-1}.

The simulation experiments demonstrate the ability of the proposed adaptive system to correct pointing error of a flexible spacecraft.

9. Conclusions

A least-squares, simultaneous state and parameter estimation method for a general state space model of multivariable discrete-time linear system has been presented. This method separates state and parameter estimation functions based on the special structure of the least squares problem. A robust and efficient adaptive scheme follows from this separation. Another advantage of the least squares approach is that, it does not require knowledge of the statistics of random processes.

We have obtained specific algorithms for on-line state and parameter estimation of general state-space models for multivariable discrete-time linear systems subject to an arbitrary known input, random input (driving noise), and with noisy measurements. These algorithms have been shown to be computationally efficient and robust when compared with other algorithms.

Asymptotic properties of the parameter estimates have been established. Simulation studies demonstrated the robustness of the derived algorithms under various experimental conditions.

We have considered the problem of designing an adaptive control system for fine pointing of a flexible spacecraft. The on-line simultaneous state and parameter estimation algorithm was used together with a bang-off-bang type control law. Simulation experiments demonstrated the ability of the adaptive control system to control initial pointing errors and at the same time to identify spacecraft parameters.

10. Recommendations for Further Research

The least squares approach to simultaneous state and parameter estimation shows the essential nature of the problem and opens the way for further developments in estimation techniques. Specifically the following recommendations are made for further investigations in line with this work.

1. Our analysis of the derived estimates has been limited to open loop systems, it appears that this approach can be extended to closed loop systems.

2. The least squares formulation can be extended to a weighted least squares formulation, when the statistics of the process are available. An analysis of the estimates obtained by such a modified algorithm would be of interest. One should also consider the relationship of least squares approach to other statistical approaches to the problem.

3. The separation of variables in the original optimization problem may be made into more than two sets. Then the separation theorem 2.1 can be applied recursively and the original problem will be decomposed into a sum of less complicated least squares minimization problems. This will result in a multistage least squares estimation scheme.

4. The techniques for separable nonlinear least squares have been extended [44] to the separable nonlinear least squares problem subject to separable nonlinear equality constraints. This may be used to incorporate a priori information on the parameters and on the system into the formulation of the identification problem.

Appendix A. The Extended Kalman Filter Algorithm

In this appendix we give the algorithm that results when the extended Kalman filter is applied to the simultaneous state and parameter estimation problem for linear discrete-time system modeled by (2.1).

$$\hat{x}(k+1) = A_k \hat{x}(k) + B_k u(k) + K(k)[y(k) - C_k \hat{x}(k)],$$

$$\hat{x}(0) = 0;$$

$$\hat{\theta}(k) = \hat{\theta}(k-1) + L(k)\,[y(k) - C_{k-1} \hat{x}(k)],$$

$$\hat{\theta}(0) = \theta_0$$

$$K(k) = \Big[A_k P_1(k) C_k^T + M_k P_2^T(k) C_k^T + A_k P_2(k) D_k^T + M_k P_2(k) D_k^T + R_{12}\Big] S_k^{-1}$$

$$S_k = C_k P_1(k) C_k^T + C_k P_2(k) D_k^T + D_k P_2^T(k) C_k^T + D_k P_3(k) D_k^T + R_2;$$

$$L(k) = \Big[P_2^T(k) C_{k-1}^T + P_3(k) D_k^T\Big] S_k^{-1} \;;$$

$$P_1(k+1) = A_k P_1(k) A_k^T + A_k P_2(k) M_k^T + M_k P_2^T(k) A_k^T + M_k P_3(k) M_k^T - K(k) S_k K^T(k) + R_1,$$

$$P_1(0) = \Pi_0(\theta_0)\;;$$

$$P_2(k+1) = A_k P_2(k) + M_k P_3(k) - K(k) S_k L^T(k),$$

$$P_2(0) = 0;$$

$$P_3(k+1) = P_3(k) - L(k)\, S_k\, L^T(k),$$

$$P_3(0) = P_0 .$$

Here

$$A_k = A\left(\hat{\theta}(k)\right)$$

$$B_k = B\left(\hat{\theta}(k)\right)$$

$$C_k = C\left(\hat{\theta}(k)\right)$$

$$M_k = M\left(\hat{\theta}(k), \hat{x}(k), u(k)\right)$$

with

$$M\left(\hat{\theta}\,,\ x\,,\ u\right) = \frac{\partial}{\partial\theta}\,[A(\theta)x + B(\theta)u]\ \big|_{\theta = \hat{\theta}}$$

and

$$D_k = D\left(\hat{\theta}(k-1), \hat{x}(k)\right)$$

with

$$D\left(\hat{\theta}\,, x\right) = \frac{\partial}{\partial\theta}\,[C\ (\theta)\ x]\ \big|_{\theta = \hat{\theta}}$$

Appendix B. An RPE Algorithm for a General State-Space Model

In this appendix we give the formulas for the recursive prediction error algorithm for a general state-space model (2.1, 2.2).

$$f(k) = y(k) - \hat{y}(k), \tag{B.1}$$

$$R(k) = R(k-1) + \gamma(k) \left[\psi(k) S^{-1}(k)\, \psi^T(k) - R(k-1)\right], \tag{B.2}$$

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \gamma(k)\, R^{-1}(k)\, \Psi(k)\, S^{-1}(k)\, f(k)\,, \tag{B.3}$$

$$\hat{x}(k+1) = \hat{A}_k\, \hat{x}(k) + \hat{B}_k\, u(k) + K(k)\, f(k) \tag{B.4}$$

$$y(k+1) = \hat{C}_k\, x(k+1) \tag{B.5}$$

$$M_k^* = M\left(\hat{\theta}(k), x(k), u(k)\right) + \dot{K}_k \cdot f(k)\,, \tag{B.6}$$

$$W(k+1) = \left[\hat{A}_k - K(k)\, \hat{C}_k\right] W(k) + M^* - K_k\, D_k \tag{B.7}$$

$$\psi(k+1) = W^T(k+1)\, \hat{C}_k^T + D^T\left(\hat{\theta}(k), \hat{x}(k+1)\right) \tag{B.8}$$

here

$$K(k) = \left[\hat{A}_k\, P_1(k)\, \hat{C}_k^T + R_{12}(k)\right] S^{-1}(k)\,, \tag{B.9}$$

$$P_1(k+1) = \hat{A}_k\, P_1(k)\, \hat{A}_k^T + R_1(k) - K(k)\, S(k)\, K^T(k)\,, \tag{B.10}$$

$$S(k) = \hat{C}_k\, P_1(k)\, \hat{C}_k + R_2(k) \tag{B.11}$$

and

$$\dot{K}_k^{(i)} = \left[\frac{\partial}{\partial\theta_i} \hat{A}_k P_1(k) \hat{C}_k^T + \hat{A}_k \Pi_k^{(i)} C_k^T + \right.$$

$$\left. \hat{A}_k P_1(k) \frac{\partial}{\partial\theta_i} C_k^T + \frac{\partial}{\partial\theta_i} R_{12} \right] \Big|_{\theta = \hat{\theta}(k)} \tag{B.12}$$

$$- K(k)\, \sigma_k^{(i)}\, S^{-1}(k)$$

$$\sigma_k^{(i)} = \left[\frac{\partial}{\partial\theta_i} C_k P_1(k) C_k^T + C_k \Pi_k^{(i)} C_k^T \right.$$

$$\left. + C_k P_1(k) \frac{\partial}{\partial\theta_i} C_k^T + \frac{\partial}{\partial\theta_i} R_2 \right] \Big|_{\theta = \hat{\theta}(k)} \tag{B.13}$$

$$\Pi_{k+1}^{(i)} = \left[\frac{\partial}{\partial\theta_i} A_k P_1(k) A_k^T + A_k \Pi_k^{(i)} A_k^T \right. \tag{B.14}$$

$$+ A_k P_1(k) \frac{\partial}{\partial\theta_i} A_k + \frac{\partial}{\partial\theta_i} R_1$$

$$- \dot{K}_k^{(i)} S(k) K^T(k) - K(k)\, \sigma_k^{(i)} K^T(k)$$

$$\left. - K(k) S(k) \left(\dot{K}_k^{(i)} \right)^T \right] \Big|_{\theta = \hat{\theta}(k)}$$

where $D[\hat{\theta},x]$, D_k are defined in Appendix A.

Appendix C. Definitions of Pseudo-Inverse, Frechet Derivative, Level Set and Mean Value Stationary Process

Theorem. [The Penrose Conditions] [5]

The $n \times m$ matrix A^+ is the unique matrix X that satisfies the following conditions:

a) $AXA = A$

b) $XAX = X$

c) $(AX)^T = AX$

d) $(XA)^T = XA$

Because this unique generalized inverse had previously been studied (though defined in a different way) by E. H. Moore, it is commonly known as the Moore-Penrose pseudoinverse, and is denoted by A^+.

Definition. The mapping $F : D \subset R^n \to R^m$ is Frechet - differentiable at $x \in$ int (D) if there is an $A \in L(R^n, R^m)$ such that

$$\lim_{h \to 0} (1/\|h\|) \, \| F(x+h) - f(x) - Ah \| = 0.$$

The linear operator A is again denoted by $F'(x)$ and is called the Frechet - derivative of F at x.

Definition. The Frechet derivative of $F : D \subset R^n \to R^m$ at $x^0 \subset D$ is strong if, given any $\epsilon > 0$, there is a $\delta > 0$ so that $\bar{S}(x^0, \delta) \subset D$ and

$$\| F(y) - F(x) - F'(x^0)(y-x) \| \le \epsilon \|y-x\|, \ \forall x,y \in \bar{S}(x^0,\delta).$$

Definition. If $g : D \subset R^n \to R^1$, then any nonempty set of the form $L(\gamma) = \{ x \in D \mid g(x) \le \gamma \}$, $\gamma \in R^1$, is a level set of g.

Definition. A stochastic process is mean value stationary if the mean value is not a function of time such that

$$\mu_x(k) = E\{X(k)\} = \text{const.}$$

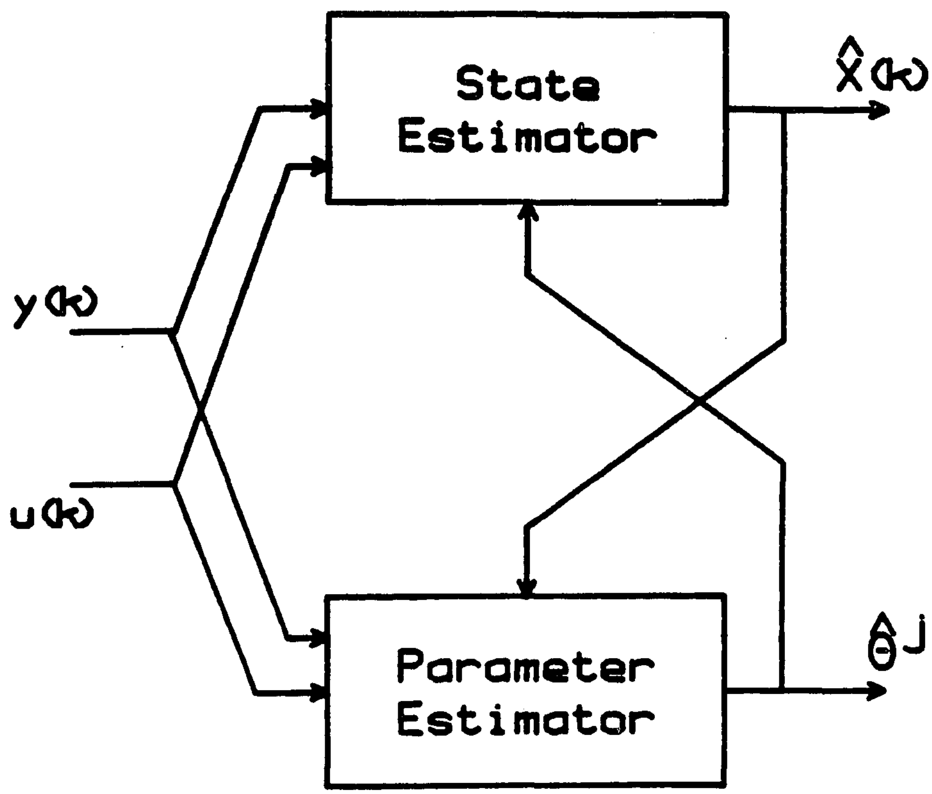


FIGURE 1 : State and Parameter Estimator

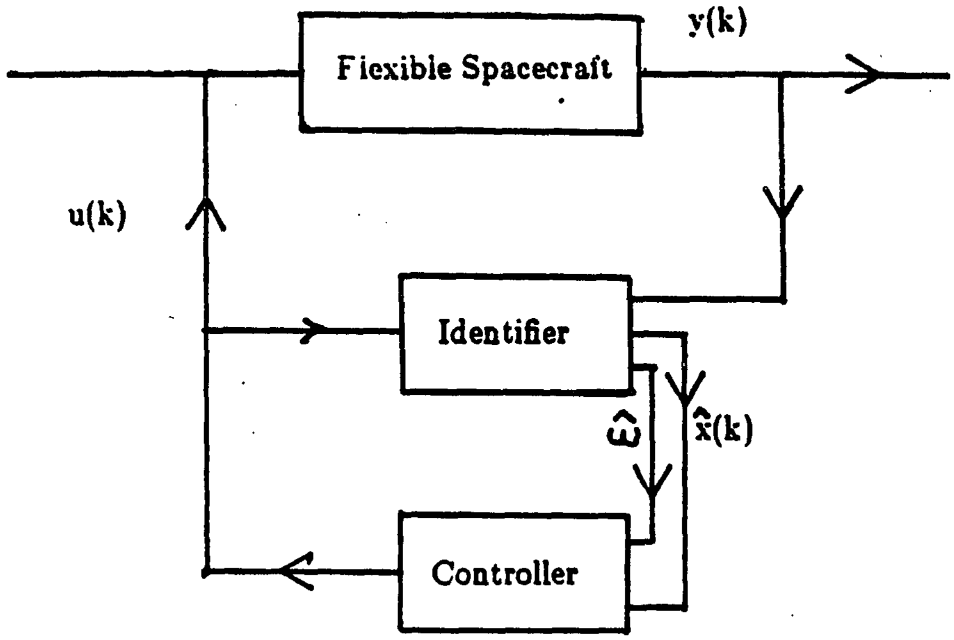

Flexible Spacecraft
y(k)
u(k)
Identifier
$\hat{\omega}$
$\hat{x}(k)$
Controller

FIGURE 2.

