A BAYESIAN APPROACH
TO
SYSTEM IDENTIFICATION

by

Chun Loo

(July, 1966)

Submitted to the Department of Electrical Engineering in partial fulfilment of the requirements for the degree of Master of Science.

Department of Electrical Engineering
Faculty of Pure and Applied Science
University of Ottawa
Ottawa, Canada

1966

A BAYESIAN APPROACH TO SYSTEM IDENTIFICATION

by

Chun Loo

ABSTRACT

A Bayesian approach to system identification is considered in this thesis. The method uses all a priori knowledge of the plant under investigation. Indirectly, it is shown that the method unifies and extends the works of Levin [1] and Lindenlaub [2]. This thesis also shows that the limit on the identification time is greatly reduced when a priori knowledge is used.

Several examples of a second order plant were simulated on a digital computer to confirm the method. The results indicated that the effect of a priori knowledge used in the identification process is critical only in the first few iterations. Moreover, these examples showed that the method converges rapidly to a value which is very close to the true value of the plant.

ACKNOWLEDGEMENTS

The author wishes to express his sincere thanks to Professor Jaan Kruus for his guidance, suggestions and constant encouragement throughout this research.

The helpful discussions with Professor S. G. Shiva and his colleague Mr. M. Arozullah are gratefully acknowledged. This work was supported by N. R. C. grant, No. A-2580.

TABLE OF CONTENTS

Table of Contents -- Continued --

LIST OF FIGURES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Outline of Problem:

This investigation considers the problem of identifying a linear system from deterministic inputs and observed outputs over a finite interval of time, making use of all prior knowledge of the system. In order to develop a practical procedure, it is necessary to assume that there is some uncertainity in the observed outputs. This uncertainity could be due to random noise, measurement errors and internal system noise. With such an assumption, the problem of identification becomes an estimation problem which requires the estimation theory of mathematical statistics. In particular, Bayesian statistics will be used to estimate the system impulse response.

In this thesis, the procedure for identifying the system impulse response will be derived. Systems with only a single input and one output will be treated. With suitable modifications to notation, the procedure is applicable to multi-dimensional systems.

1.2 Previous Work:

Levin [1] has considered the optimum estimate of the impulse response of a linear system. His estimate can be obtained by putting the problem into a form for which least squares and Markov estimates are appropriate. These estimates, together with their sampling variances, are determined without any assumption regarding the form of the system input. The least squares estimate requires the solution of a matrix equation having the form of a discrete version of the Wiener-Hopf equation, whereas the Markov estimate requires the solution of a generalization of this equation.

Lindenlaub [2] has extended this analysis to include a time limit on the identification of system impulse response using a known input signal.

Hill and McMurtry [3] have developed a digital technique for the above method. Their technique uses a discrete interval binary test signal and is applicable to on line system identification.

All previous methods failed to make use of any prior knowledge of the plant. However, the work in this thesis represents an extension of Levin's method, making use of all prior knowledge of the plant.

1.3 Statement of the Problem :

It is desired to form an estimate of the discrete impulse response of a linear system corrupted by additive noise. This estimate uses all prior knowledge of the system. The system noise is assumed to be white and Gaussian.

1.4 Specification of Linear Systems:

The usual methods for specifying characteristics of a linear system are as follows:

(1) Differential equations

(2) Frequency response (for continuous system only)

(3) Transfer function (continuous or discrete)

(4) Impulse or Step Response (continuous or discrete)

(5) State-Space representation i.e. transistion matrix.

1.5 System Model:

The dynamic characteristics of the system are determined by analyzing an input x(t) to the system and the resulting output of this system, z(t). It is assumed that the input x(t) is known exactly, but z(t) is assumed to be corrupted by white Gaussian noise v(t). It is further assumed that v(t) is independent of x(t). Thus the only available data for analysis is at the output

$$y(t) = z(t) + v(t). \qquad (1.1)$$

The resulting model is shown in Fig. 1.1.

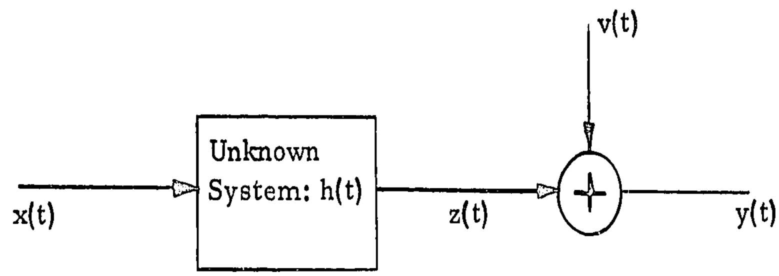


Fig. 1.1 Model for System Identification.

1.6 Examples of Engineering Systems with Unknown Parameters:

Numerous examples of engineering systems with unknown parameters exist in our scientific world. A few of the examples is given below:

(1) Industrial processes - e.g. distillation column and evaporator [4]

(2) Communication systems - e.g. transmission path and space probe [9].

(3) Adaptive control systems - e.g. automated flight director [5].

1.7 Outline of New Results:

A Bayesian solution to the problem of identifying system impulse response is derived in this thesis. This solution uses all prior knowledge of the system in combination with all experimental observations. The law of combining the prior and observed information depends on the confidence index of each i.e. it uses Bayes' Theorem. The results obtained with this procedure are far better than those using any method not taking into account the *a priori* knowledge of the system. From a practical point of view, there always exist a certain amount of prior knowledge in any system to be identified. In addition, the method developed in this thesis is very well suited for numerical computation using a digital computer. Moreover, this technique identifies the system impulse response sequentially making on-line identification possible. On-line identification is an essential part in any adaptive control system. Finally this procedure is easily implemented in practice.

CHAPTER II

STATISTICAL ESTIMATION METHODS RELATED TO SYSTEM IDENTIFICATION

In this chapter, we shall describe several statistical methods pertinent to system identification.

2.1 Linear Regression Method:

Consider n samples (observations) y(0), y(1) .... y(N-1). These samples can be represented as components of a column vector $\underline{Y}$. Now suppose there exist functions of the regressors x_{ik} which are completely known and unknown parameters h_k,

$k = 0, 1, 2, \ldots\ldots K$

called regression coefficients. For convenience, we define the following:

$$Y \triangleq \begin{bmatrix} y(K) \\ y(K+1) \\ \vdots \\ y(K+N) \end{bmatrix} \qquad h \triangleq \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(K) \end{bmatrix}$$

$$X \triangleq \begin{bmatrix} x(K) & x(K-1) & \ldots & x(0) \\ x(K+1) & x(K) & \ldots & x(1) \\ \vdots & \vdots & & \vdots \\ x(K+N) & x(K+N-1) & \ldots & x(N) \end{bmatrix}$$

Then $\underline{Y}$ may be expressed as

$$\underline{Y} = X\underline{h} + \underline{V} \qquad (2.1)$$

where $\underline{V}$ is a column vector

$$\underline{V} = \begin{bmatrix} v(0) \\ v(1) \\ \vdots \\ v(N-1) \end{bmatrix}$$

It can be shown [10] that a linear unbiased estimate* can be obtained by minimizing a quadratic loss function of the following form.

$$\ell = (\underline{Y} - X\underline{h})' \, W(\underline{Y} - X\underline{h}) \tag{2.2}$$

where W is a symmetric non-negative weighting matrix and X'X is non-singular and the prime indicate transpose. Then ℓ may be expressed as

$$\begin{aligned}\ell &= Y' \, W\,[I - X(X'WX)^{-1}(WX)'\,]\,\underline{Y} \\ &+ [(X'WX)^{\frac{1}{2}}\underline{h} - (X'WX)^{-\frac{1}{2}}(WX)'\,\underline{Y}\,] \,\circ \\ &\quad [(X'WX)^{\frac{1}{2}}\underline{h} - (X'WX)^{-\frac{1}{2}}(WX)'\,\underline{Y}\,]\end{aligned} \tag{2.3}$$

where

$$(X'WX)^{\frac{1}{2}}(X'WX)^{\frac{1}{2}} = X'WX$$

Since the second term is the sum of squares and must be non-negative, ℓ is minimized when

$$(X'WX)^{\frac{1}{2}}\underline{h} - (X'WX)^{-\frac{1}{2}}(WX)'\,\underline{Y} = 0.$$

This is satisfied when

$$\underline{h} = \hat{\underline{h}} = (X'WX)^{-1}(WX)'\,\underline{Y} \;. \tag{2.4}$$

The covariance of $\underline{h}$ is given by

$$\text{Cov}\;\hat{\underline{h}} = X'WX)^{-1}(WX)'\,R(WX)\,(X'WX)^{-1} \tag{2.5}$$

where R is the covariance of the vector $\underline{V}$.

* Linear unbias estimate is a minimum variance estimate.