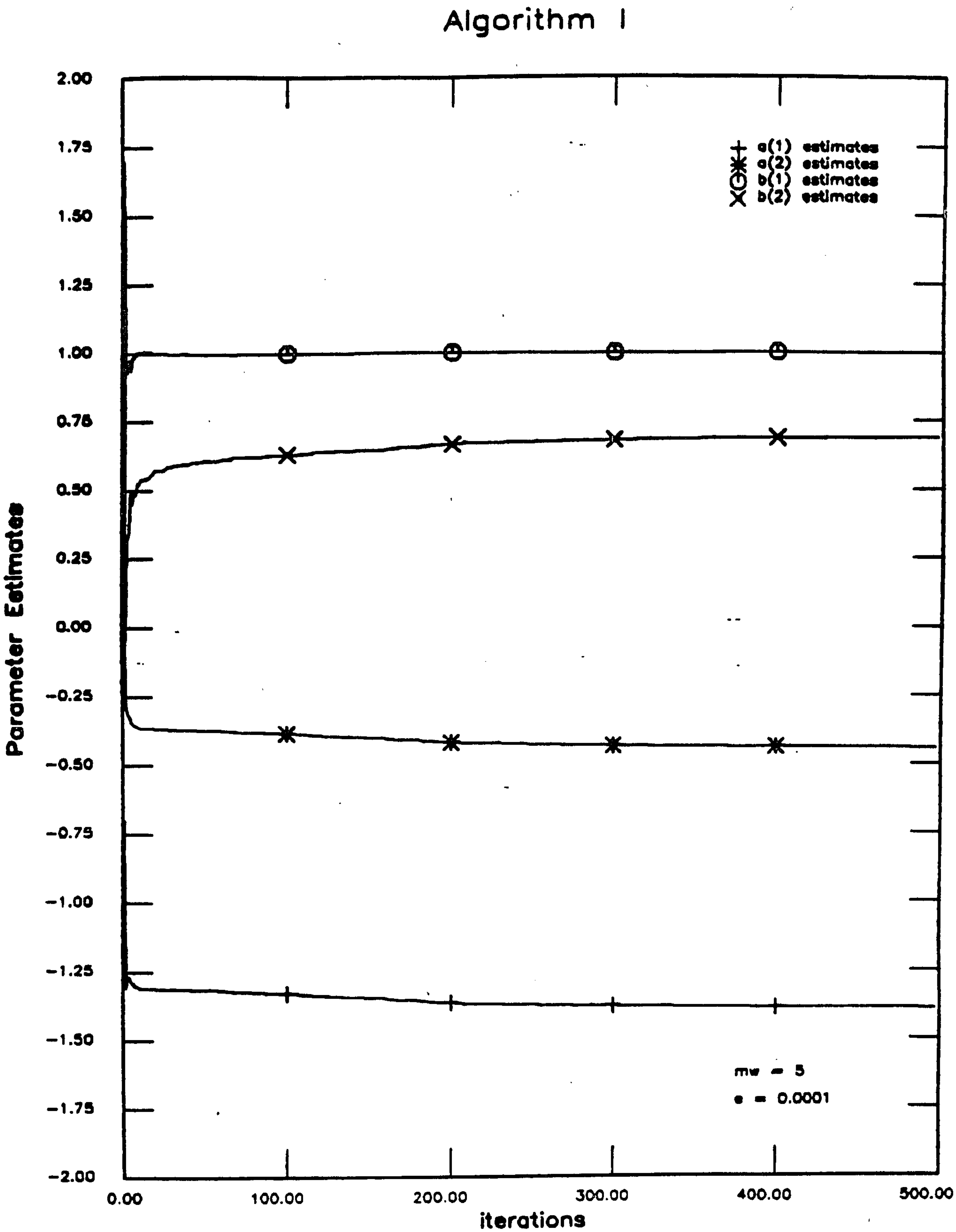


FIGURE 3a

5/ 2/88
16:28:06

Algorithm I

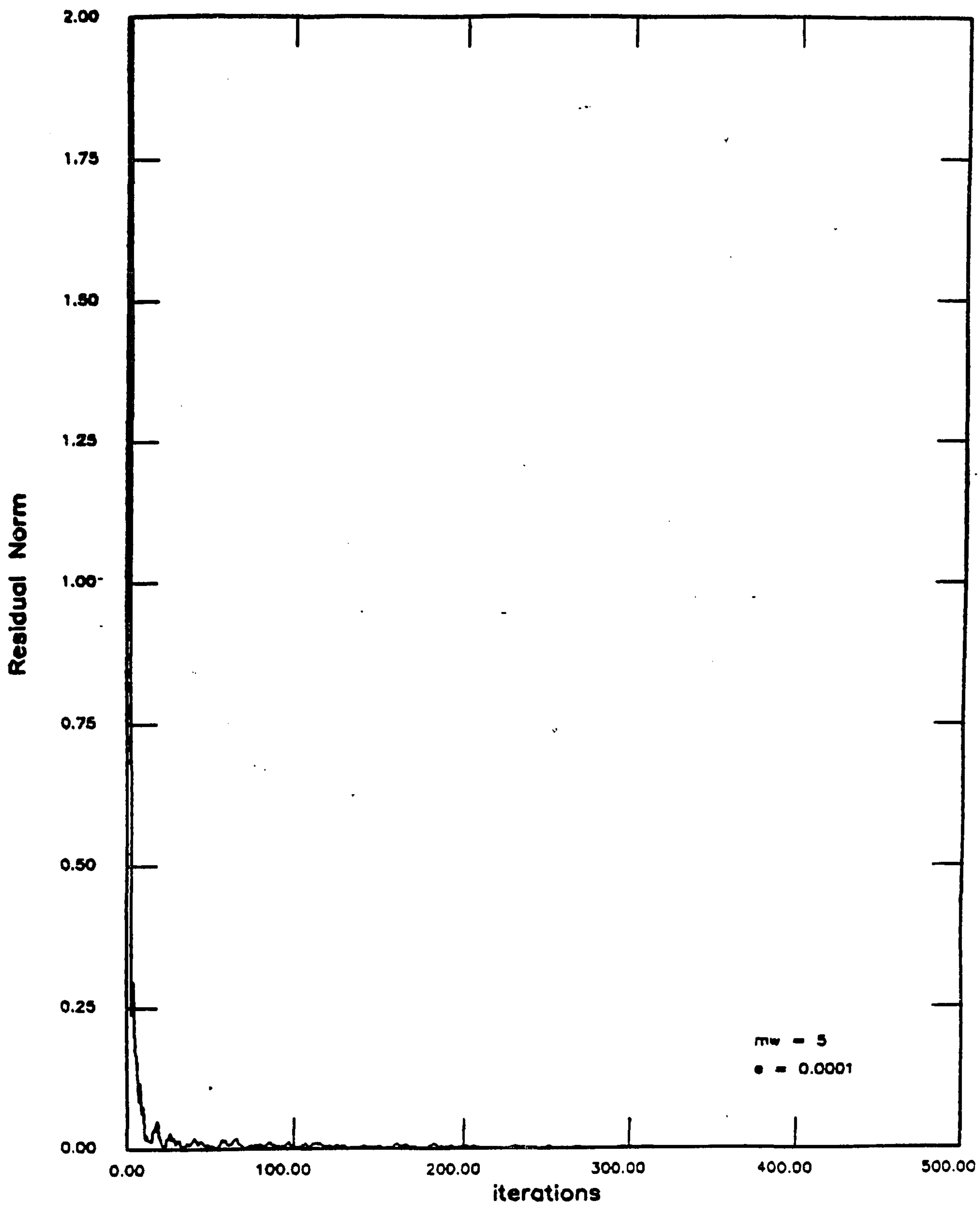


FIGURE 3b

5/ 2/86
18:31:54

POINTS OUTSIDE PLOT BOUNDARY: 2

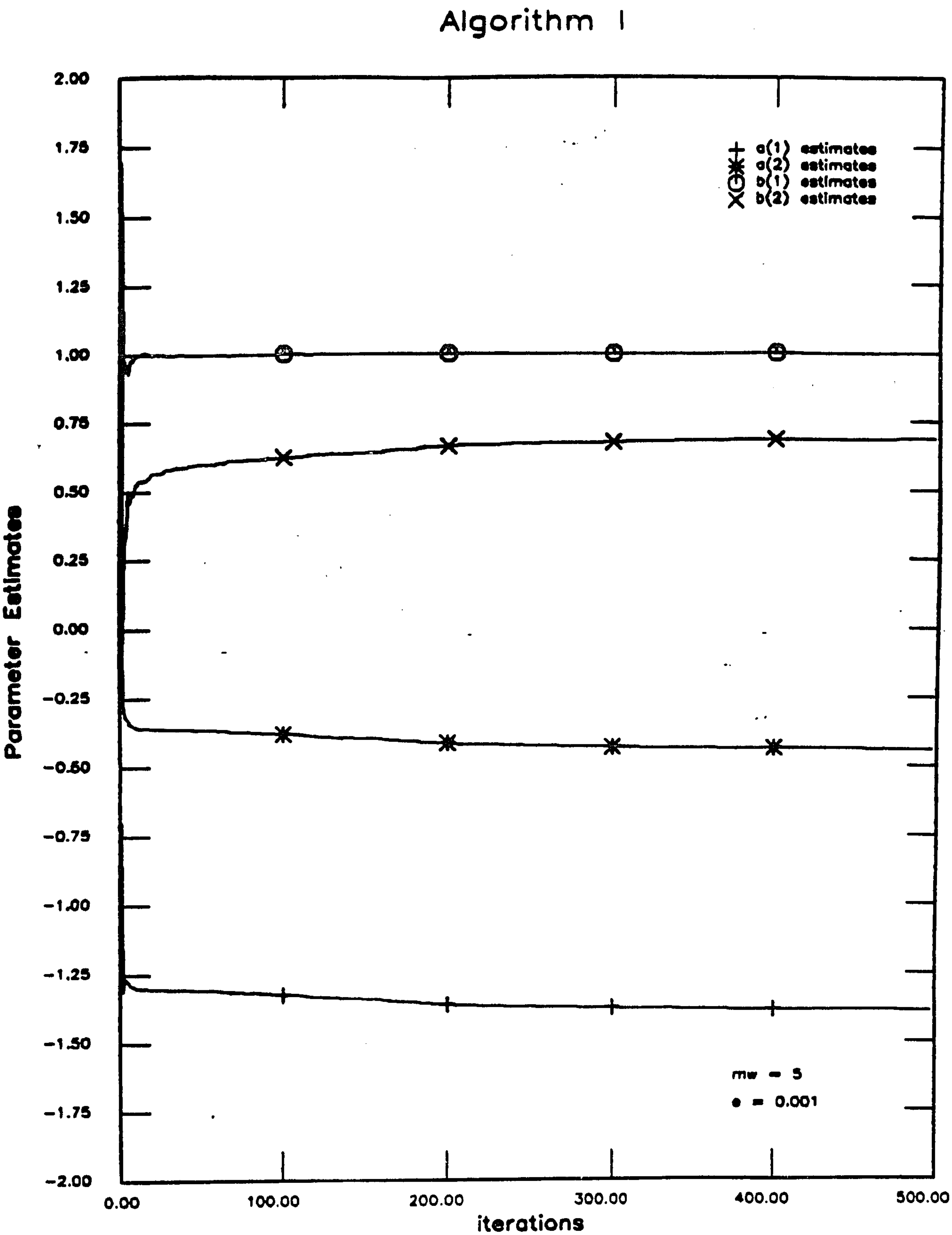


FIGURE 4a

4/30/88
13:34:43

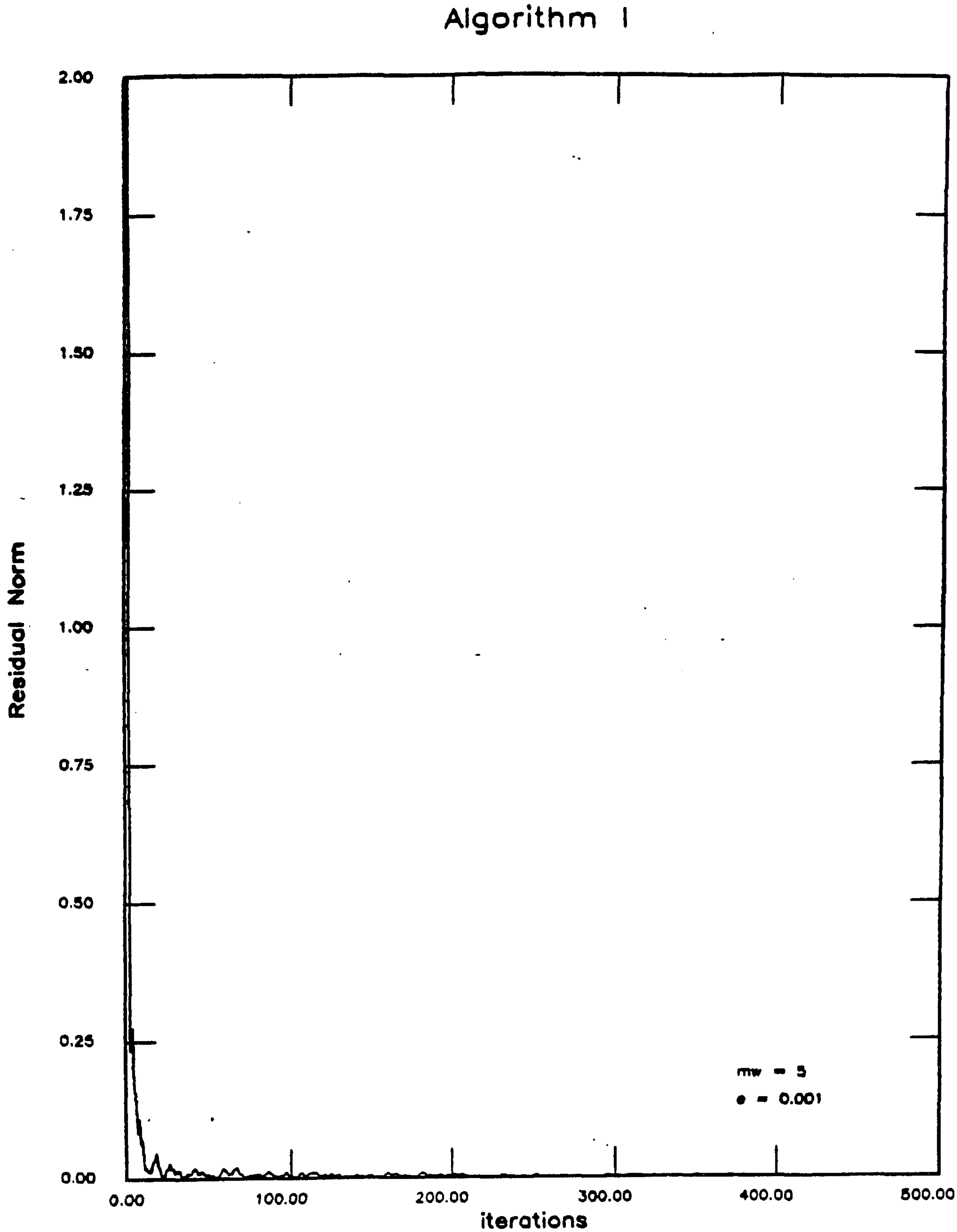


FIGURE 4b

4/30/86
13:43:47

POINTS OUTSIDE PLOT BOUNDARY: 2

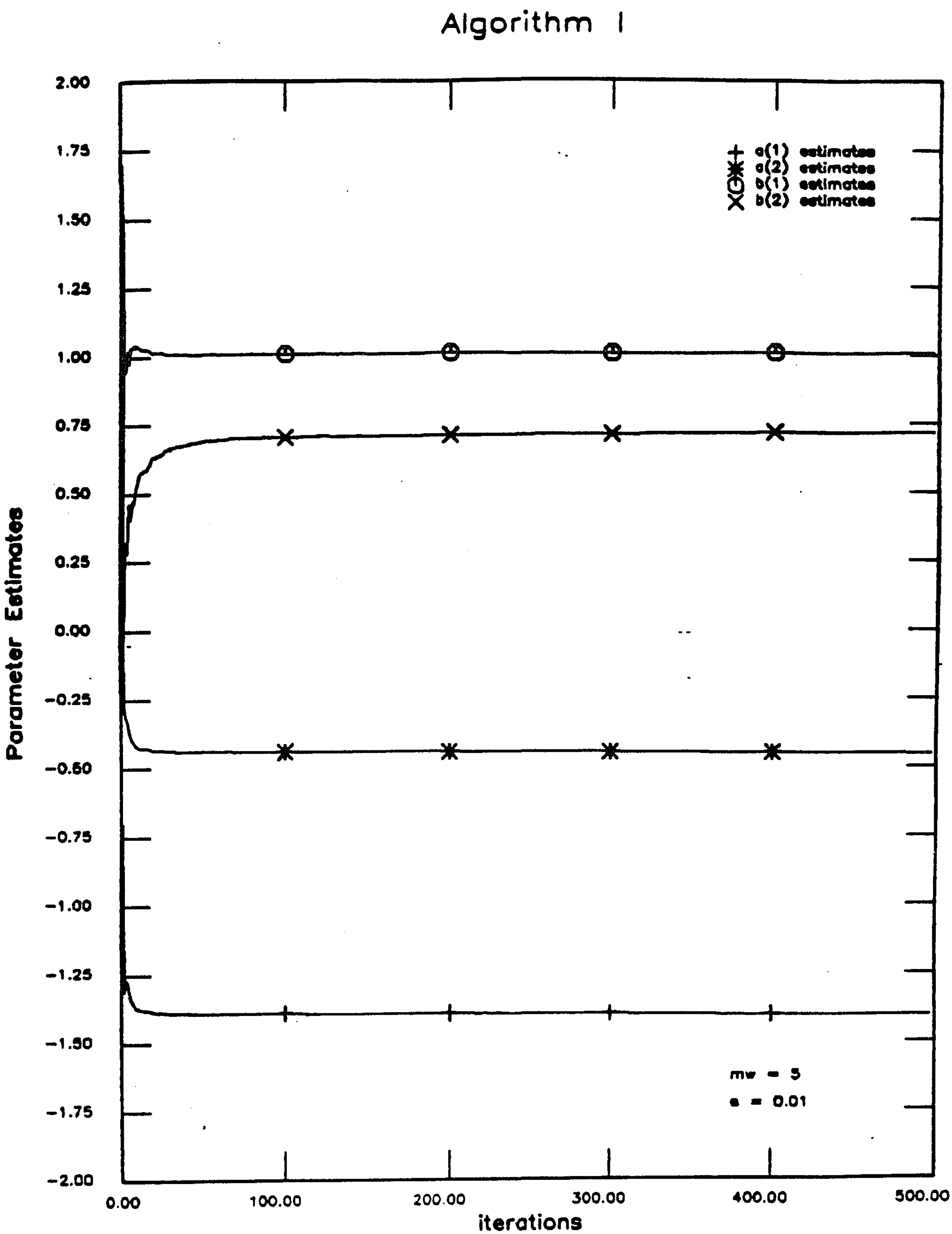


FIGURE 5a

4/30/86
11:11:46

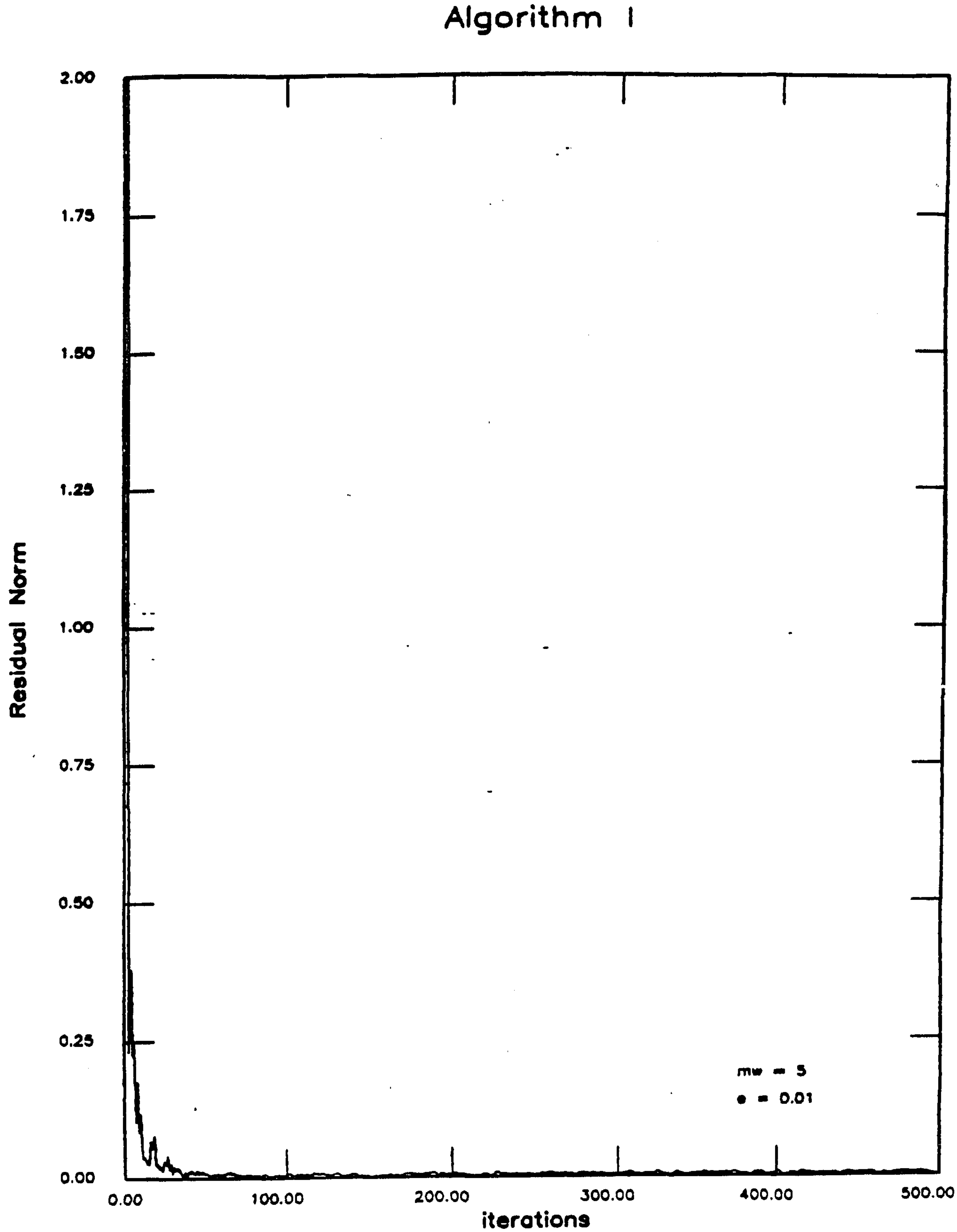


FIGURE 5b

5/ 2/86
18:50:19

POINTS OUTSIDE PLOT BOUNDARY: 2

Algorithm I

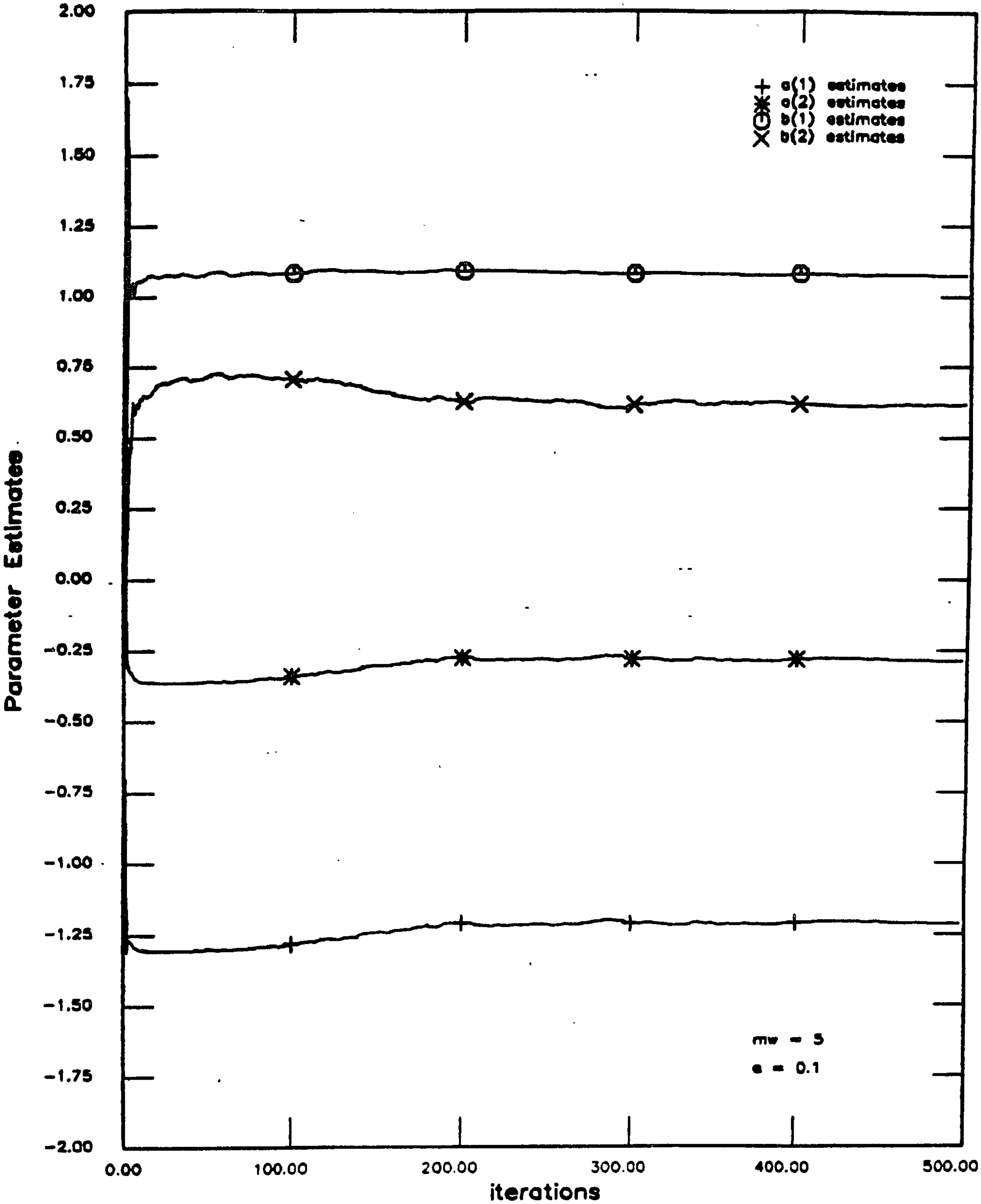


FIGURE 6a

4/30/86
11:35:37

Algorithm I

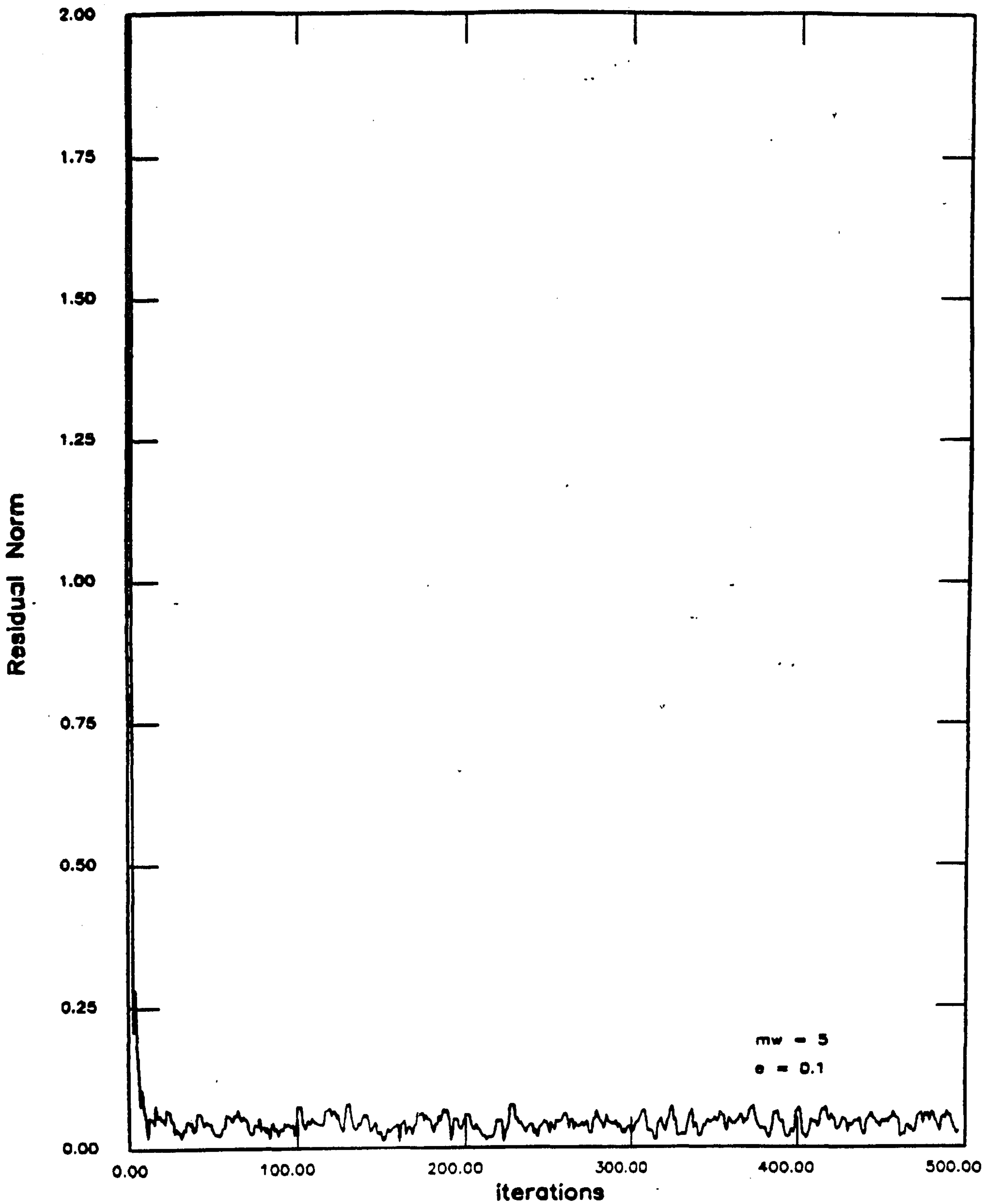
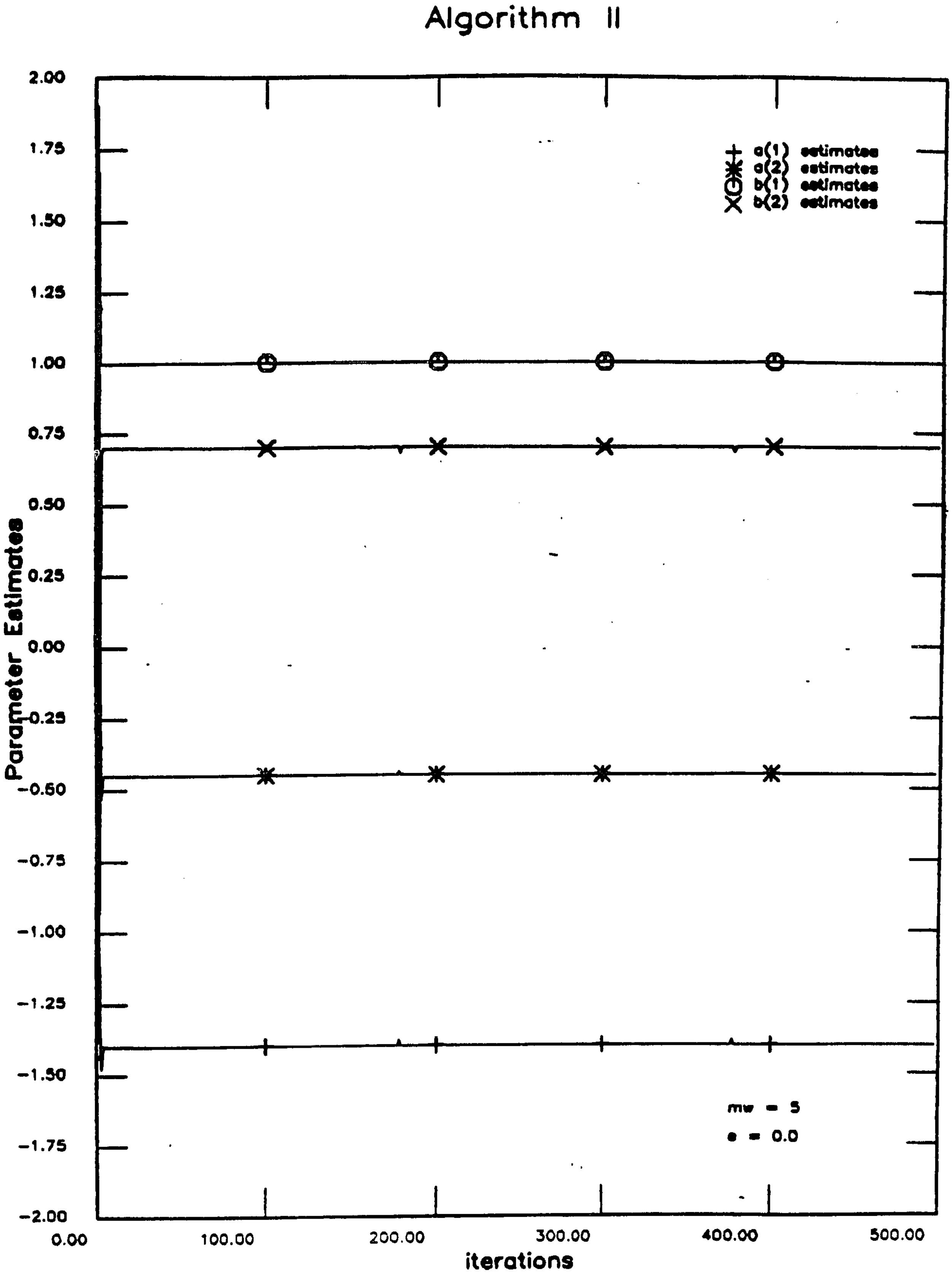


FIGURE 6b

4/30/86
12:01:19

POINTS OUTSIDE PLOT BOUNDARY: 2



Algorithm II
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
-1.50
-1.75
-2.00
Parameter Estimates
a(1) estimates
a(2) estimates
b(1) estimates
b(2) estimates
mw = 5
e = 0.0
0.00
100.00
200.00
300.00
400.00
500.00
iterations
4/27/86
18:44:21