2.2 Least Squares Estimate:

When the weighting matrix W defined previously becomes an identity matrix i.e. W = I, it can be shown [10] that the least squares estimate is obtained by minimizing the following quadratic loss function

$$\ell = (\underline{Y} - X\underline{h})' (\underline{Y} - X\underline{h}) .$$

The least squares estimate is now

$$\hat{\underline{h}} = (X'X)^{-1} X'\underline{Y} \tag{2.6}$$

and the covariance of $\hat{\underline{h}}$ is given by

$$\text{Cov } \hat{\underline{h}} = (X'X)^{-1} X'RX(X'X)^{-1}. \tag{2.7}$$

2.3 Markov Estimate:

When the weighting matrix W is made equal to the inverse of the covariance of $\underline{V}$, i.e. $W = R^{-1}$, assuming R is non-singular, the Markov estimate [10] is obtained by minimizing the following loss function:

$$\ell = (\underline{Y} - X\underline{h})' R^{-1} (\underline{Y} - X\underline{h}).$$

The estimate is

$$\hat{\underline{h}} = (X'R^{-1}X)^{-1} (R^{-1}X)' \underline{Y} \tag{2.8}$$

with covariance given by

$$\text{Cov } \hat{h} = (X'R^{-1}X)^{-1}. \tag{2.9}$$

2.4 Maximum Likelihood Estimate:

The method of maximum likelihood is, from a theoretical point of view, the most important general method of estimation. This method was originally developed by Fisher who used it in point estimation of parameters. The basic concept of the method is quite simple and can be found in most standard texts of statistics and probability e.g. [12]. First, we define a likelihood function, which is usually the conditional probability $P[{}^{Y}/h]$ relating the parameter $\underline{h}$ and the measurement $\underline{Y}$.

Consider a random process which has a multi-variate normal distribution. Then the conditional probability density function of the random variable $\underline{Y}$ is given by

$$P[{}^{y}/h] = \frac{1}{2\pi^{h/2} \, \|R\|^{\frac{1}{2}}} \exp - \frac{1}{2} Q(y)$$

where $Q(y) = (\underline{Y} - X\underline{h})' R^{-1} (\underline{Y} - X\underline{h})$

The likelihood function of $\underline{Y}$ is defined as

$$L = \log \frac{1}{2\pi^{n/2} |R|^{\frac{1}{2}}} \exp - \frac{1}{2} Q(y) \qquad (2.10)$$

To find the best estimate of $\underline{h}$ in the maximum likelihood sense, we simply maximize the likelihood function with respect to $\underline{h}$. This is equivalent to minimizing the quadratic function $Q(y)$. This is the same procedure used for finding the Markov estimate. Thus for a random process which has normal distributions, the maximum likelihood estimate is identical to the Markov estimate. Therefore, the maximum likelihood estimate $\hat{\underline{h}}$ of $\underline{h}$ is

$$\hat{\underline{h}} = (X'R^{-1}X)^{-1} (R^{-1} X)' \underline{Y} \qquad (2.11)$$

with covariance given by

$$\text{Cov } \hat{\underline{h}} = (X'RX)^{-1} \qquad (2.12)$$

2.5 Bayes Estimate:

Suppose an *a priori* distribution function is available not only for measurement errors but also for the values of the unknown parameter $\underline{h}$, then it is possible to find an estimate $\underline{h}^*$ which minimizes the expected loss. Such an estimate is called a Bayes estimate [11].

In general, the Bayes estimate depends upon the form of both the loss function and the a priori distributions of parameters and measurement noise. In particular, we shall consider only the special case of a positive semi-definite quadratic loss function. Then the Bayes estimate is just the mean of $\underline{h}$ conditioned on the observation $\underline{Y}$. This is true regardless of the distribution of parameters or measurement [11]. In order to compute the Bayes estimate, it is necessary to determine the conditional probability density function. This can be obtained from Bayes Theorem

$$P[{}^{h}/y] = \frac{P[{}^{y}/h]\,P[h]}{P[y]} \tag{2.13}$$

where

$P[{}^{h}/y]$ is the conditional (a posteriori) probability density function.,

$P[h]$ is the a priori probability density,

and $P[{}^{y}/h]$ is the likelihood function.

The term $P[y]$ in the denominator is a normalizing factor.

For a Gaussian process, Parzen [7] has obtained the following results; Consider the probability density (likelihood) function of the form

$$P[{}^{y}/h] = \exp\left[hU - \frac{1}{2}h^2 G\right]$$

and the a priori probability density of h is given by

$$P[h] = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2\sigma^2}(h - h_p)\right]$$

with known mean h_p and known variance σ^2 . The a posteriori probability density of h becomes

$$P[h/y] = \frac{1}{2\pi\gamma} \exp\left[- \frac{1}{2\gamma}\left(h - \frac{\alpha}{\gamma}\right)^2 \right] \quad (2.14)$$

where

$$\alpha = U + \sigma^2 h_p$$

$$\gamma = G + \sigma^{-2}$$

Thus the conditional density is normal with mean $\frac{\alpha}{\gamma}$ and variance $\frac{1}{\gamma}$.
The Bayes estimate, h^*, of h is given by the conditional mean

$$h^* = \frac{U + \sigma^{-2} h_p}{G + \sigma^{-2}} \quad (2.15)$$

with mean square estimation error

$$E[|h - h^*|^2] = [G + \sigma^{-2}]^{-1} \quad (2.16)$$

Now, suppose we are very uncertain about the possible range of values of the parameter h. This corresponds to assuming that the prior variance is very large i.e

$$\sigma^2 \rightarrow \infty$$

$$h^* \rightarrow G^{-1} U$$

and

$$E[|h - h^*|^2] \rightarrow G^{-1}$$

If the variance σ^2 of the parameter h is very large, the Bayes estimate is approximately given by the estimate

$$\hat{h} = G^{-1} U$$

It can be shown [7] that $\hat{h}$ is the maximum likelihood estimate which is also minimum variance linear unbiased estimate .

To put the Bayes estimate into a more suitable form which is useful for application, we define the following:

$$I_p = \{E[|h - h_p|^2]\}^{-1} = \sigma^{-2} \qquad (2.17)$$

$$\hat{I} = \{E[|h - \hat{h}|^2]\}^{-1} = G \qquad (2.18)$$

$$I^* = \{E[|h - h^*|^2]\}^{-1} \qquad (2.19)$$

where I_p, $\hat{I}$ and I^* represents the confidence indices attached respectively to the prior estimate h_p, the maximum likelihood estimate $\hat{h}$ and the Bayes estimate h^*. Now the Bayes estimate can be rewritten as

$$h^* = (I^*)^{-1} (\hat{I}\hat{h} + I_p h_p) \qquad (2.20)$$

where

$$I^* = \hat{I} + I_p.$$

Thus the Bayes estimate is a weighted average of the maximum likelihood estimate and of the prior estimate. The weight is proportional to the confidence index in each estimate.

The above derivation could easily be extended to a multi-parameter $\underline{h}$. Then the maximum likelihood estimate $\hat{\underline{h}}$ of $\underline{h}$ is given by

$$\hat{\underline{h}} = G^{-1} \underline{U} \qquad (2.21)$$

with mean square estimation error matrix

$$E[|(\underline{h} - \hat{\underline{h}})'(\underline{h} - \hat{\underline{h}})|] = G^{-1} \qquad (2.22)$$

and $\hat{I} = G.$

If the vector parameter $\underline{h}$ possess a prior distribution which is multivariate Gaussian with mean $E[\underline{h}] = \underline{h}_p$ and the prior confidence index

$$I_p^{-1} = E[|(\underline{h} - \underline{h}_p)'(\underline{h} - \underline{h}_p)|], \qquad (2.23)$$

The Bayes estimate $\underline{h}^*$ becomes

$$\underline{h}^* = (\hat{I} + I_p)^{-1} (\hat{I}\hat{\underline{h}} + I_p \underline{h}_p) \quad (2.24)$$

with mean square estimation error

$$E\left[\left|(\underline{h} - \underline{h}^*)'(\underline{h} - \underline{h}^*)\right|\right] = (I^*)^{-1} \quad (2.25)$$

where

$$I^* = \hat{I} + I_p$$

2.6 Equivalent Properties of Statistical Estimation Methods:

1. If the system noise is white i.e has covariance matrix $\sigma_v^2 I$

   a) The least squares estimate is a minimum variance unbiased estimate of $\underline{h}$ [1]

   b) Markov and least squares estimates are equivalent [1].

2. If the system noise is white and Gaussian, the least squares and maximum likelihood estimates are equivalent, [1]

3. If the system noise is white and Gaussian and the prior variance is very large $\sigma^2 \to \infty$, the Bayes and maximum likelihood estimates are coincident [7].

CHAPTER III

IDENTIFICATION OF SYSTEM IMPULSE RESPONSE

In this chapter, a statistical estimate is derived in terms of a sampled-data system with sampling interval T.

3.1 Least Squares Estimate of System Impulse Response:

Using the model assumed in Section 1.5, we sample the input and the output. Our model is as shown in Fig. 3.1

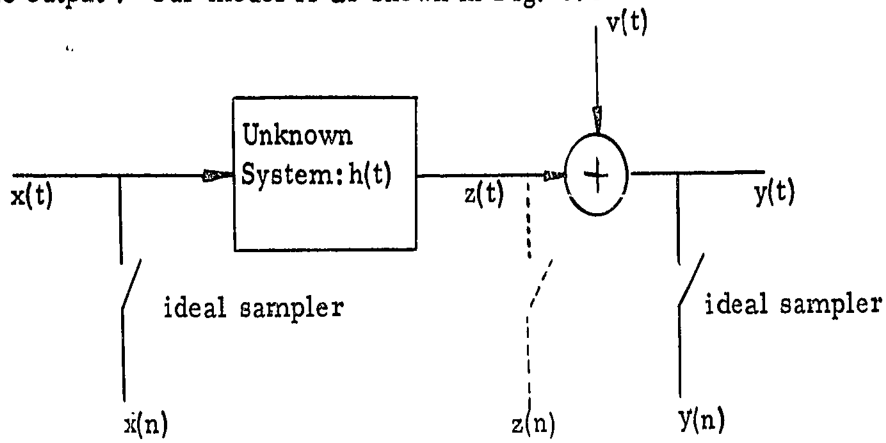


Fig. 3.1 System Model for Least Squares Estimate.