FIGURE 7a

Algorithm II

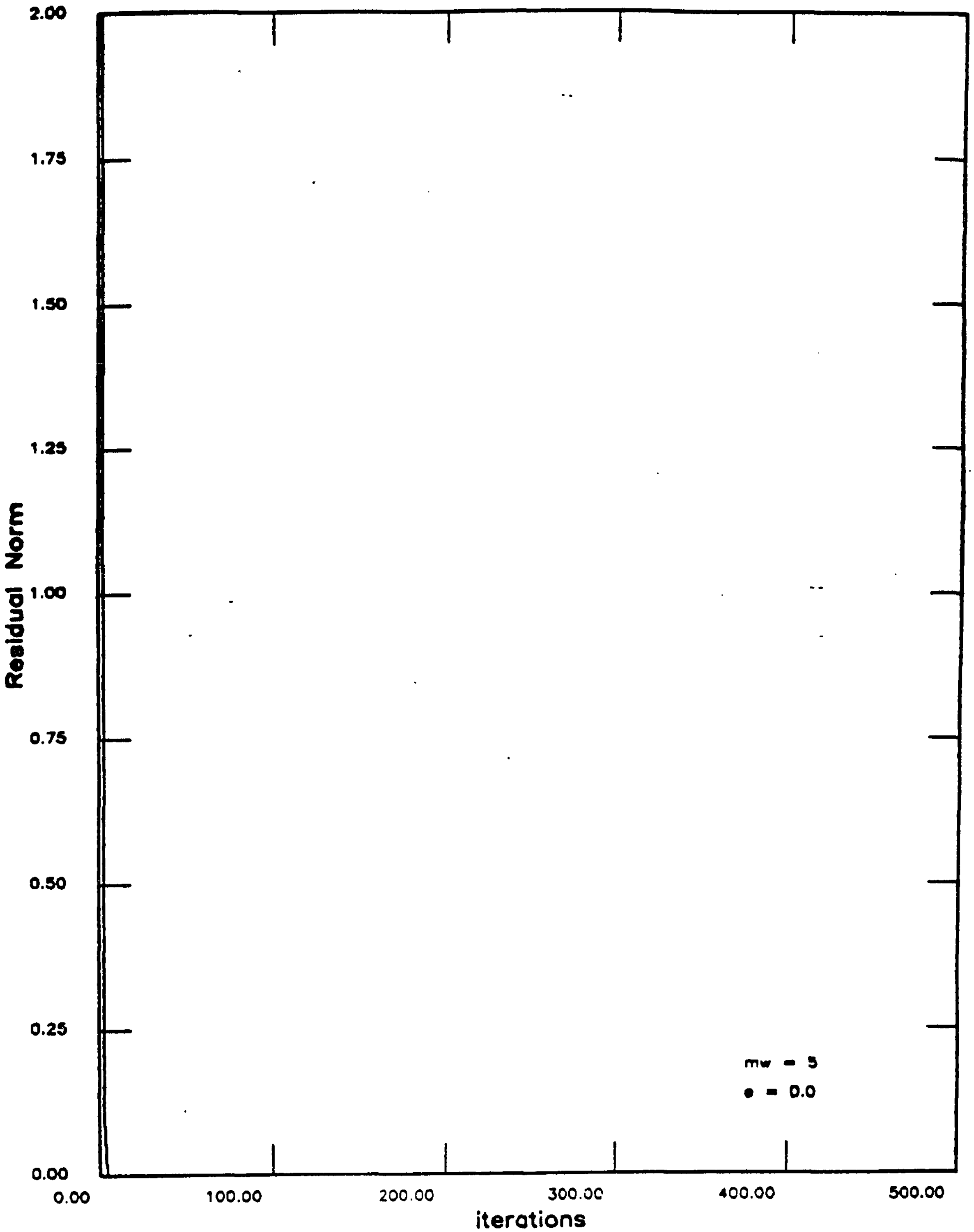


FIGURE 7b

4/27/86
19:02:33

POINTS OUTSIDE PLOT BOUNDARY 2

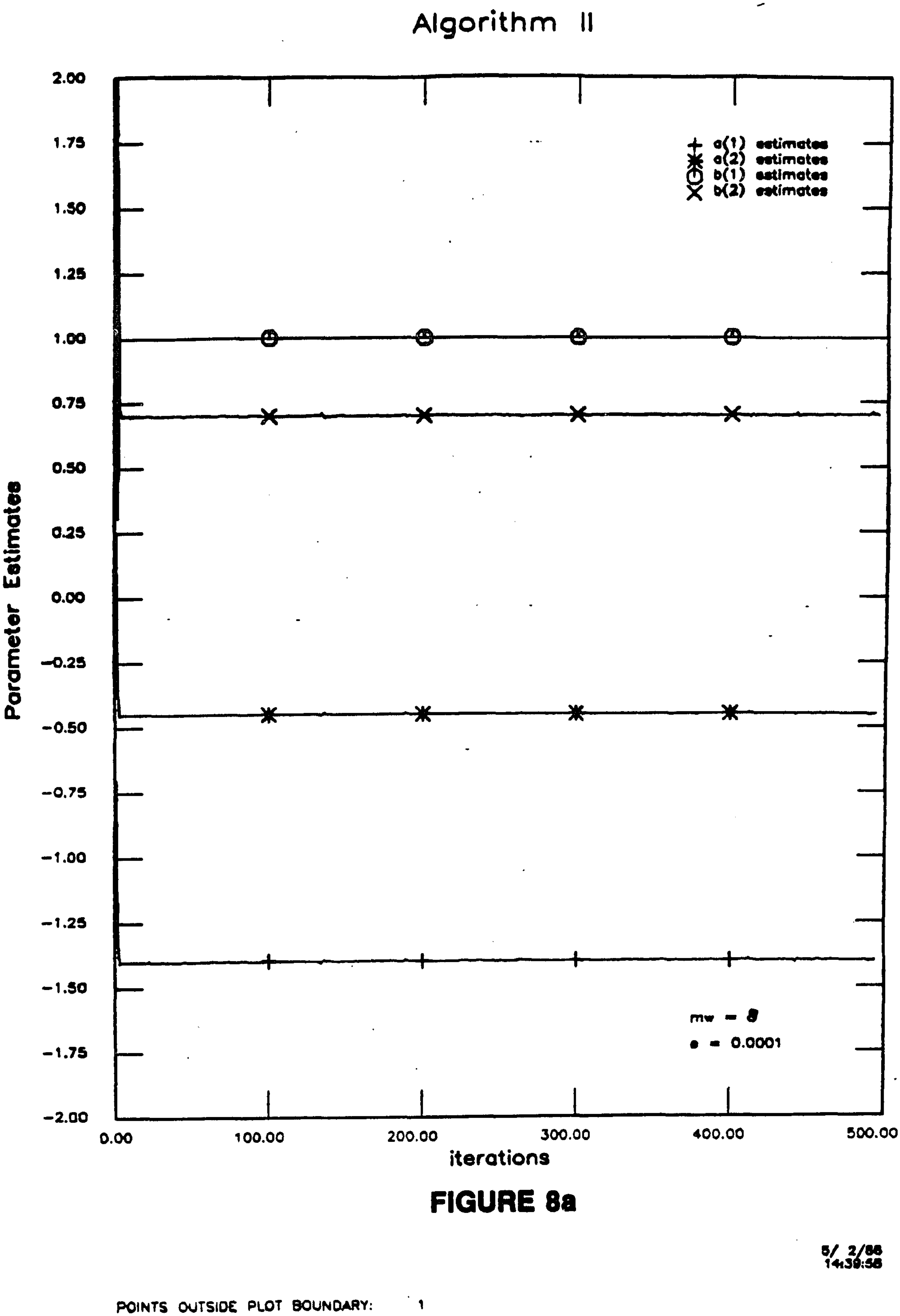

Algorithm II
a(1) estimates
a(2) estimates
b(1) estimates
b(2) estimates
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
-1.50
-1.75
-2.00
Parameter Estimates
0.00
100.00
200.00
300.00
400.00
500.00
iterations
mw = 8
e = 0.0001
5/ 2/86
14:39:58
POINTS OUTSIDE PLOT BOUNDARY: 1