The observed output can be represented by a sampled-data version of the convolution integral

$$z(n) = \sum_{t=o}^{\infty} x(n-k)\, h(k)\, T \tag{3.1}$$

and

$$y(n) = z(n) + v(n) \tag{3.2}$$

where

$x(n)$ is the sampled input

$y(n)$ is the sampled output

$v(n)$ is a Gaussian white noise sample

and $\underline{h}$ is the sampled system impulse response.

First, we can represent the impulse response h(t) to a good approximation by the following vector

$$\underline{h} = \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(K) \end{bmatrix} \tag{3.3}$$

where K is chosen so that KT is greater than the significant durantion of h(t). An estimate of $\underline{h}$ is computed from a sequence of N+K sampled inputs and a sequence of N sampled outputs. The delay caused by the system being identified makes it necessary to record N + K sample values of the input x(t).

For convenience, we introduce the following notation:

$$Z = \begin{bmatrix} z(K) \\ z(K+1) \\ \vdots \\ z(K+N) \end{bmatrix} \quad \underline{V} = \begin{bmatrix} v(K) \\ v(K+1) \\ \vdots \\ v(K+N) \end{bmatrix} \quad \underline{Y} = \begin{bmatrix} y(K) \\ y(K+1) \\ \vdots \\ y(K+N) \end{bmatrix} \tag{3.4}$$

and the matrix

$$\underline{X} = \begin{bmatrix} x(K) & x(K-1) & \cdots & x(0) \\ x(K+1) & x(K) & \cdots & x(1) \\ \vdots & & & \\ x(K+N) & x(K+N-1) & \cdots & x(N) \end{bmatrix} \tag{3.5}$$

With this notation, we can replace the sampled-data version of the convolution integral

$$z(n) = \sum_{k=o}^{K} x(n-k)\, h(k)\ T \tag{3.6}$$

First, we can represent the impulse response h(t) to a good approximation by the following vector

$$\underline{h} = \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(K) \end{bmatrix} \tag{3.3}$$

where K is chosen so that KT is greater than the significant durantion of h(t). An estimate of $\underline{h}$ is computed from a sequence of N+K sampled inputs and a sequence of N sampled outputs. The delay caused by the system being identified makes it necessary to record N + K sample values of the input x(t).

For convenience, we introduce the following notation:

$$Z = \begin{bmatrix} z(K) \\ z(K+1) \\ \vdots \\ z(K+N) \end{bmatrix} \quad \underline{V} = \begin{bmatrix} v(K) \\ v(K+1) \\ \vdots \\ v(K+N) \end{bmatrix} \quad \underline{Y} = \begin{bmatrix} y(K) \\ y(K+1) \\ \vdots \\ y(K+N) \end{bmatrix} \tag{3.4}$$

and the matrix

$$\underline{X} = \begin{bmatrix} x(K) & x(K-1) & \cdots & x(0) \\ x(K+1) & x(K) & \cdots & x(1) \\ \vdots & & & \\ x(K+N) & x(K+N-1) & \cdots & x(N) \end{bmatrix} \tag{3.5}$$

With this notation, we can replace the sampled-data version of the convolution integral

$$z(n) = \sum_{k=o}^{K} x(n-k)\, h(k)\ T \tag{3.6}$$

by the matrix equation

$$\underline{Z} = X \underline{h} T \tag{3.7}$$

Due to external disturbance we can not obtain $\underline{Z}$ directly. The best we can do is to approximate $\underline{Z}$ from the observed output $\underline{Y}$.

The method of least squares is used to estimate the impulse response. This method chooses that value of $\underline{h}$ which will minimize the sum of squared deviations.

$$[\underline{Y} - X \underline{h} T]' [\underline{Y} - X \underline{h} T] = \text{minimum}$$

Levin [1] has shown that a least square estimate of the impulse response is given by a set of so-called normal equations

$$X' X \underline{h} T = X' Y \tag{3.8}$$

One can put this set of equation into a more familiar form by defining pseudo correlation function in terms of the sampled input and output sequences;

1. Pseudo auto-correlation function

$$\hat{\psi}_{xx}(i) = \frac{1}{N+1} \sum_{n=o}^{N} x(n)\, x(n+i) \tag{3.9}$$

2. Pseudo cross-correlation function

$$\hat{\psi}_{xy}(i) = \frac{1}{N+1} \sum_{n=o}^{N} x(n)\, y(n+i) \tag{3.10}$$

These pseudo correlation functions are not statistical parameters but are directly calculated from a set of input and output sequences. However, if x(t) comes from a stationary random process then, in the limit as $N \rightarrow \infty$ $\hat{\psi}_{xx}$ and $\hat{\psi}_{xy}$ become true correlation functions

$$\lim_{N \rightarrow \infty} \hat{\psi}_{xx}(i) = \psi_{xx}(i) \tag{3.11}$$

by the matrix equation

$$\underline{Z} = X \underline{h} T \tag{3.7}$$

Due to external disturbance we can not obtain $\underline{Z}$ directly. The best we can do is to approximate $\underline{Z}$ from the observed output $\underline{Y}$.

The method of least squares is used to estimate the impulse response. This method chooses that value of $\underline{h}$ which will minimize the sum of squared deviations.

$$[\underline{Y} - X \underline{h} T]' [\underline{Y} - X \underline{h} T] = \text{minimum}$$

Levin [1] has shown that a least square estimate of the impulse response is given by a set of so-called normal equations

$$X' X \underline{h} T = X' Y \tag{3.8}$$

One can put this set of equation into a more familiar form by defining pseudo correlation function in terms of the sampled input and output sequences;

1. Pseudo auto-correlation function

$$\hat{\psi}_{xx}(i) = \frac{1}{N+1} \sum_{n=o}^{N} x(n)\, x(n+i) \tag{3.9}$$

2. Pseudo cross-correlation function

$$\hat{\psi}_{xy}(i) = \frac{1}{N+1} \sum_{n=o}^{N} x(n)\, y(n+i) \tag{3.10}$$

These pseudo correlation functions are not statistical parameters but are directly calculated from a set of input and output sequences. However, if x(t) comes from a stationary random process then, in the limit as $N \to \infty$ $\hat{\psi}_{xx}$ and $\hat{\psi}_{xy}$ become true correlation functions

$$\lim_{N \to \infty} \hat{\psi}_{xx}(i) = \psi_{xx}(i) \tag{3.11}$$

and

$$\lim_{N \to \infty} \hat{\psi}_{xy}(i) = \psi_{xy}(i) \qquad (3.12)$$

It is seen that the i-jth element of the matrix X'X is $\psi_{xx}(i-j)$

$$X'X = \begin{bmatrix} \sum_{K}^{K+N} x(n)^2 & \sum_{K}^{K+N} x(n)Xn+1) \;.\;.\;.\;.\;.\;. & \sum_{K}^{K+N} x(n)\, x(n+K) \\ \sum_{K-1}^{K+N-1} x(n)x(n+1) & \sum_{K-1}^{K+N-1} x(n)^2 \;.\;.\;.\;.\;.\;. & \sum_{K-1}^{K+N-1} x(n)x(n+K) \\ \vdots & \vdots & \vdots \\ \sum_{o}^{N} x(n)x(n+K) & \sum_{o}^{K} x(n)x(n+K-1) & \sum_{o}^{K} x(n)^2 \end{bmatrix} \qquad (3.13)$$

and

$$\phi_{xx}(k-i) = \begin{bmatrix} \psi_{xx}(0) & \psi_{xx}(1) \;.\;.\;.\;.\;.\;.\;. & \psi_{xx}(N-1) \\ \psi_{xx}(-1) & \psi_{xx}(0) \;.\;.\;.\;.\;.\;.\;. & \psi_{xx}(N-2) \\ \vdots & \vdots & \vdots \\ \psi_{xx}(-N+1) & \psi_{xx}(-N+2) \;.\;.\;.\;.\;.\;.\;. & \psi_{xx}(0) \end{bmatrix} \qquad (3.14)$$

The normal equation then becomes

$$T \sum_{k=o}^{K} \hat{\psi}_{xx\,(k-i)} \; \hat{h}_{(k)} = \hat{\psi}_{xy(i)} \tag{3.15}$$

This equation appears similar to the sampled-data version of the Wiener-Hopf equation. However, the correlation functions used here are not true in the statistical sense and we can only call this equation the pseudo Wiener-Hopf equation.

3.2 Optimum Test Signals:

So far, we have not mentioned anything about the type of test signals to be used for identifying system impulse response. In order to obtain an optimum in the least square sense test signal, it is necessary to have a signal which will satisfy the criterion as defined by Levin [1]. His criterion is as follows:

If x(n) is a deterministic signal then its value must be such as to satisfy the following two conditions

1. $\hat{\psi}_{xx}(0) \neq 0$
2. $\hat{\psi}_{xx}(i) = 0 \qquad 0 < i \leq K$

A class of binary signals which is known as pseudo-random binary sequence gives a good approximation to the optimum test signal. In addition, a pseudo binary random sequence has the following properties

1. The signal can be generated by means of linear feedback shift register
2. The multiplication of the signal can be done by a gating circuit
3. The delayed signal can be obtained with ease from a cyclic shift of its original signal by means of modulo 2 addition

4. The pseudo auto-correlation function is a good approximation to a unit impulse response and is fixed in form.

   We can estimate its error easily.