FIGURE 8a

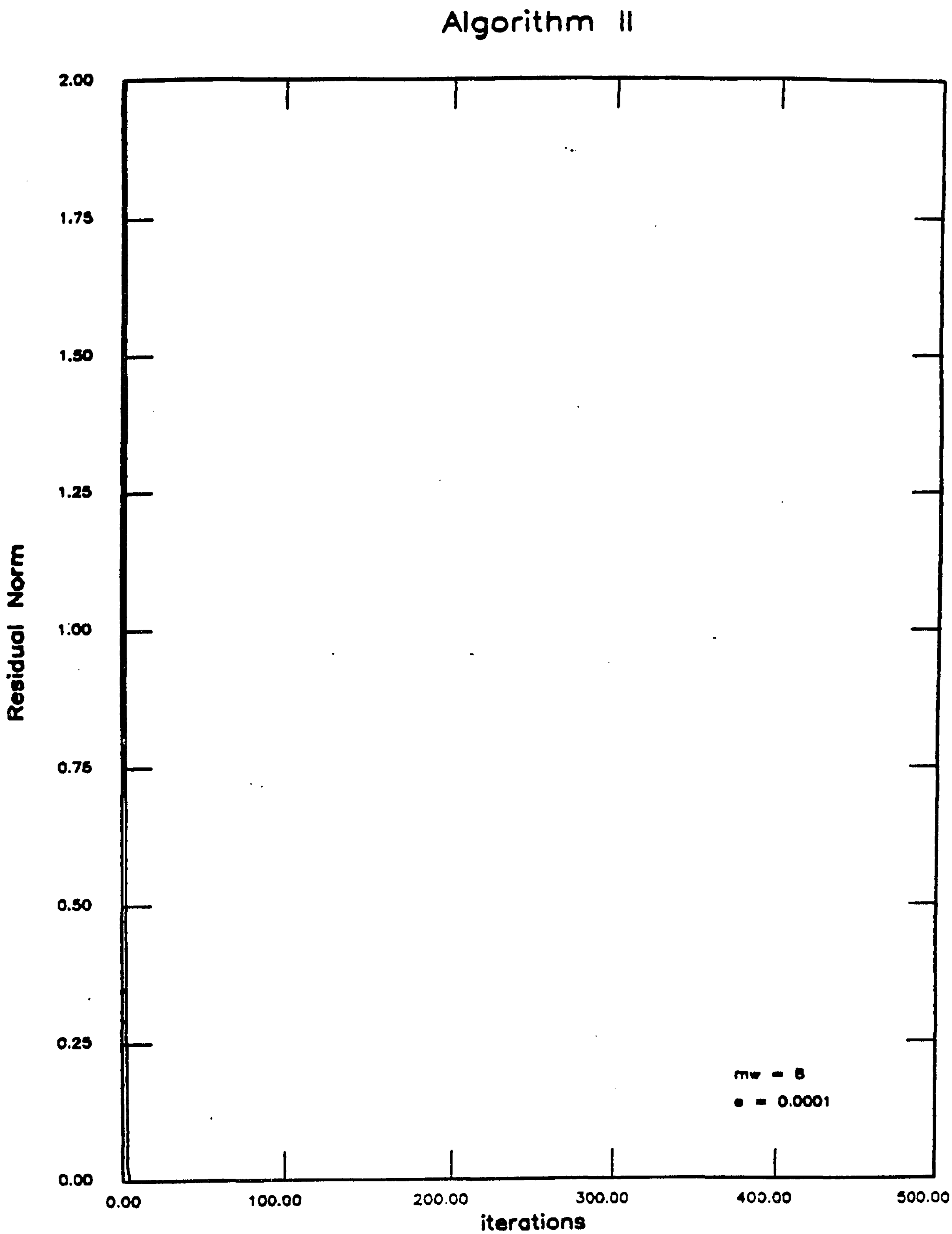


FIGURE 8b

5/ 2/86
15:03:03

POINTS OUTSIDE PLOT BOUNDARY: 1

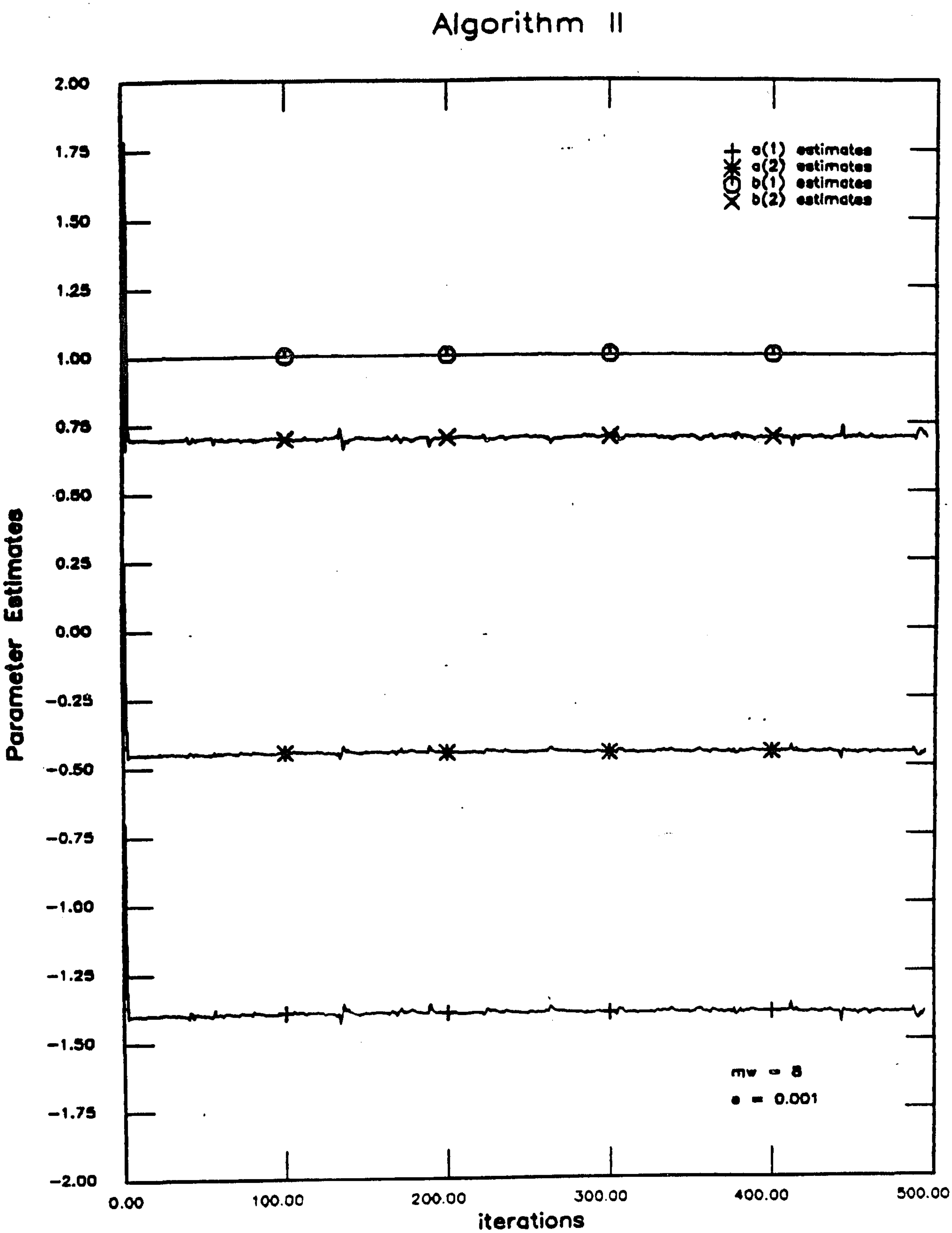


FIGURE 9a

5/ 2/86
16:41:50

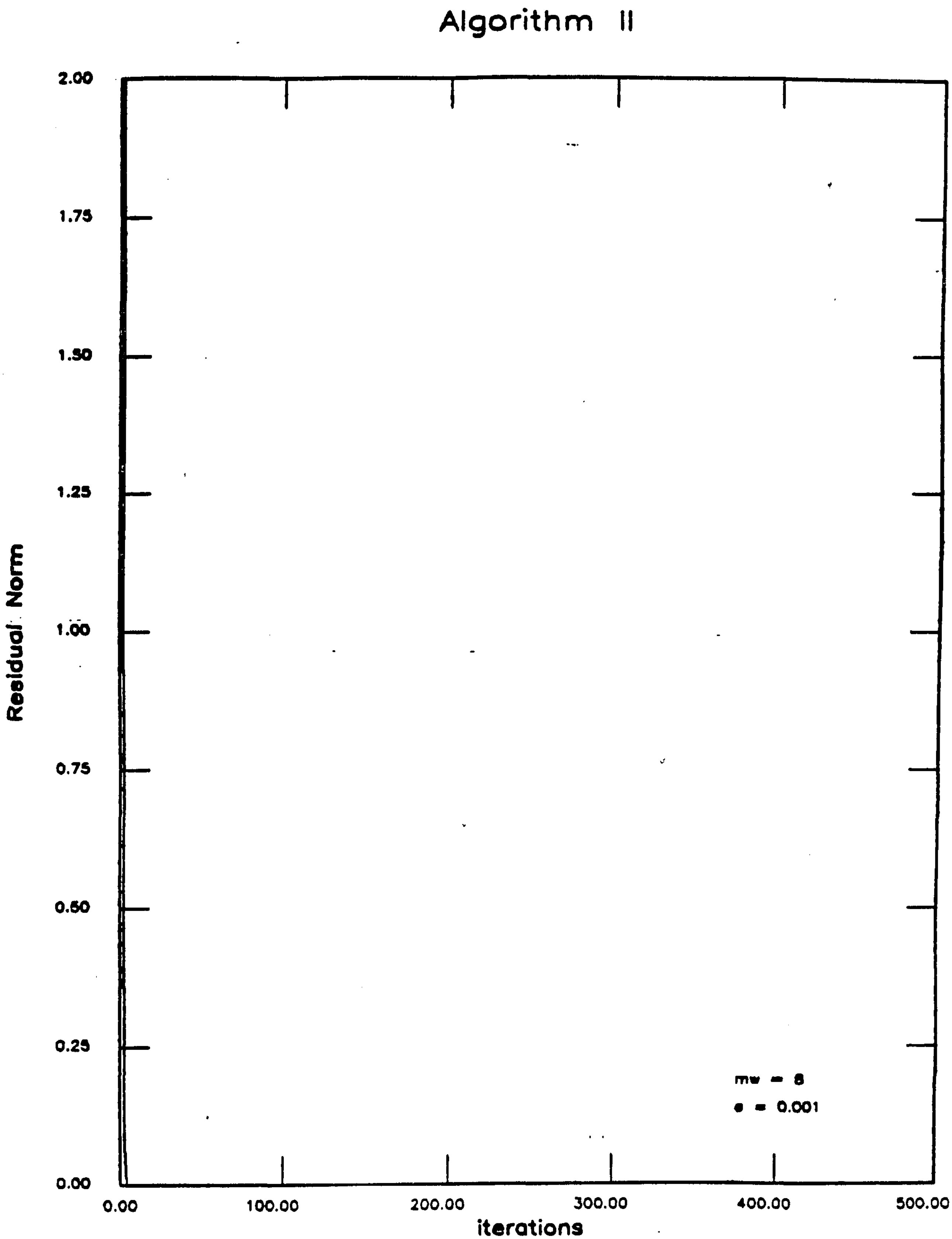


FIGURE 9b

5/ 2/86
16:53:31

POINTS OUTSIDE PLOT BOUNDARY: 1

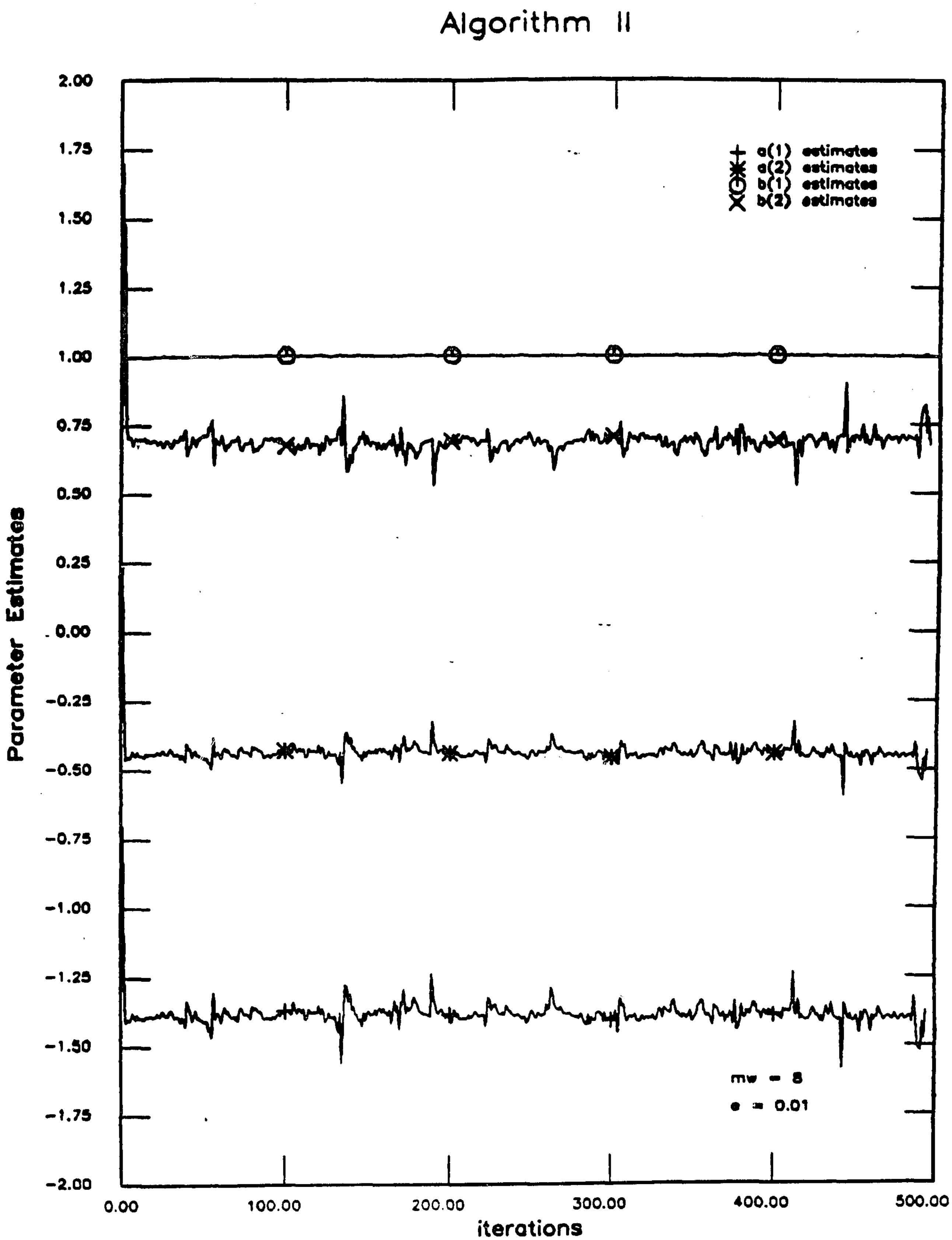


FIGURE 10a

5/ 2/86
17:53:37

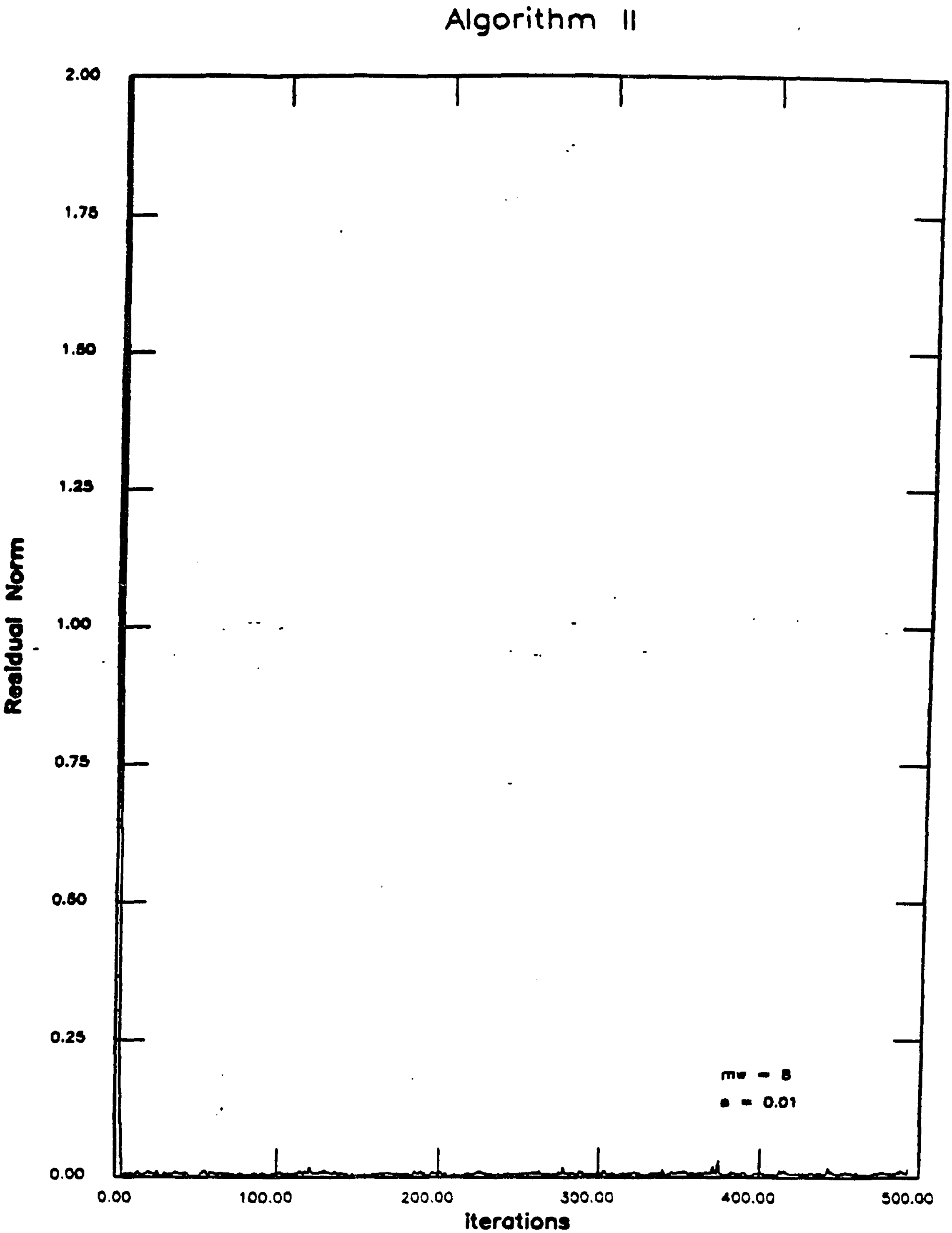


FIGURE 10b

5/ 2/88
17:57:06

POINTS OUTSIDE PLOT BOUNDARY: 1

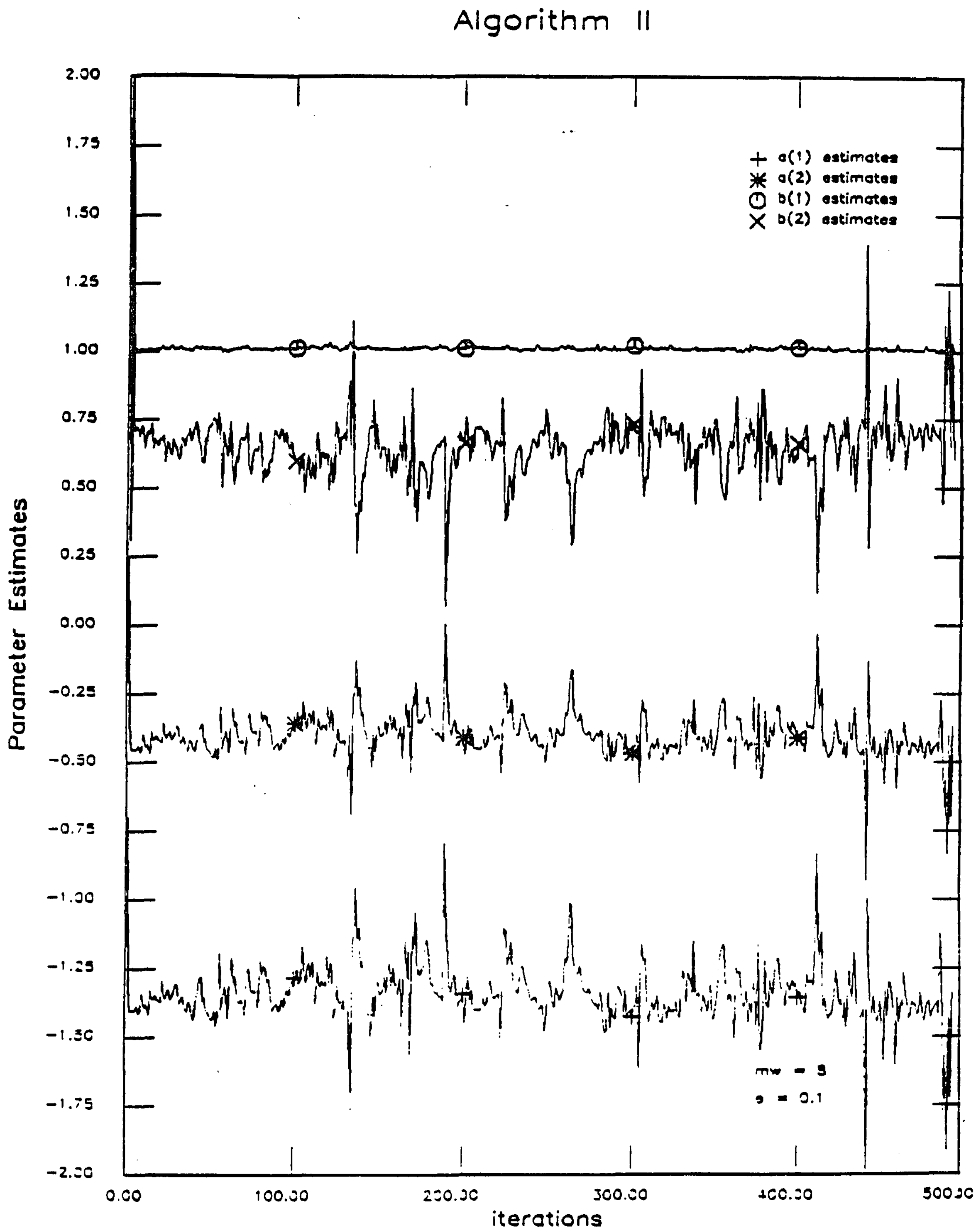


FIGURE 11a

5/16/88
21:41:35

POINTS OUTSIDE PLOT BOUNDARY: 2

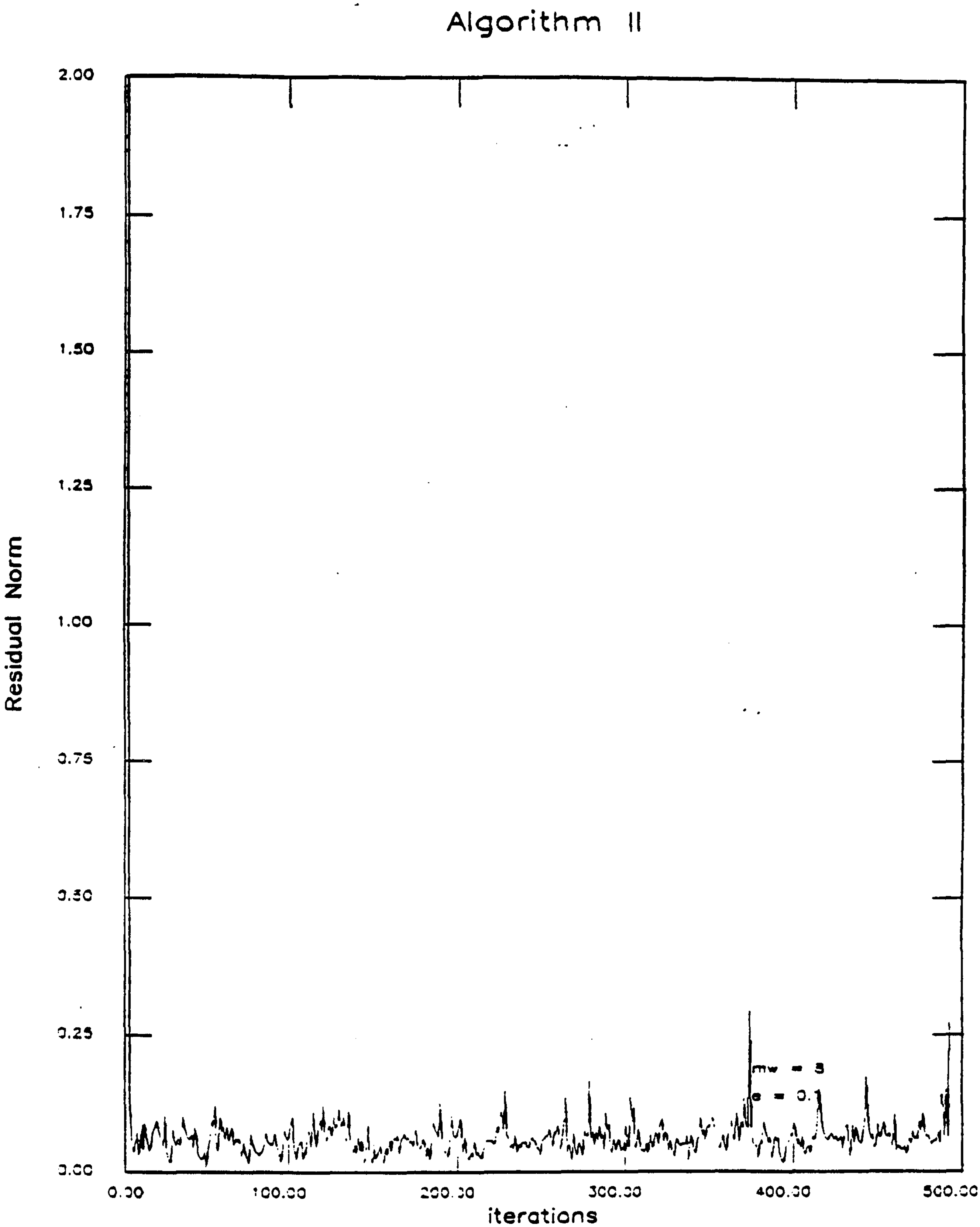


FIGURE 11b

5/16/86
21:28:36

POINTS OUTSIDE PLOT BOUNDARY: 1

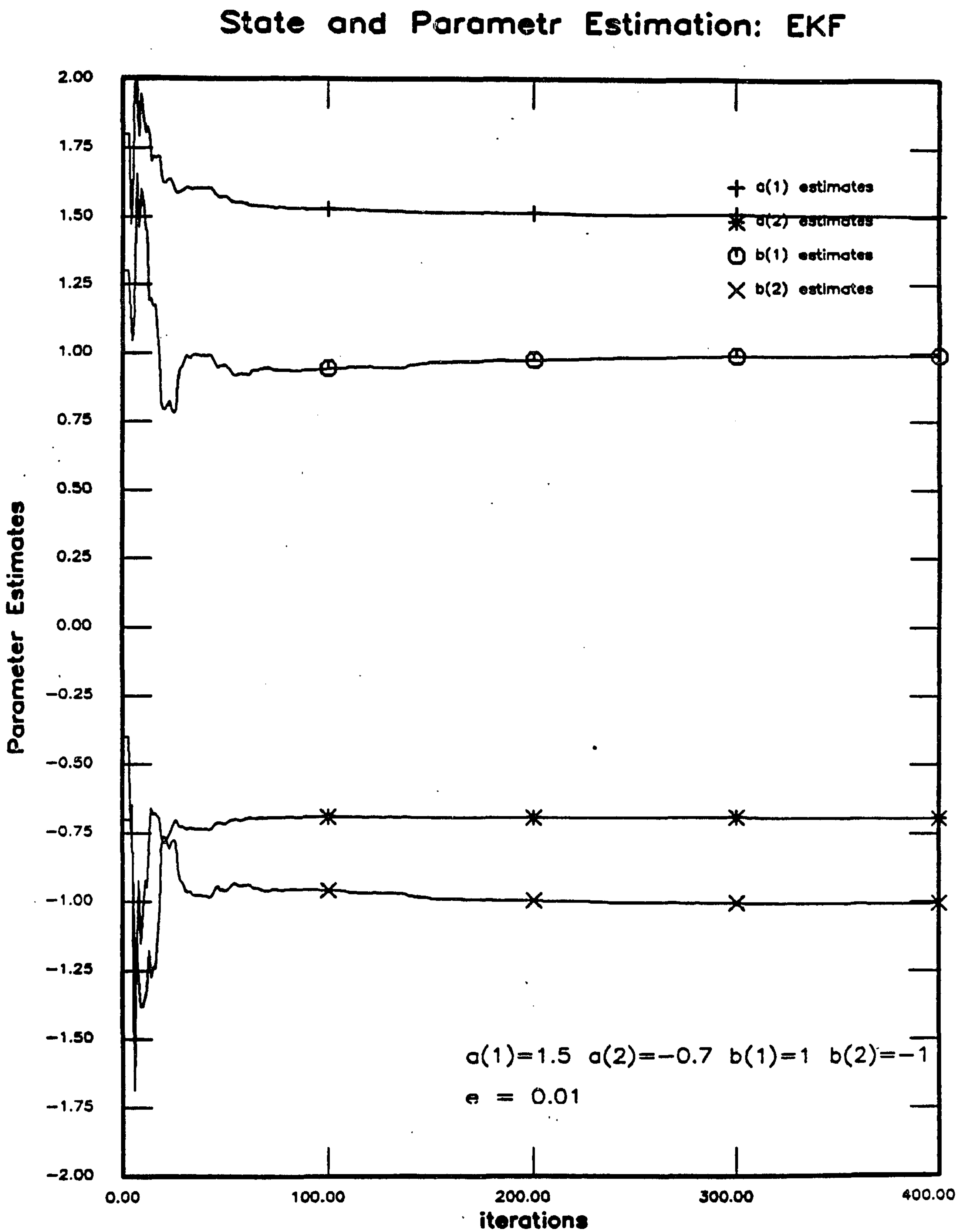


FIGURE 12a

4/ 4/87
11:42:52

]PLT.EXE;5

POINTS OUTSIDE PLOT BOUNDARY:

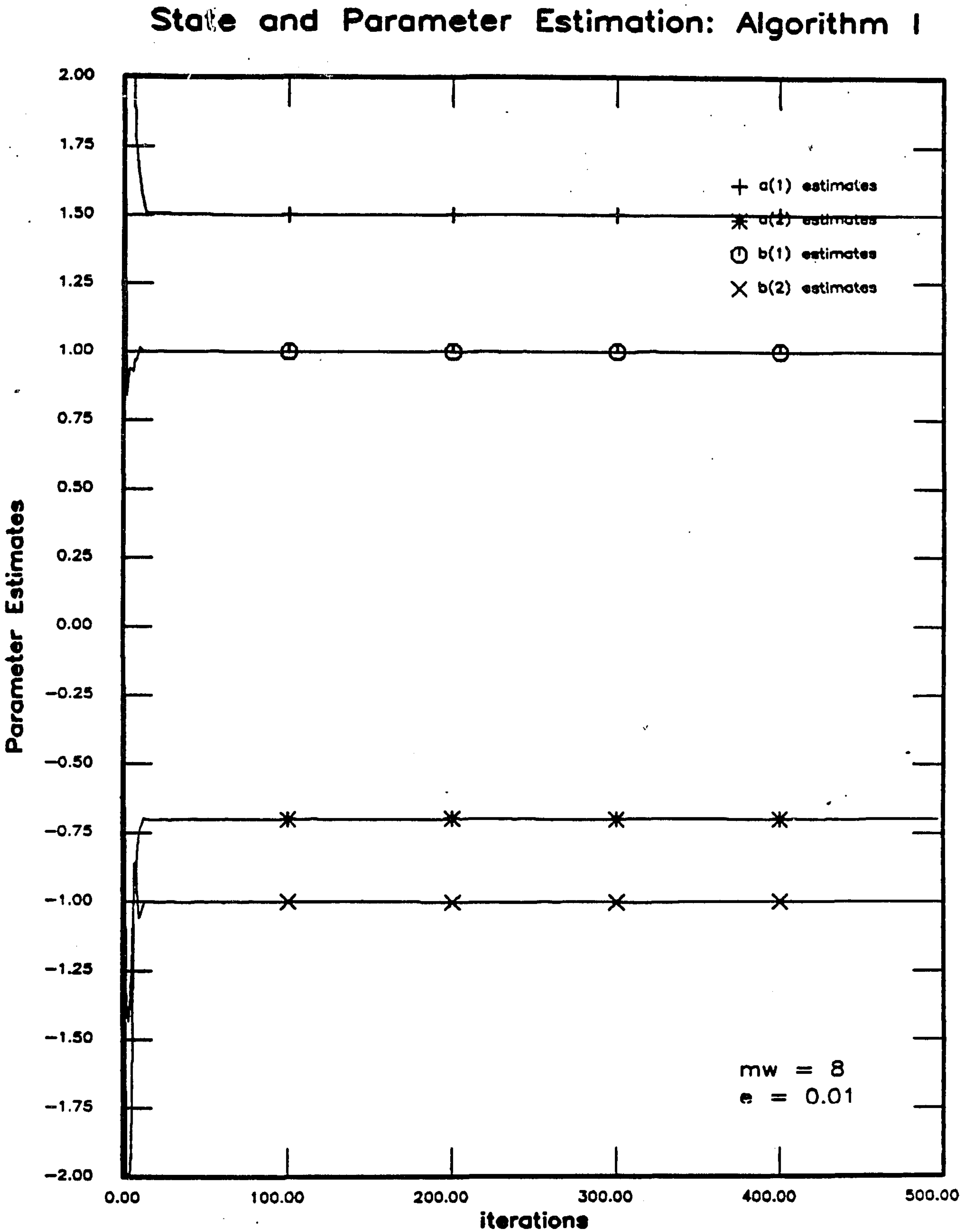


FIGURE 12b

4/ 4/87
12:18:33

[TIM.ESTM]PLT.EXE;36

POINTS OUTSIDE PLOT BOUNDARY:

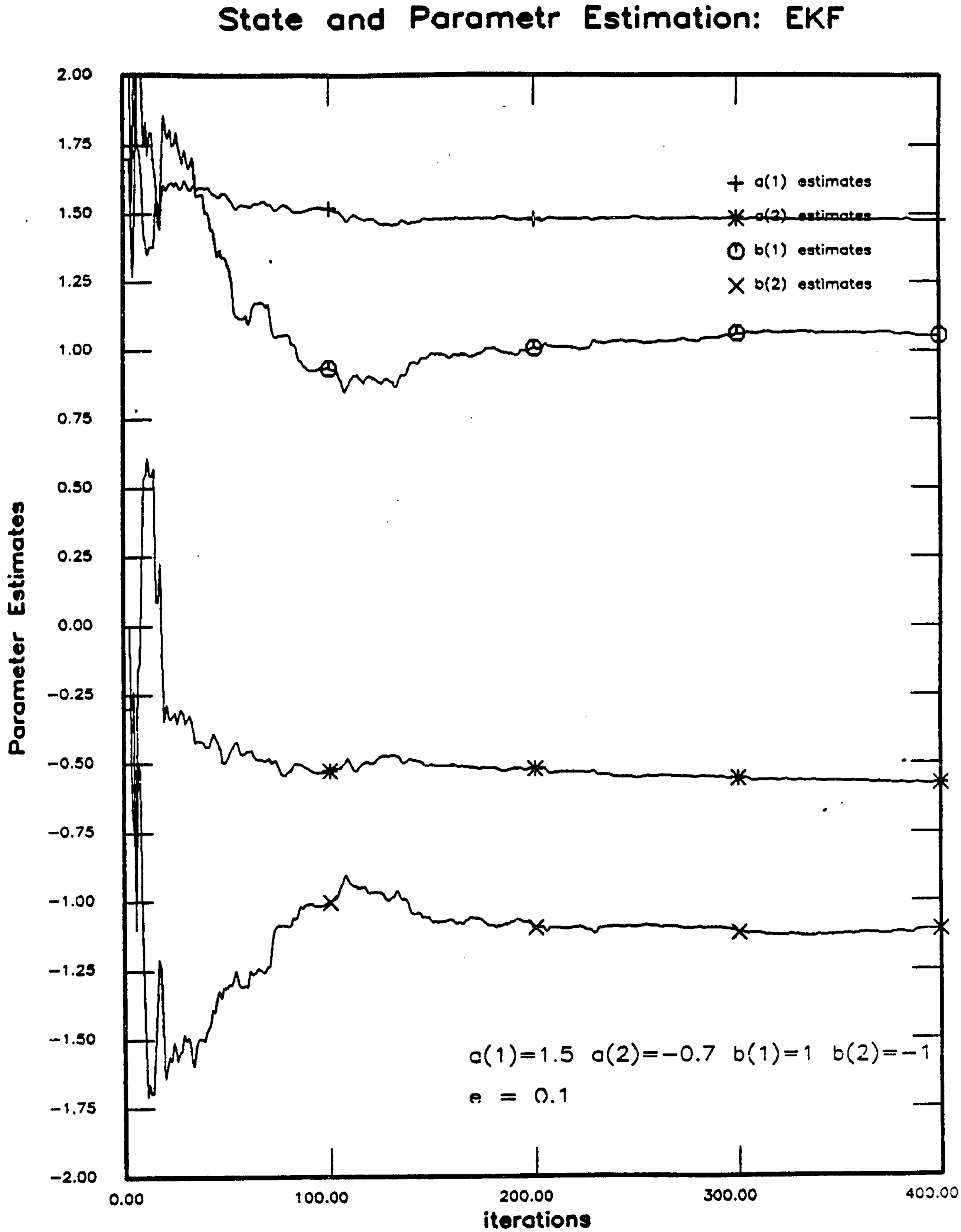


FIGURE 13a

4/ 4/87
12.59.15

TIM.KALMAN]PLT.EXE;5

POINTS OUTSIDE PLOT BOUNDARY: 14

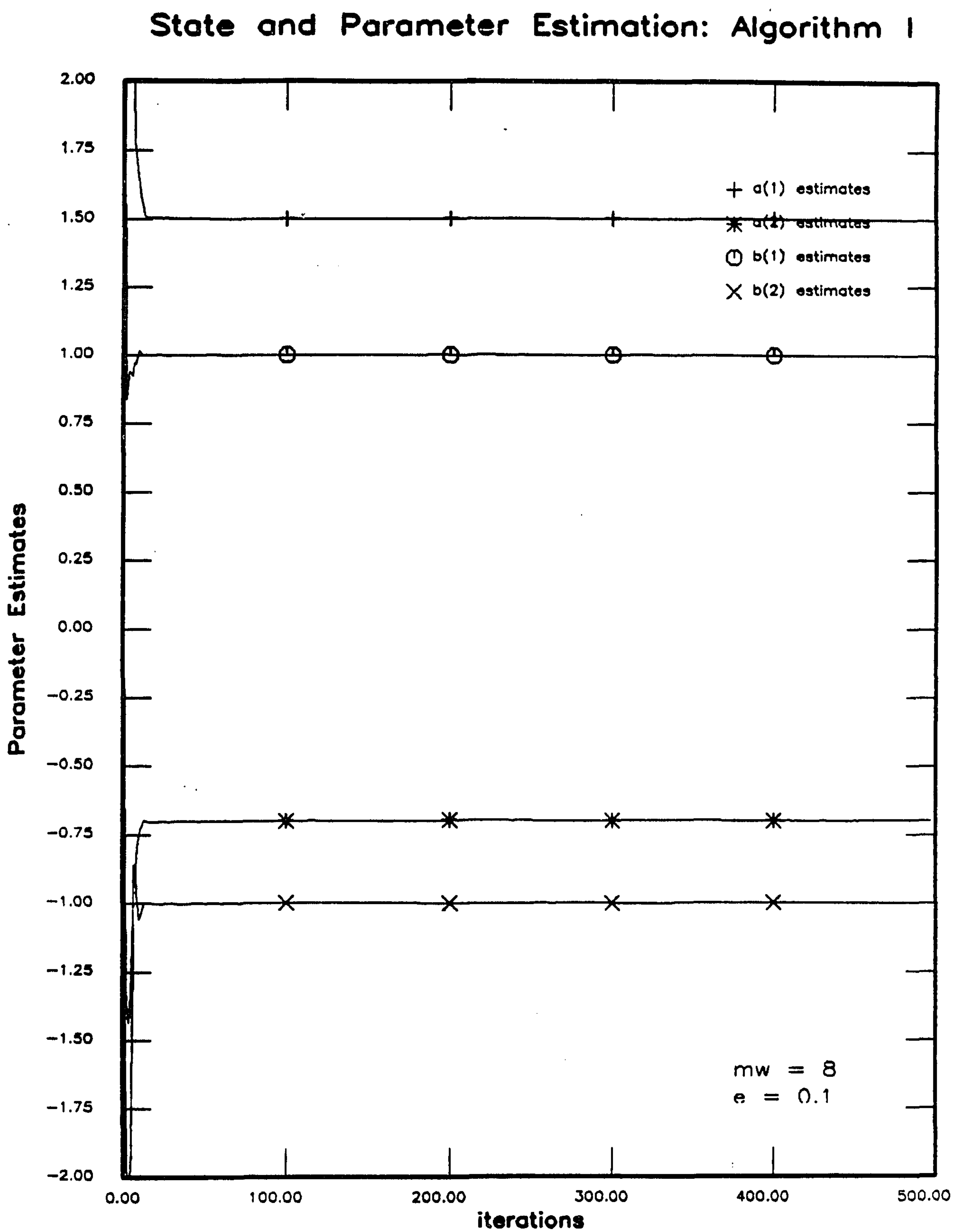


FIGURE 13b

4/ 4/87
12:11:46

[TIM.ESTM]PLT.EXE;35

POINTS OUTSIDE PLOT BOUNDARY: 8 · 5

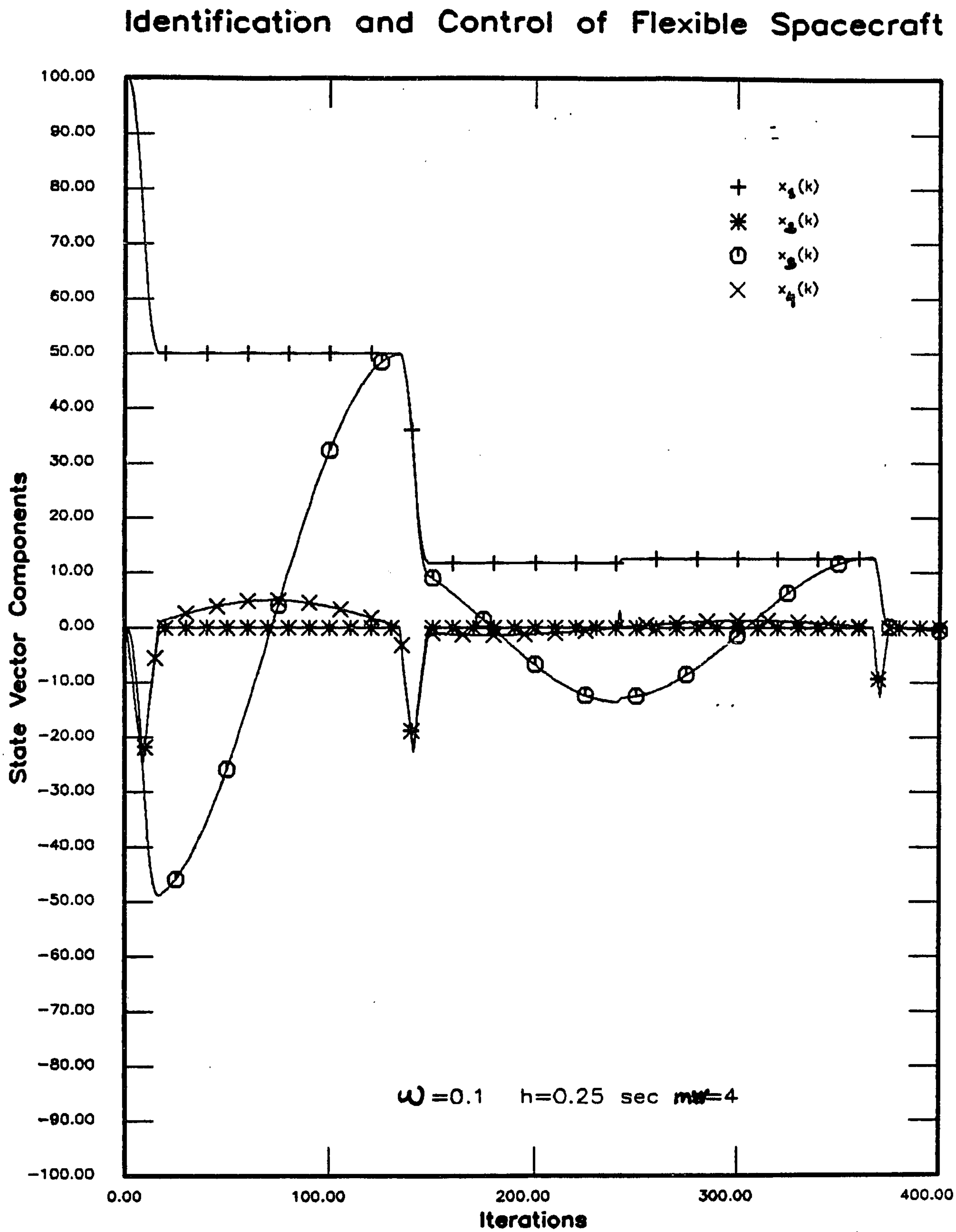


FIGURE 14

5/ 1/87
14:23:40

[TIM.SPACE]PLTS1.EXE;6

POINTS OUTSIDE PLOT BOUNDARY: 5

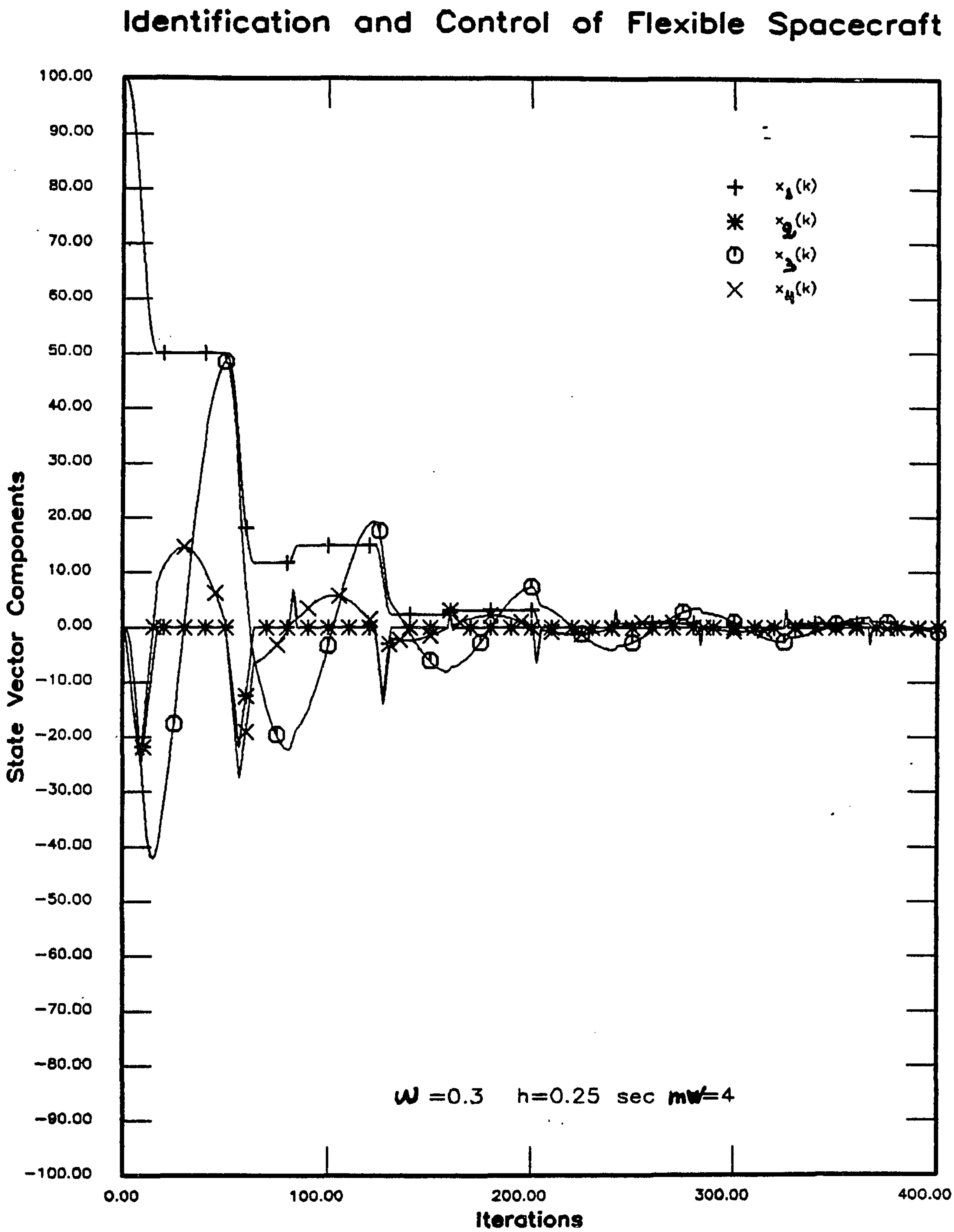


FIGURE 15

5/ 1/87
14:16:37

[TIM.SPACE]PLTS1.EXE;5

POINTS OUTSIDE PLOT BOUNDARY: 5

Identification and Control of Flexible Spacecraft

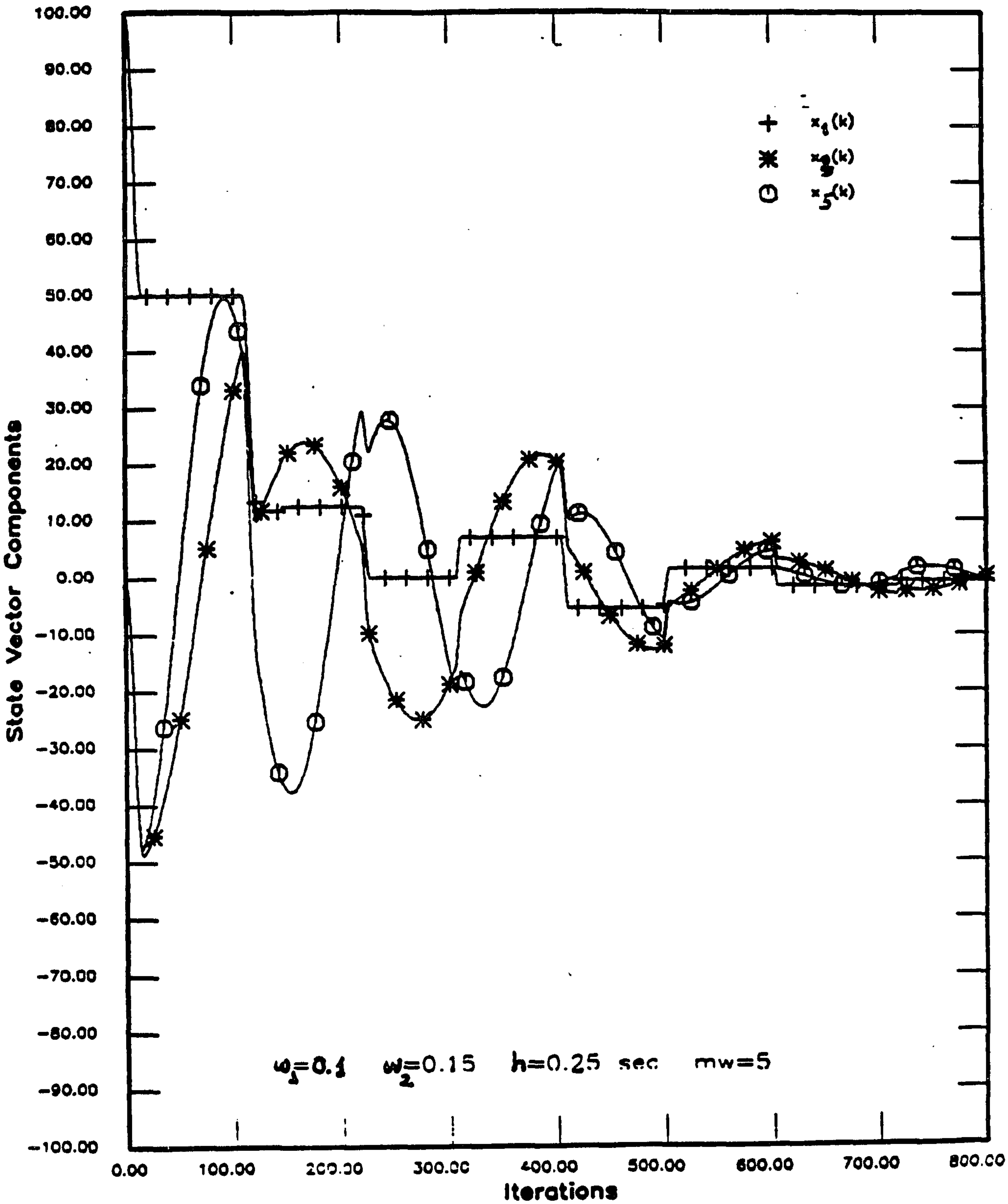


FIGURE 16a

2/13/87
16:48:39

[TIM.SPACE]PLTS.EXE;37

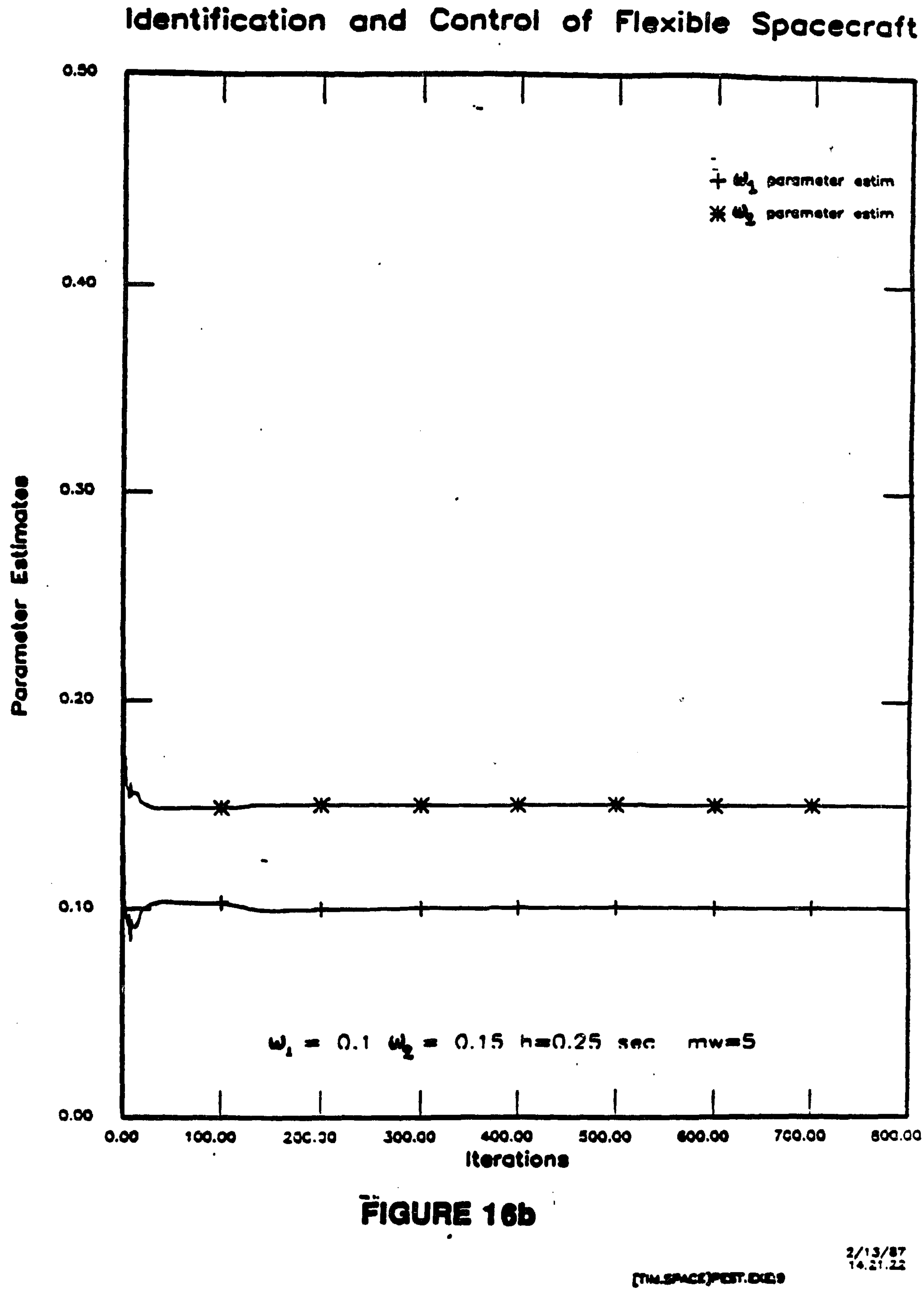

Identification and Control of Flexible Spacecraft
Parameter Estimates
0.50
0.40
0.30
0.20
0.10
0.00
0.00
100.00
200.00
300.00
400.00
500.00
600.00
700.00
800.00
Iterations
+ ω1 parameter estim
✳ ω2 parameter estim
ω1 = 0.1 ω2 = 0.15 h=0.25 sec mw=5
2/13/87
14.21.22
[TIM.SPACE]PEST.EXE;9