5. Since the signal is periodic, it shortens the time required to calculate the pseudo-correlation function.

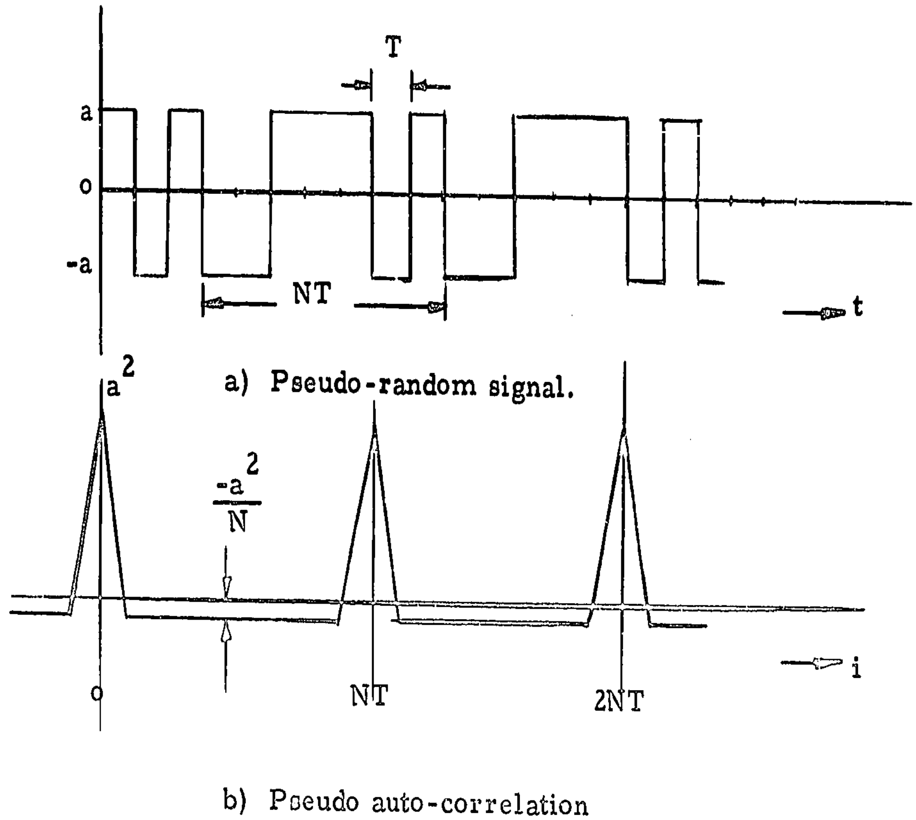


Fig. 3.2 Pseudo random signal and its auto-correlation function.

For a pseudo-random binary sequence with period (lenghth) N, then

$$\hat{\psi}_{xx}(0) = a^2 \quad \text{and} \quad \hat{\psi}_{xx}(i) = -\frac{a^2}{N}.$$

This condition is closely satisfies Levin's criterion for an optimum test signal. Thus we have

$$\hat{\phi}_{xx}(k-i) = \begin{bmatrix} a^2 & -a^2/_N & \cdots\cdots\cdots & -a^2/_N \\ -a^2/_N & a^2 & \cdots\cdots\cdots & a^2/_N \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ -a^2/_N & -a^2/_N & \cdots\cdots\cdots & a^2 \end{bmatrix} \qquad (3.16)$$

It can be shown [3] that

$$\hat{\phi}^{-1}_{xx}(k-i) = \frac{N}{a^2(N+1)} \begin{bmatrix} 2 & 1 & \cdots\cdots & 1 \\ 1 & 2 & \cdots\cdots & 1 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & 1 & \cdots\cdots & 2 \end{bmatrix} \qquad (3.17)$$

We now define the k^{th} vector estimate of the impulse response and the k^{th} pseudo cross-correlation function as follows:

$$\hat{h}_k = \begin{bmatrix} h(k,0) \\ h(k,1) \\ \vdots \\ h(k,N-1) \end{bmatrix} \qquad \hat{\phi}_{xy}(k,i) = \begin{bmatrix} \hat{\psi}_{xy}(k,0) \\ \hat{\psi}_{xy}(k,1) \\ \vdots \\ \hat{\psi}_{xy}(k,N-1) \end{bmatrix} \qquad (3.18)$$

The pseudo Wiener-Hopf equation becomes

$$\hat{h}_k = \hat{\phi}_{xx}^{-1}(k-i) \; \hat{\phi}_{xy}(k,i) \; \frac{1}{T} \quad . \tag{3.19}$$

Substitution of (3.17) gives

$$\hat{h}_k = \frac{N}{a^2(N+1)\;T} \begin{bmatrix} 2 & 1 & \cdots\cdots & 1 \\ 1 & 2 & \cdots\cdots & 1 \\ \vdots & \vdots & & \\ 1 & 1 & \cdots\cdots & 2 \end{bmatrix} \begin{bmatrix} \hat{\psi}_{xy}(k,0) \\ \hat{\psi}_{xy}(k,1) \\ \vdots \\ \hat{\psi}_{xy}(k,N-1) \end{bmatrix} \tag{3.20}$$

3.3 Bayes Estimate of System Impulse Response:

In the previous section, we derived an estimate of the system impulse response using the method of least squares. This method, however, does not consider any prior knowledge about the plant being identified. In practice, a certain amount of prior knowledge is available in any plant under the process of identification. A complete disregard for prior knowledge about the plant means that a great deal of information is lost, especially when information is at a premium. For these reasons, a Bayesian estimate is derived here which takes into account all prior knowledge about the plant being identified.

To begin, a confidence index is assigned to all the prior knowledge which we have on the plant under investigation. This confidence index is $I^*_{k-1} = \frac{1}{\sigma^2}$ which is inversely proportional to the variance of the prior impulse response. The prior impulse response is represented by the following vector

$$h^*_{k-1} = \begin{bmatrix} h_{-n} \\ h_{-n+1} \\ \cdot \\ \cdot \\ \cdot \\ h_{-1} \end{bmatrix} \tag{3.21}$$

For Gaussian processes, Parzen [7] has shown that the Bayes estimate is a weighted sum of the prior estimate and the maximum likelihood estimate, the weight being proportional to the confidence index in each estimate. The confidence index of the maximum likelihood estimate $\hat{h}_k$ after the k^{th} observation is given by

$$\hat{I}_k = \{E [\mid (\hat{h}_k - h_k)' (\hat{h}_k - h_k) \mid] \}^{-1} \tag{3.22}$$

where h_k is the actual value, $\hat{I}_k$ is inversely proportional to the mean square error. The complete derivation of this confidence index will be described in a later section. The Bayesian estimate of the system impulse response, after k observations is

$$h^*_k = I^{*-1}_k (\hat{I}_k \hat{h}_k + I^*_{k-1} h^*_{k-1}) \tag{3.23}$$

$$\text{where} \quad I^*_k = \hat{I}_k + I^*_{k-1}$$

Similarly

$$h^*_{k+1} = I^{*-1}_{k+1} (\hat{I}_{k+1} \hat{h}_{k+1} + I^*_k h^*_k)$$

$$\vdots$$

$$h^*_{k+j} = I^{*-1}_{k+j} (\hat{I}_{k+j} \hat{h}_{k+j} + I^*_{k+j-1} h^*_{k+j-1})$$

and

$$I^*_{k+1} = \hat{I}_{k+1} + I^*_k$$

$$\vdots \tag{3.24}$$

$$I^*_{k+j} = \hat{I}_{k+j} + I^*_{k+j-1}$$

Thus the resulting estimate will be a weighted sum of the preceding estimate with the latest estimate.

For a process with Gaussian white noise, the least squares estimate and the maximum likelihood estimate coincide. Therefore, the least squares estimate derived in the previous section is the required maximum likelihood estimate. It is

$$\hat{\underline{h}}_k = \begin{bmatrix} \hat{g}(k,0) \\ \hat{g}(k,1) \\ \vdots \\ \hat{g}(k,N-1) \end{bmatrix} = \frac{N}{a^2(N+1)\,T} \begin{bmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ \vdots & & & \\ 1 & 1 & \dots & 2 \end{bmatrix} \begin{bmatrix} \hat{\psi}_{xy}(k,0) \\ \hat{\psi}_{xy}(k,1) \\ \vdots \\ \hat{\psi}_{xy}(k,N-1) \end{bmatrix} \tag{3.25}$$

To make computation easy, it is necessary to obtain each succeeding vector from the preceding vector by addition of some perturbation terms. i.e, to obtain h_{k+1} from h_k. In this way we do not require to calculate each estimate from our original estimate

Since

$$\hat{h}_{k+1} = \begin{bmatrix} \hat{g}(k+1,0) \\ \hat{g}(k+1,1) \\ \vdots \\ \hat{g}(k+1,N) \end{bmatrix} = \frac{N}{a^2(N+1)\,T} \begin{bmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \dots & 2 \end{bmatrix} \begin{bmatrix} \hat{\psi}_{xy}(k+1,0) \\ \hat{\psi}_{xy}(k+1,1) \\ \vdots \\ \hat{\psi}_{xy}(k+1,N) \end{bmatrix} \tag{3.26}$$

It is necessary to partition the above matrix equation in order to obtain a recursive relation between each estimate. After a bit of manipulation, we obtain the following relation.

$$\hat{h}_{k+1} = \begin{bmatrix} \underline{\hat{g}_k} \\ \cdots\cdots \\ \hat{g}(k+1, N) \end{bmatrix} \tag{3.27}$$

where $\underline{\hat{g}_k} = \underline{\hat{h}_k} + C\, \hat{\psi}_{xy}(k+1,\ N)$

for $C = \dfrac{N}{a^2 (N+1)\, T}$

and

$$\hat{g}\,(k+1, N) = C\ \{\sum_{i=o}^{N} \hat{\psi}_{xy}(k+1, i) + \hat{\psi}_{xy}(k+1,\ N)\}$$

Similarly

$$\underline{\hat{g}_{k+j-1}} = \underline{\hat{h}_{k+j-1}} + C\, \hat{\psi}_{xy}(k+j,\ N+j)$$

and

$$\hat{g}(k+j, N+j) = C\ \{\sum_{i=o}^{N+j} \hat{\psi}_{xy}(k+j, i) + \hat{\psi}_{xy}(k+j, N+j)$$

Finally the $k+j^{th}$ iteration of the system impulse response will depend only on the previous vector $\underline{\hat{g}}_{k+j-1}$ and a succeeding term $\hat{g}\,(k+j, N+j)$ i.e.,

$$\underline{\hat{h}_{k+j}} = \begin{bmatrix} \underline{\hat{g}}_{k+j-1} \\ \cdots\cdots \\ \hat{g}(k+j, N+j) \end{bmatrix} \tag{3.29}$$

We see in this procedure that the dimensions of the input matrix keep on growing after more and more observations are obtained.

Therefore we must decide how large a dimension should be used in order to obtain the required accuracy.

Another approach is to keep the dimension of the input matrix fixed. This can be done quite easily, since the matrix is symmetrical with identical elements at its diagonal and with all off-diagonal elements equal to unity. It has been shown that after k+jth iteration, the estimate of the impulse response is

$$\underline{\hat{h}}_{k+j} = \{\hat{\phi}_{xx}^{-1}\}_{kk} \, \hat{\phi}_{xy}(k+j) \, \frac{1}{T} \tag{3.30}$$

where

$$\underline{\hat{h}}_{k+j} = \begin{bmatrix} \hat{g}(j) \\ \hat{g}(j+1) \\ \vdots \\ \hat{g}(k+j) \end{bmatrix} \qquad \phi_{xy}(k+j) = \begin{bmatrix} \hat{\psi}_{xy}(j) \\ \hat{\psi}_{xy}(j+1) \\ \vdots \\ \hat{\psi}_{xy}(k+j) \end{bmatrix}$$

The dimension of the input matrix is k by k and is fixed.

$$\{\phi_{xx}^{-1}\}_{kk} = \frac{N}{a^2(N+1)T} \begin{bmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 1 \\ \vdots & & & \\ 1 & 1 & \cdots & 2 \end{bmatrix}_{k \times k} \tag{3.31}$$

When using pseudo-random binary sequence of period (length) N as test signal, the maximum number of distinct delay version of the original signal is (N-1). Keeping the dimension of the input matrix fixed and letting k = N-1, it can be shown that $\psi_{xy}(0)$ and $\psi_{xy}(N)$ are equivalent pseudo cross-correlation functions. They differ only by the external noise and parameters variation of the plant being identified.

Similarly

$$\hat{\psi}_{xy}(N+1) \text{ is equivalent to } \hat{\psi}_{xy}(1)$$

$$\hat{\psi}_{xy}(N+2) \text{ is equivalent to } \hat{\psi}_{xy}(2)$$

$$\vdots$$

$$\hat{\psi}_{xy}(N+j) \text{ is equivalent to } \hat{\psi}_{xy}(j).$$

In this way we can obtain a sequential estimate of the system impulse response.

3.4 Derivation of Confidence Index:

A brief description of the maximum likelihood confidence index will be given in this section. The confidence index is chosen to be the inverse mean square error between the actual and the estimated impulse response. This confidence index can be obtained in the following way; Since the actual output vector of the system is given by

$$\underline{Y} = \underline{X}\,\underline{h}_k\,T + \underline{V} \quad , \tag{3.32}$$

Multiply both side by $\underline{X}'$ gives

$$\underline{X}'\,\underline{Y} = \underline{X}'\,\underline{X}\,\underline{h}_k\,T + \underline{X}'\,\underline{V} \quad . \tag{3.33}$$

The estimated impulse response is given by the following equation

$$\hat{\phi}_{xy} = \hat{\phi}_{xx}\,\hat{\underline{h}}_k\,T \tag{3.34}$$

where

$$\hat{\phi}_{xx} = \underline{X}' \; \underline{X}$$

$$\hat{\phi}_{xy} = \underline{X}' \; \underline{Y}$$

Substituting equation (3. 34) into equation (3. 33),

$$\hat{\phi}_{xx} \; \underline{\hat{h}_k} \; T = \hat{\phi}_{xx} \; \underline{\hat{h}_k} \; T + \underline{X}' \; \underline{V}$$

and (3. 35)

$$(\hat{h}_k - h_k) = \hat{\phi}_{xx}^{-1} \; \underline{X}' \; \underline{V} \; \frac{1}{T} \; .$$

$$E\,(\underline{\hat{h}_k} - \underline{h_k})'\,(\underline{\hat{h}_k} - \underline{h_k}) = E \; \{(\hat{\phi}_{xx}^{-1} \; \underline{X}' \; \underline{V} \; \frac{1}{T})'(\hat{\phi}_{xx}^{-1} \; \underline{X}' \; \underline{V} \; \frac{1}{T})\}$$

$$= \frac{\hat{\phi}_{xx}^{-1}}{T^2} \, E(\underline{V}' \; \underline{V})$$

$$= \frac{\hat{\phi}_{xx}^{-1}}{T^2} \, \sigma_v^{2} \qquad (3.36)$$

where σ_v^{2} is the noise variance.

The maximum likelihood confidence index is chosen to be

$$\underline{I_k} = \{E\,[\,|\,(\underline{h_k} - \underline{h_k})'\,(\underline{h_k} - \underline{h_k})\,|\,]\}^{-1} \qquad (3.37)$$

When using pseudo-random binary sequence as test signal the term $\frac{\hat{\phi}_{xx}^{-1}}{T^2}$ is closely approximated by a constant. Thus the maximum likelihood confidence index is inversely proportional to the noise variance. In practice, however, the statistic of the noise variance is often an unknown quantity and we must approximate it as follows:

Let the mean estimated impulse response be

$$\bar{\underline{h}} = \frac{1}{N+1} \sum_{k=o}^{N} \hat{\underline{h}}_k \tag{3.38}$$

Then the approximate variance is given by

$$S^2 = \frac{1}{N+1} \sum_{k=o}^{N} (\hat{\underline{h}}_k - \bar{\underline{h}})' \, (\hat{\underline{h}}_k - \bar{\underline{h}}) \tag{3.39}$$

and

$$\sigma_v^2 \, \phi_{xx}^{-1} \simeq S^2$$

$$\hat{I}_k \simeq 1/S^2 \tag{3.40}$$

3.5 A Criterion for Choosing Sampling Time:

As indicated in Section 3.1, the system impulse response can be represented to a good approximation by the vector

$$\underline{h} = \begin{bmatrix} h(0) \\ h(1) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ h(K) \end{bmatrix} \tag{3.41}$$

provided KT is chosen to be greater than the significant duration of the impulse response. We can use $T = \frac{1}{2\,fc}$, where fc is the assumed upper cut-off frequency. If, due to our ignorance of the characteristics of the system being identified, T is chosen to be smaller than this, then a larger number of parameters will apparently be required for a given system settling time T_s.

$K = {}^{T}s/_{T}$ is the number of values h_k which must be estimated.
If T were chosen to be greater than $^1/2fc$, we would introduce a system structural error and the fine structural of the system i.e, the high frequency response characteristics would be lost. The number of impulse response parameters to be estimated is $K = {}^{T}s/_{T}$. Therefore the minimum number of input samples observed must be 2K-1. If more data is observed we have to use a longer measurement time interval Tm. This will allow us to form an excess number of equations relative to the number of independent parameters to be estimated. A redundancy is obtained from the excess equations which can be used for noise smoothing. Kerr and Surber [6] have defined a "redundancy factor" R.

$$R = \frac{\pi}{K} = \frac{M-K}{K} = \left[\frac{M}{k} - 1 \right] \qquad (3.42)$$

where

$K = {}^{T}s/_{T}$ = minimum number of output samples

$(2K-1)$ = minimum number of input samples

M = total number of observed output samples

$\pi = (M-K)$ = number of excess output samples.

If M and K are expressed in terms of time intervals of T_s and T_m, the "redundancy factor" then becomes

$$R = \left[\frac{T_m - 2T_s}{T_s} + \frac{1}{K} \right] \qquad (3.43)$$

For continuous signals with no bandwidth limitation i.e. $T = 0$ and $K \to \infty$, the minimum measuring time for zero redudndancy becomes $T_m = 2T_s$. For a bandlimited system, the minimum observation time is of the form

$$T_m = [(2 - \frac{1}{K}) T_s] \tag{3.44}$$

3.6 Comparison of Bayes and Maximum Likelihood Estimates:

Bayes and Maximum likelihood estimates are compared in terms of their mean square errors. It has been shown [1], that for a process with Gaussian white noise the maximum likelihood estimate of h_k is given by;

$$\underline{\hat{h}}_k = (X' X)^{-1} \underline{X}' \underline{Y} \tag{3.45}$$

and the mean square estimation error is

$$\underline{\hat{I}}_k^{-1} = (X' X)^{-1} \cdot \sigma_v^2$$

When the a priori probability density $p(h_k)$ of h_k is known along with the statistic of the noise, the Bayes estimate h_k^* for the given measurement $\underline{Y}$ has been shown to be the mean of h_k. For a process with Gaussian white noise and with mean h_{k-1}^* and variance σ^2, Parzen [7] has shown that the Bayes estimate h_k^* is given by

$$h_k^* = I_k^{*\,-1} (\hat{I}_k \hat{h}_k + I_{k-1}^* h_{k-1}^*) \tag{3.46}$$

Where $I_k^* = \hat{I}_k + I_{k-1}^*$ is the inverse mean square error and $I_{k-1}^* = \frac{1}{\sigma^2}$.