FIGURE 16b

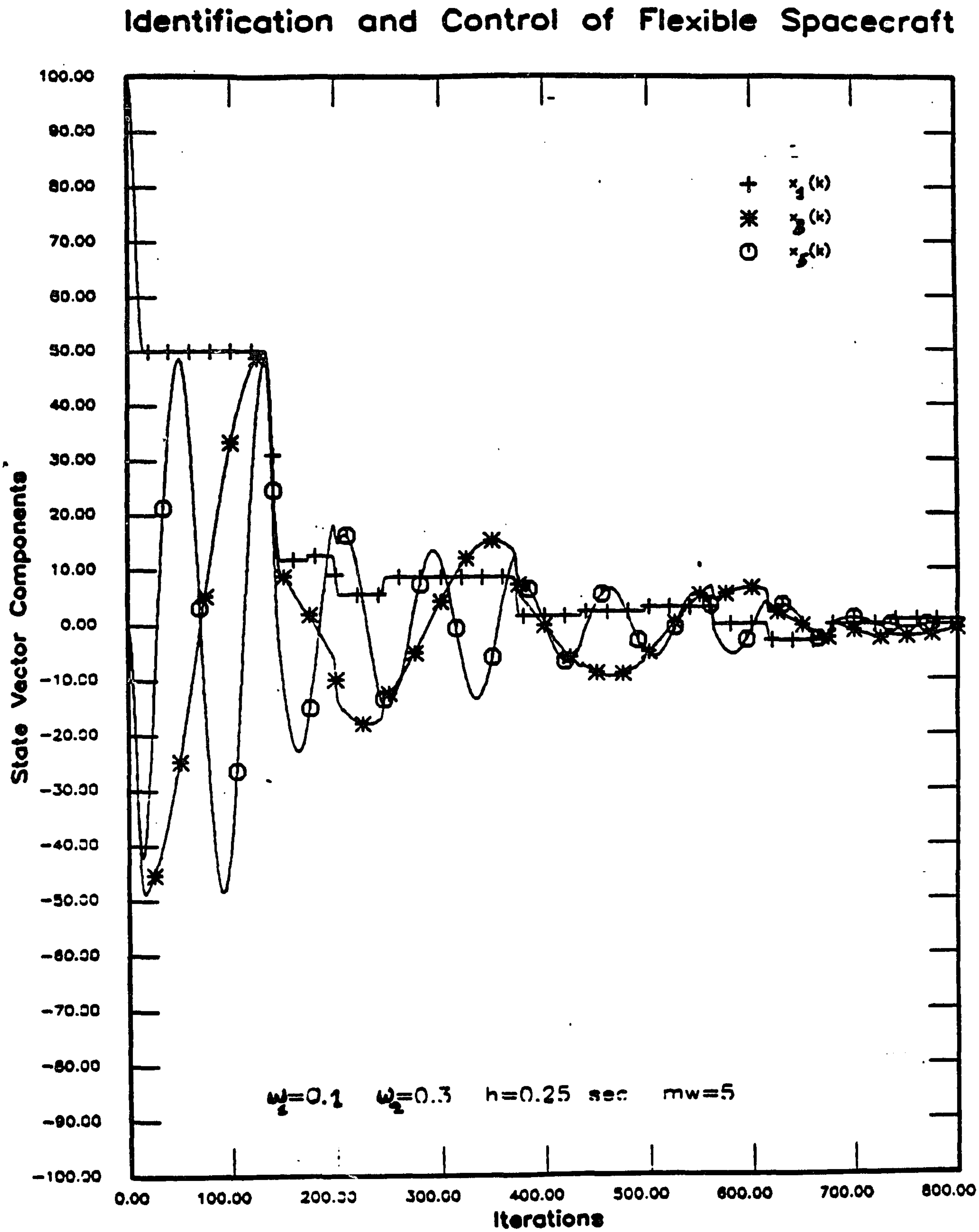


FIGURE 17a

2/13/87
12:44:37

[TIM.SPACE]PLTS.EXE;38

Identification and Control of Flexible Spacecraft

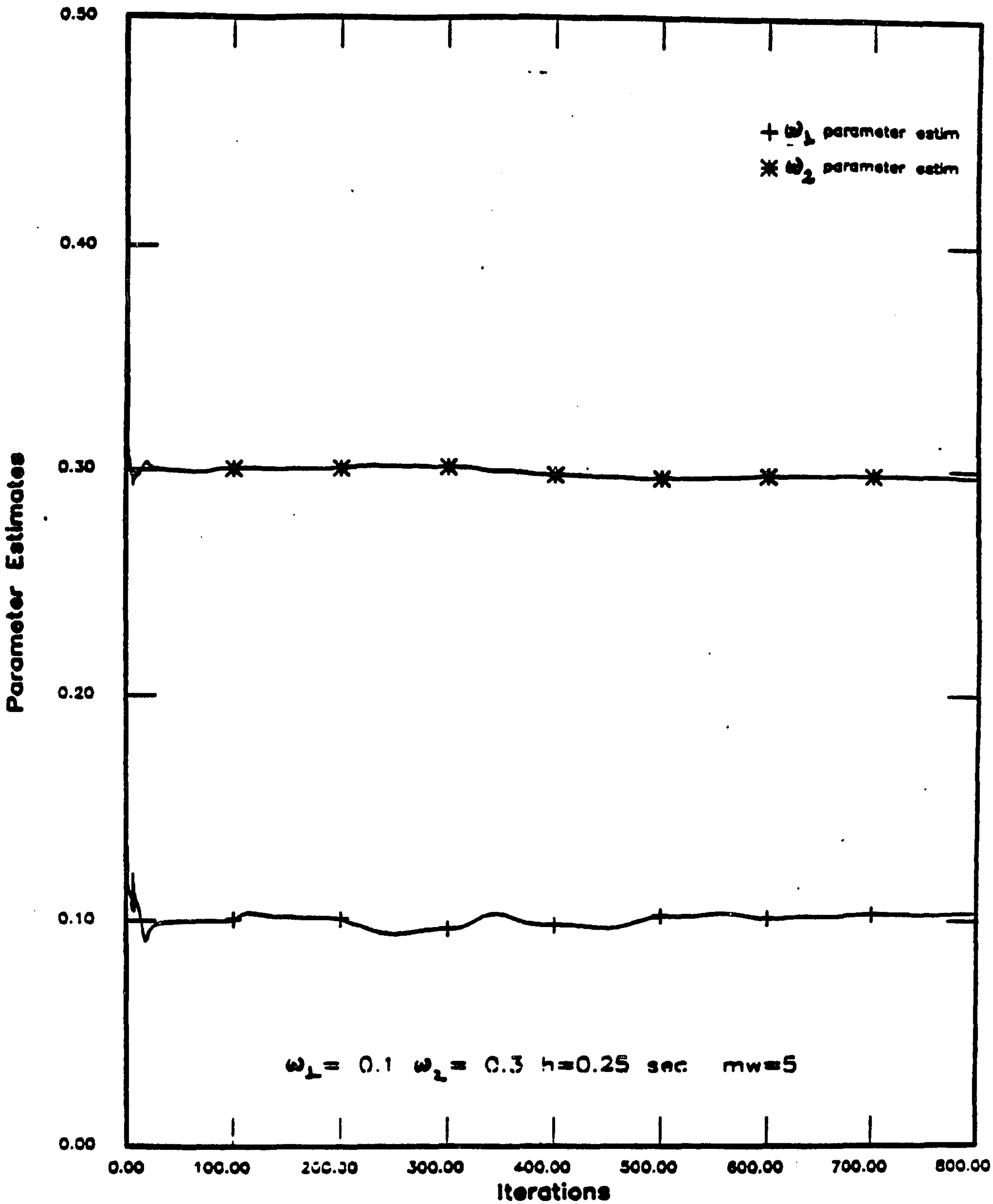


FIGURE 17b

2/13/87
12:56:01

[TIM.SPACE]PEST.DIZ;6

References

[1] Duncan, D. B. and Horn, S. D., "Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis," J. Amer. Statist. Assoc., 67(1972) pp. 815-821.

[2] Paige, C. C. and Saunders, M. A., "Least Squares Estimation of Discrete Linear Dynamic Systems Using Orthogonal Transformations," SIAM J. Numer. Anal., Vol. 14, No. 2, April 1977, pp. 180-193.

[3] Golub, G. H. and Pereyra, V., "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.

[4] Kaufman, L., "A Variable Projection Method for Solving Separable Nonlinear Least Squares Problems," BIT, 15, (1975), pp. 49-57.

[5] Hanson, R. J. and Lawson, C. L., "Extensions and Applications of the Householder Algorithm for Solving Linear Least Squares Problems," Math. Comp., Vol. 23, No. 108, pp. 787-812.

[6] Moore, J. B. and Ledwich, G., "Multivariable Adaptive Parameter and State Estimators with Convergence Analysis," J. Austr. Math. Soc., Vol. 21, Series B, 1979, pp. 176-197.

[7] Lawson, C. L., Hanson, R. J., Solving Least Squares Problems, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1974.

[8] Fogel, E., "A Fundamental Approach to the Convergence Analysis of Least-Squares Algorithms," IEEE Trans., Vol. AC-26, No. 3, June 1981, pp. 646-655.

[9] Marquardt, D., "An Algorithm for Least Squares Estimation of Nonlinear Parameters," J. Soc. Indust. Appl. Math. 11, June 1963.

[10] Osborne, M. R., "A Class of Nonlinear Regression Problems," Data Representation, pp. 94-101, R. S. Anderssen and M. R. Osborne eds., Queensland Univ. Press, 1970.

[11] Anderson, B. D. O. and Moore, J. B., Optimal Filtering, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1979.

[12] Eulrich, B. J., et al., "Partitioning Identification Algorithms," IEEE Trans. Vol. AC-25, No. 3, June 1980, pp. 521-528.

[13] Ljung, L., "Asymptotic Behavior of the Extended Kalman Filter as a Parameter Estimator for Linear Systems," IEEE Trans. Vol. AC-24, No. 1, February 1979.

[14] Jazwinski, A. H., Stochastic Processes and Filtering Theory, New York: Academic Press, 1970.

[15] Kopp, R. E. and Orford, R. J., "Linear Regression Applied to System Identification for Adaptive Control Systems," AIAA J. Vol. 1, No. 10, pp. 2300-2306, 1963.

[16] Nelson, L. W. and Stear, E., "The Simultaneous On-line Estimation of Parameters and States in Linear Systems," IEEE Trans. Vol. AC-21, pp. 94-98, February 1976.

[17] Bar-Shalom, Y., "Optimal Simultaneous State Estimation and Parameter Identification in Linear Discrete-Time Systems," IEEE Trans. Automatic Control Vol. AC-17 p. 308-320, 1972.

[18] Warren, A. W., "Partitioned State Algorithms for Recursive System Identification," Proceedings American Control Conference 1983, pp. 742-747.

[19] Halme, A. and Selkainaho, J., "Recursive Simultaneous State and Parameter Estimation by A Robust Filtering Method and Orthogonal Polynomials," Proceedings IEEE Conference on Decision and Control 1982.

[20] Irwin, G. W. and Roberts, A. P., "The Luenberger Canonical Form in the State/Parameter Estimation of Linear Systems," Int. J. Control, Vol. 23, No. 6, pp. 851-864, 1976.

[21] Leonedes, C. T. and J. O. Pearson, "Kalman Filtering of Systems With Parameter Uncertainties A Survey," Int. J. Control, Vol. 17, 1973, pp. 785-792.

[22] Thau, F. E., Eliazov, T., and Montgomery, R. C., "Least-Squares Sequential Parameter and State Estimation for Large Space Structures," presented at the 1982 American Automatic Control Conference.

[23] Farison, J. B., Graham, R. E., Shelton, R. C., "Identification and Control of Linear Discrete Systems," IEEE Trans. Automatic Control, Vol. AC-12, pp. 438-442, August 1967.

[24] Krogh, F. T., "Efficient Implementation of a Variable Projection Algorithm for Nonlinear Least Squares Problems," Comm. ACM 17, March 1974.

[25] Golub, G. H., "Matrix Decompositions and Statistical Calculations," Statistical Computation, Roy C. Milton and John A. Nelder, eds., Academic Press, New York, 1969.

[26] Peters, G. and Wilkinson, J. H., "Generalized Inverse of Matrices and Its Applications," Comp. J. 13.3, August 1970, pp. 309-316.

[27] Leonedes, C. T. ed., "Control and Dynamic Systems Advances in Theory and Applications," Vol. 20: Nonlinear and Kalman filtering techniques.

[28] Rune, A., "Accelerated Gauss-Newton Algorithms for Nonlinear Least Squares Problems," BIT 19, 1979, pp. 356-367.

[29] Ramsin, H. and Wedin, P. A., "A Comparison of Some Algorithms for the Nonlinear Least Squares Problem," BIT 17, 1977, pp. 72-90.

[30] Bard, Y., Nonlinear Parameter Estimation, Academic Press, New York, 1974.

[31] Wedin, P. A., "On The Gauss-Newton Method for the Nonlinear Least Squares Problem," Rep. 24, The Swedish Institute of Applied Mathematics,

Stockholm, 1974.

[32] Wacker, H., Continuation Methods, Academic Press, 1978.

[33] Stewart, G. W., Introduction to Matrix Computations, Academic Press, New York, 1973.

[34] Sage, A. P. and Melsa, J. L., System Identification, Academic Press, New York, 1971.

[35] Rheinboldt, W. C., "An Adaptive Continuation Process for Solving Systems of Nonlinear Equations," TR-393 Computer Science, University of Maryland, 1975.

[36] Ortega, J. M. and Rheinboldt, W. C., "Iterative Solution of Nonlinear Equations in Several Variables," Academic Press, New York and London, 1970.

[37] Thau, F. E., Montgomery, R. C., and Horner, G. C.: On-line Structural Parameter Identification. Proceedings of the 1981 AIAA Guidance and Control Conference, August 1981, pp. 530-539.

[38] Beloglazov, I. N., "Optimal Joint Estimation and Identification in Discrete Linear Systems," Dokl. Akad. Nauk SSSR 273, December 1983.

[39] Halme, A., Selkainaho, J. and Soininen, J. "Adaptive Control With Nonlinear Filtering," Automatica Vol. 21, No. 4, pp. 453-463, 1985.

[40] Linik, Y. V., Method of Least Squares and Principles of the Theory of Observations, Pergamon Press, New York, London, 1961.

[41] Bertsekas, D. P., Notes on Nonlinear Programming and Discrete-Time Optimal Control, LIDS-R-919 MIT, Cambridge, Massachusetts, 1979.

[42] Ostrowski, A. M., Solution of Equations in Euclidean and Banach Spaces, Academic Press, New York and London, 1973.

[43] Whetstone, W. T., SPAR Structural Analysis System Reference Manual, NASA.

[44] Ljung, L., "Convergence Analysis of Parametric Identification Methods," IEEE Transactions on Automatic Control, Vol. AC-23, pp. 770-783.

[45] Ljung, L., Soderstrom, T., "Theory and Practices of Recursive Identification," The MIT Press, Cambridge, Massachusetts, 1983.

[46] Dodds, S. J. and Williamson, S. E., "A Signal Switching Time Bang-Bang Attitude Control Law for Fine Pointing of Flexible Spacecraft," Int. J. Control, 1984, Vol. 40, No. 4, pp. 16-21.

[47] Halme, A., Selkainaho, J. and Soininen, J., "Implementation of Multivariable Self-Tunable PI-Type Controller in a Distributed Microcomputer System," Proceedings of the Symposium on Application of Multivariable Systems Theory, pp. 113-121, London, 1982.

[48] Lo, J. T. H., "Optimal Estimation for the Satellite Attitude Using Star Tracker Measurements," Automatica, Vol. 22, pp. 477-482, 1986.

[49] Schmidt, S. F., "Compensation for Modeling Errors in Orbit Determination Problems," Analytical Mechanics Assoc., Westbury, N.Y., Report 67-16, 1967.

[50] Neal, S. R., "Linear Estimation in the Presence of Errors in Assumed Plant Dynamics," IEEE Trans. Automatic Control, Vol. AC-12, pp. 582-594, October 1967.

[51] Luders, G. and Narendra, K. S., "Stable Adaptive Schemes for State Estimation and Identification of Linear Systems," IEEE Transactions on Automatic Control, Vol. AC-19, pp. 841-847, 1974.

[52] Carol, R. L. and Lindorff, D. P., "An Adaptive Observer for Single-Input Single-Output Linear Systems," IEEE Transactions on Automatic Control, Vol. AC-18, pp. 428-434, 1973.

[53] Dodds, S. J., "Simultaneous Plant Parameter and State Estimation by a Successive States Method," IEE Control, pp. 452-457, 1986.