Kozin [8] has shown that the mean square error of the Bayes estimate is in general less than the mean square error of the maximum likelihood estimate. For convenience, we compare the mean square errors of different estimates in terms of the mean square error of the maximum likelihood estimate. To facilitate the comparison, we define a "correction factor" Q where

$$Q = \frac{I^{*}_{k-1}}{\hat{I}_k + I^{*}_{k-1}} \tag{3.47}$$

and

$$0 \leq Q \leq 1$$

The mean square error of the Bayes estimate is given by the following identity:

$$I^{*\,-1}_{k} = \hat{I}_k^{-1}\left(1 - \frac{I^{*}_{k-1}}{\hat{I}_k + I^{*}_{k-1}}\right)$$

$$= \hat{I}_k^{-1}\,(1 - Q) \tag{3.48}$$

Since $0 \leq Q \leq 1$, the mean square error of the Bayes estimate will always be less than that of the maximum likelihood estimate except when $Q = 0$. The value of Q approaches zero only when $\sigma^2 \to \infty$ for a given σ_v^2. The value of σ^2 can be considered as a measure of the range in which h_k may be expected to lie. When the possible range of values of h_k is uncertain, this corresponds to assuming a very large a priori variance σ^2. The Bayes estimate then approaches the maximum likelihood estimate. With known a priori mean and known a priori variance, the Bayes estimate is a better estimate in the sense of less mean square error than the maximum likelihood estimate. The smaller the a priori variance available, better will be the Bayes estimate as compared to the maximum likelihood estimate.

3.7 Improvement Over Identification Time:

This section deals with an improvement on the identification time T_I when applying the Bayesian method to system identification. Lindenlaub [2] has shown that the identification time T_I using the least squares estimate is equal to the time required by his ideal identifier. The identification time of the ideal identifier represents a greatest lower bound obtainable by practical methods. This result was obtained without assuming any apriori knowledge about the system. Proper uses of any available apriori knowledge should reduce the identification time.

In general, the overall accuracy of the estimated impulse response is determined by its variance and sampling interval T. The variance, as described previously, is a direct measure of the mean square error of the estimate and a trade-off between variance and identification time can be made. For a process with Gaussian white noise, Lindenlaub [2] has shown that the identification time for his ideal identifier is given by:

$$T_I = \frac{2\, W_G K_G}{\gamma\, \sigma^2_{\hat{h}_k}} \tag{3.49}$$

where

T_I = the identification time

$W_G K_G$ = the gain-bandwidth product

γ = signal to noise ratio

$\sigma^2_{\hat{h}_k}$ = variance of the maximum likelihood estimate.

Similarly, the identification time required by the Bayesian estimate is given by:

$$T_I = \frac{2\, W_G K_G}{\gamma\, \sigma^2_{h^*_k}}$$

It is shown in the preceding section that the mean square error of the Bayes estimate is

$$\sigma^2_{h^*_k} = \sigma^2_{\hat{h}_k}\,(1 - \Omega) \tag{3.50}$$

Substituting equation (3.50) into equation (3.49), the identification time of the Bayes estimate becomes

$$(1 - Q)\,T_1 = \frac{2\,W_G\,K_G}{\gamma\,\sigma_{h_k^*}^{\,2}}$$

For the same accuracy, the identification time of the Bayes estimate will in general be less than the maximum likelihood estimate except when Q = 0. However, the condition for Q = 0 is when the Bayesian estimate approaches the maximum likelihood estimate. Thus, it is shown that proper uses of apriori knowledge of the plant will reduce the identification time for the same accuracy as compared to the case when no apriori knowledge is used.

CHAPTER IV

EXPERIMENTAL RESULTS

4.1 General Description of Experiment:

In the previous chapter, the method of Bayesian estimation is developed and applied to the problem of system identification. In this chapter, examples of a second order system, differing only in their damping factors, will be described in order to gain an insight into the method and its recursive feature. This method is readily applicable to higher order systems. The problem of identifying these examples using the Bayesian estimate has been programmed for digital computation. Computed results of each example will be shown in the following sections.

4.2 Example I

A Highly-Damped Second Order System:

The differential equation of the highly-damped second order plant is given by:

$$\frac{d^2 z}{dt^2} + \frac{dz}{dt} + z = x(t)$$

The impulse response of this plant was first identified using the method of maximum likelihood. It is assumed that the plant is corrupted by additive Gaussian white noise. Once the maximum likelihood estimate is obtained, the procedure for Bayesian estimate is used. The results of the maximum likelihood estimate and Bayesian estimate are shown in Fig. 4.2. The maximum likelihood estimate is more susceptible to external disturbance, whereas the Bayes estimate is not disturbed as much. In addition, the result of the Bayes estimate showed it converges rapidly toward a value which is very close to the actual of the plant.

4.3 Example II

A Critically-Damped Second Order System:

The differential equation of the critically-damped second order system is

$$\frac{d^2 z}{dt^2} + \frac{2dz}{dt} + z = x(t)$$

The maximum likelihood estimate of the impulse response was first obtained. Then the method of Bayesian estimate is used. Their results are shown in Fig. 4.3 and 4.4 which indicated that the Bayes estimate gave a better accuracy and has a tendency to smooth out much of the external noise.

4.4 Example III

An Under-damped Second Order System:

The differential equation of the under-damped second order system is

$$\frac{d^2 z}{dt^2} + \frac{.6\, dz}{dt} + z = x(t)$$

Again, the impulse response of this system was first estimated by the method of maximum likelihood and then using the Bayes estimator. Once more the result of the Bayesian estimate is better in accuracy than the maximum likelihood estimate. Also, the Bayes estimate converges toward a value which is very close to the true value of the plant.

4.5 Effect of External Noise at the Output:

Results indicated in Fig. 4.1-A and B showed that the effect of external noise on the maximum likelihood estimate is more pronounced than on the Bayes estimate. When the external noise level is doubled, Fig. 4.1-A and B, we observe immediately that the maximum likelihood estimate fluctuates widely, but we hardly noticed any variation in the Bayes estimate. This noise smoothing property of the Bayes estimate is one of the reasons of obtaining a better estimate.

4.6 Effect of A Priori Knowledge of Plant:

A fixed set of values which is about 10% lower than the actual value of the plant was used as a priori knowledge of the plant for each example described in the previous section. In each case, a low value of the Bayes estimate was observed. This is to be expected, since the apriori knowledge used in each plant was low. As more and more observations are taken at the output of the plant, the effect of a priori knowledges diminishes very quickly. This result is shown in Fig. 4.1, Fig. 4.2, and Fig. 4.3, When a correct apriori knowledge is used for the plant Fig. 4.2-B, the Bayes estimate of the first several iteration give a better result than if wrong apriori knowledge were used, Fig. 4.2-B. Again as more and more observations are obtained the effect of apriori knowledge disappeared. In both cases, the Bayes estimate converges rapidly to a value which is very close to the actual value of the plant.

4.7 Calculating the Improved Identification Time:

It is shown in the previous chapter that the identification T_I is reduced by a factor of 1 - Q when using the Bayesian method. In this section, several values of Q will be shown based on the experimental results of the previous examples. For example: Results of the Second Order Critically Damped System is shown below:

No. of Iteration on the same point	Value of Q $Q = \frac{I^*_{k-1}}{I^*_{k-1} + i_k}$	Bayesian Identification Time T_{IB}
1	.507	$.493\ T_I$
2	.664	$.336T_I$
3	.846	$.154T_I$
4	.974	$.026T_I$
5	.875	$.0125T_I$
6	.981	$.019T_I$
7	.970	$.030T_I$

where T_I is the identification defined by Lindenlaub [2].

NOTE: standard apriori knowledge in Fig. 4.1 to 4.4 means 10% low of the actual value and an unnormalize apriori confidence index of 100.

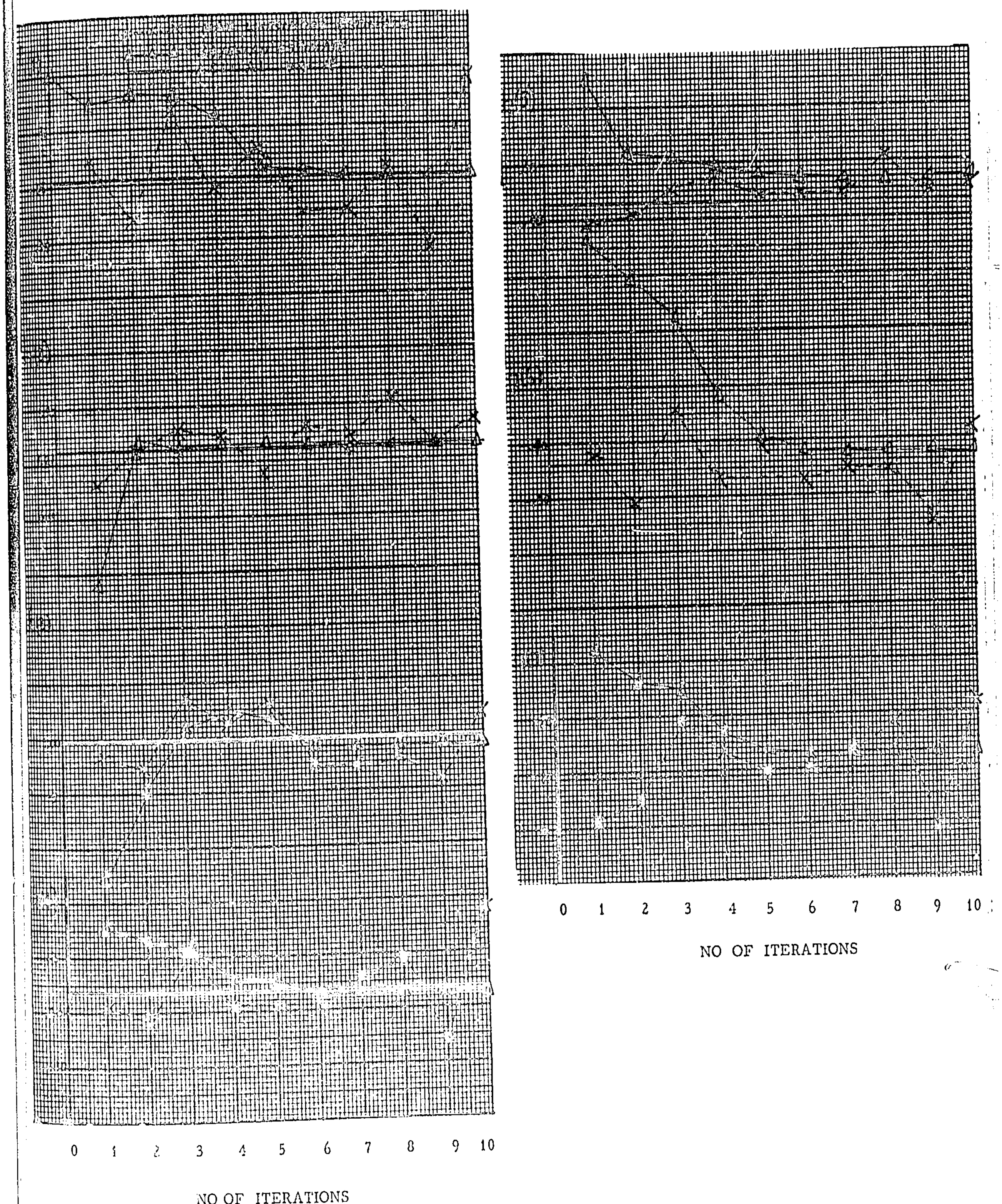
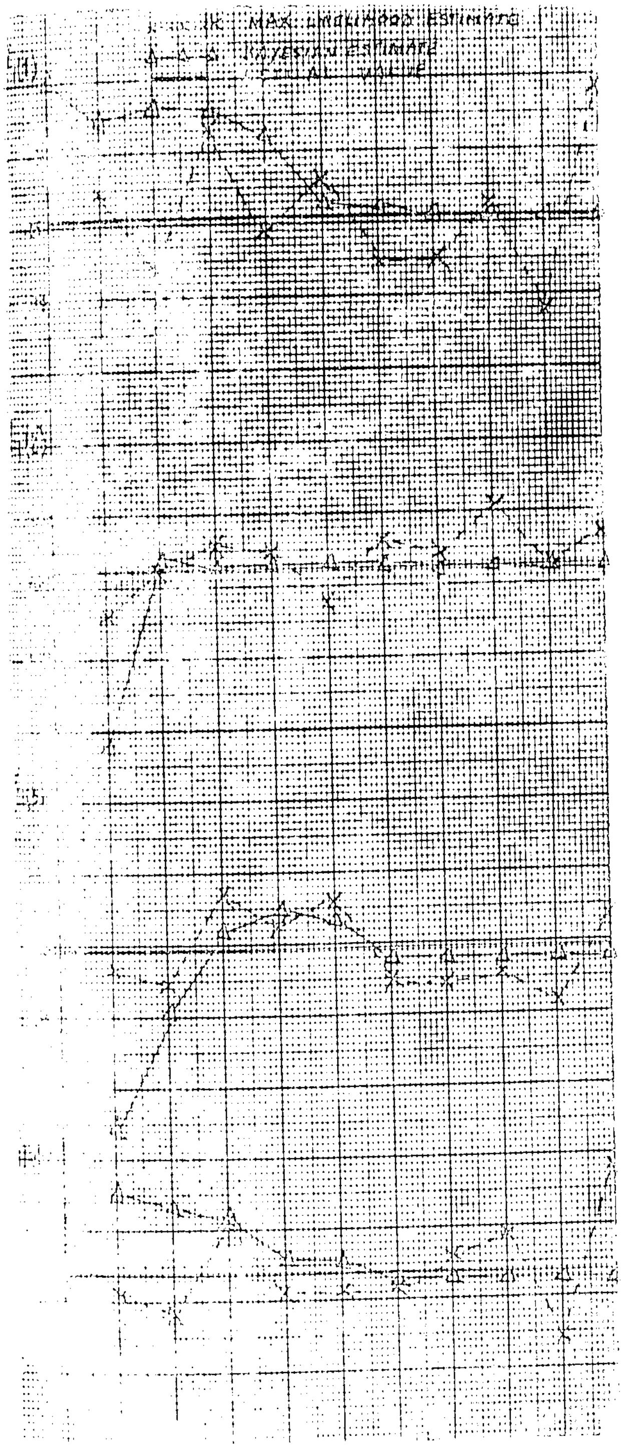
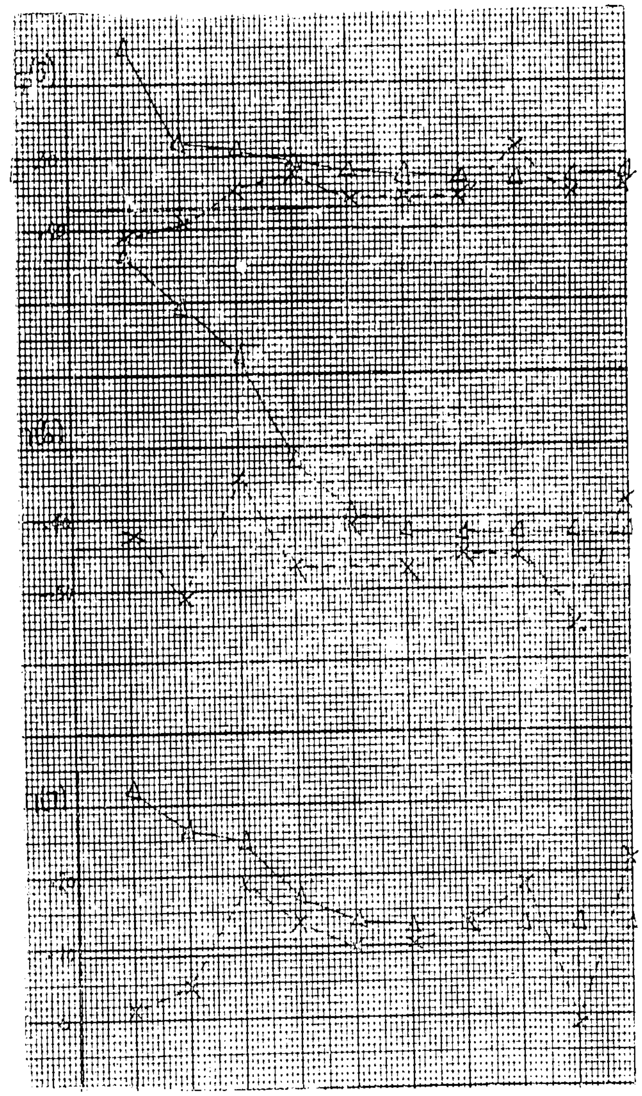
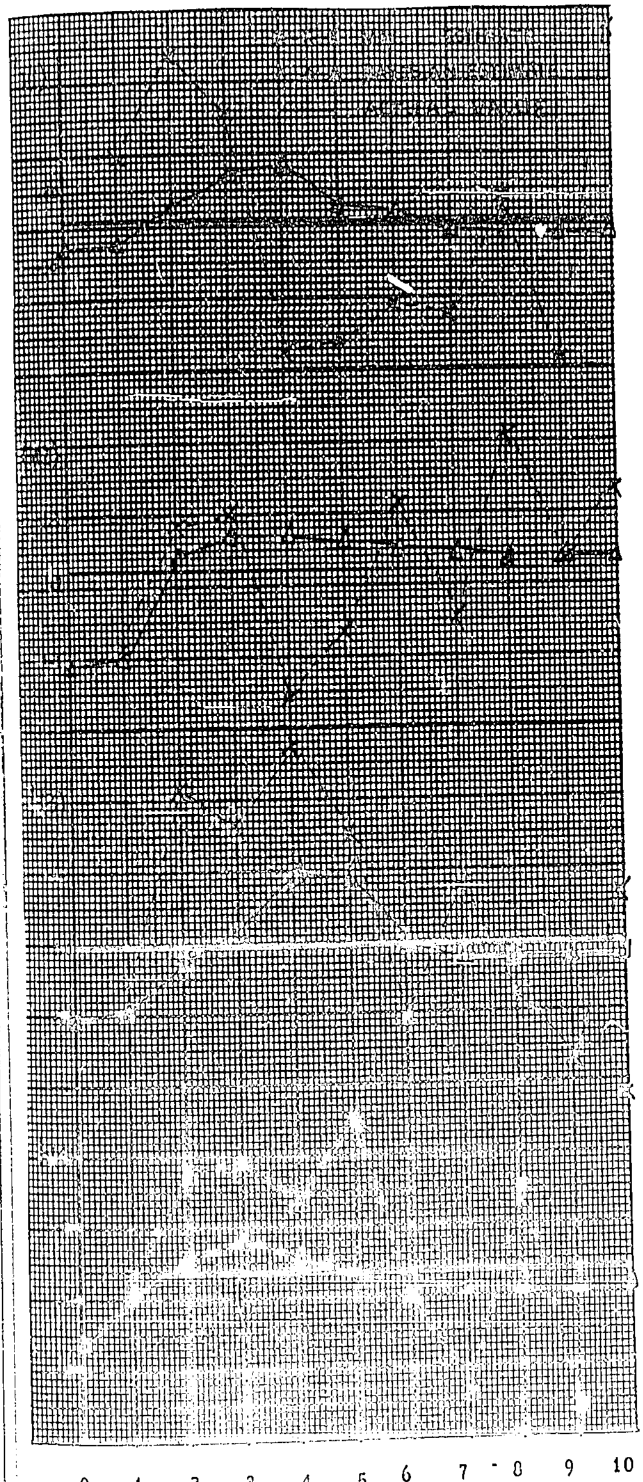


Fig 4-1-A Impulse Response of an underdamped second order system.

Maximum Likelihood and Bayesian estimates with standard *a priori* knowledge and Noise Level.


MAX. LIKELIHOOD ESTIMATE
BAYESIAN ESTIMATE





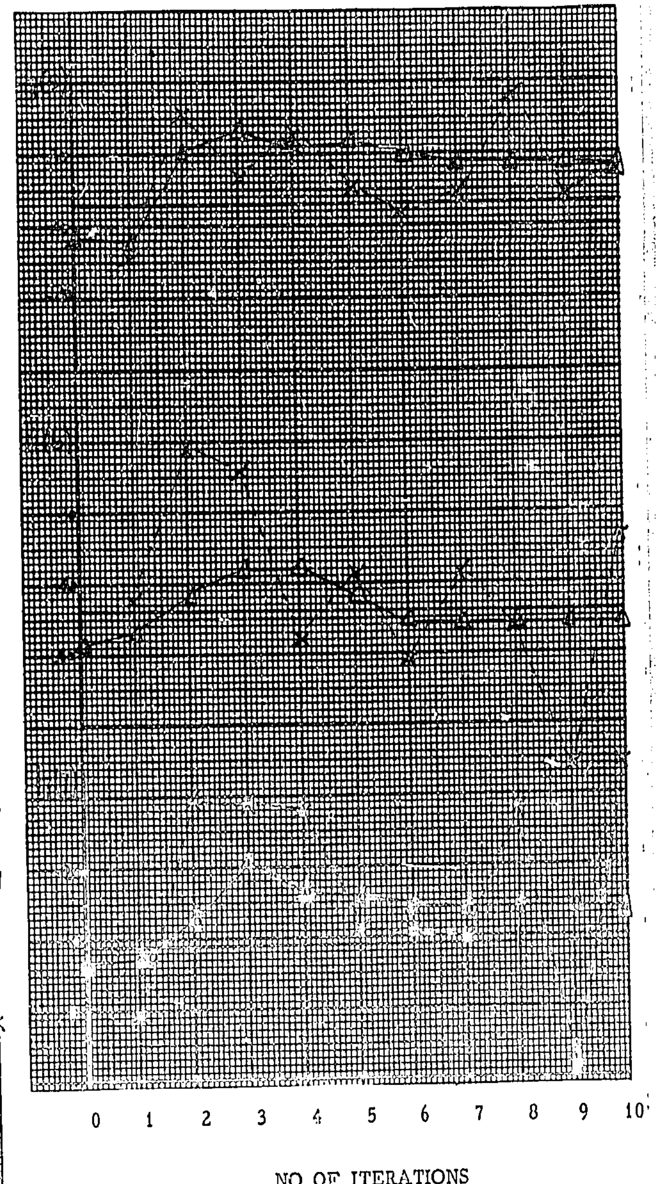
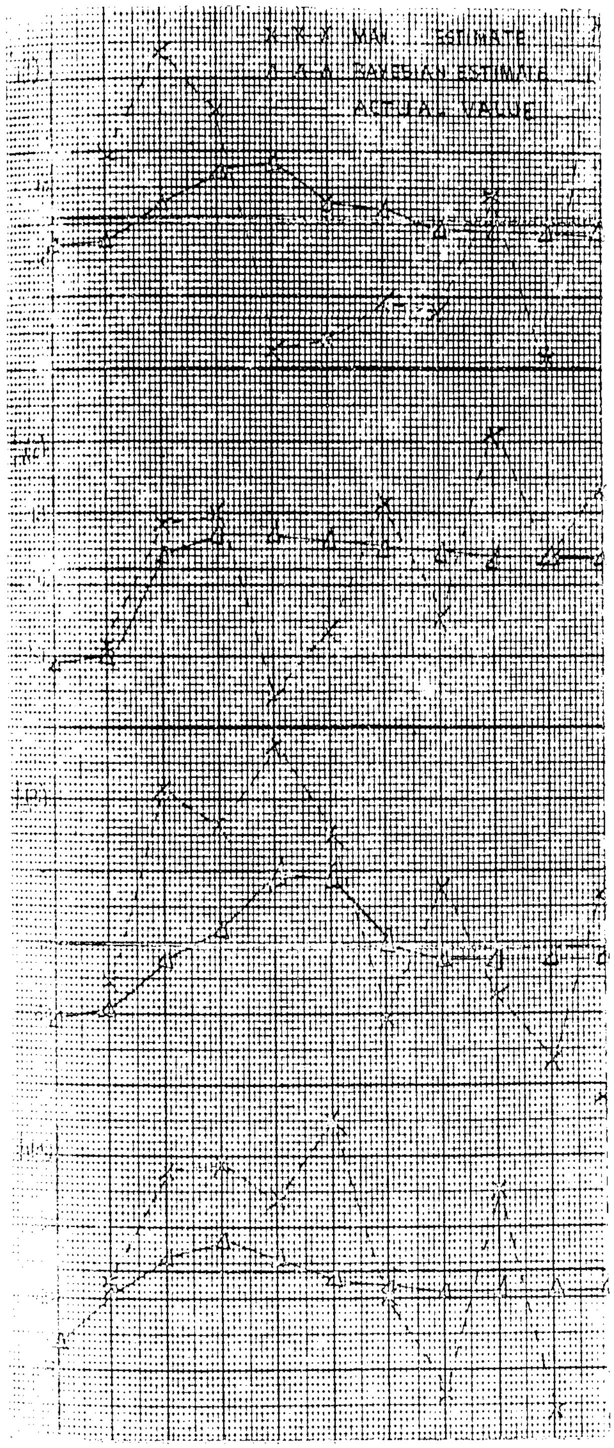
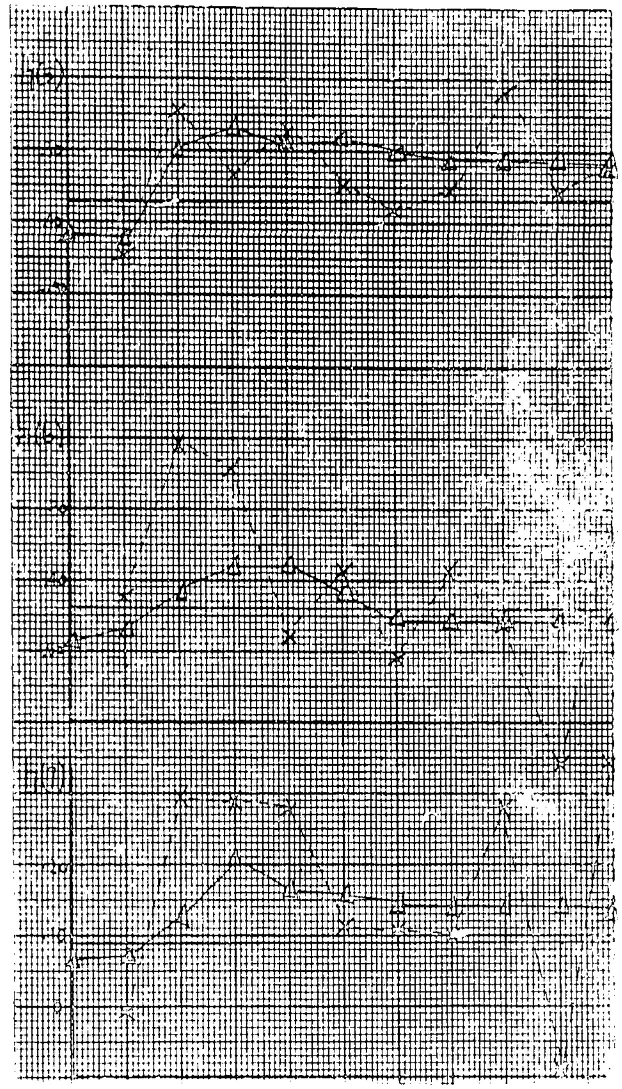


Fig 4-1-B Impulse Responce of an underdamped second order system.

Maximum Likelihood and Bayesian estimate with standard a priori knowledge and double Noise Level.


X-X-X MAX ESTIMATE
Δ-Δ-Δ BAYESIAN ESTIMATE
ACTUAL VALUE



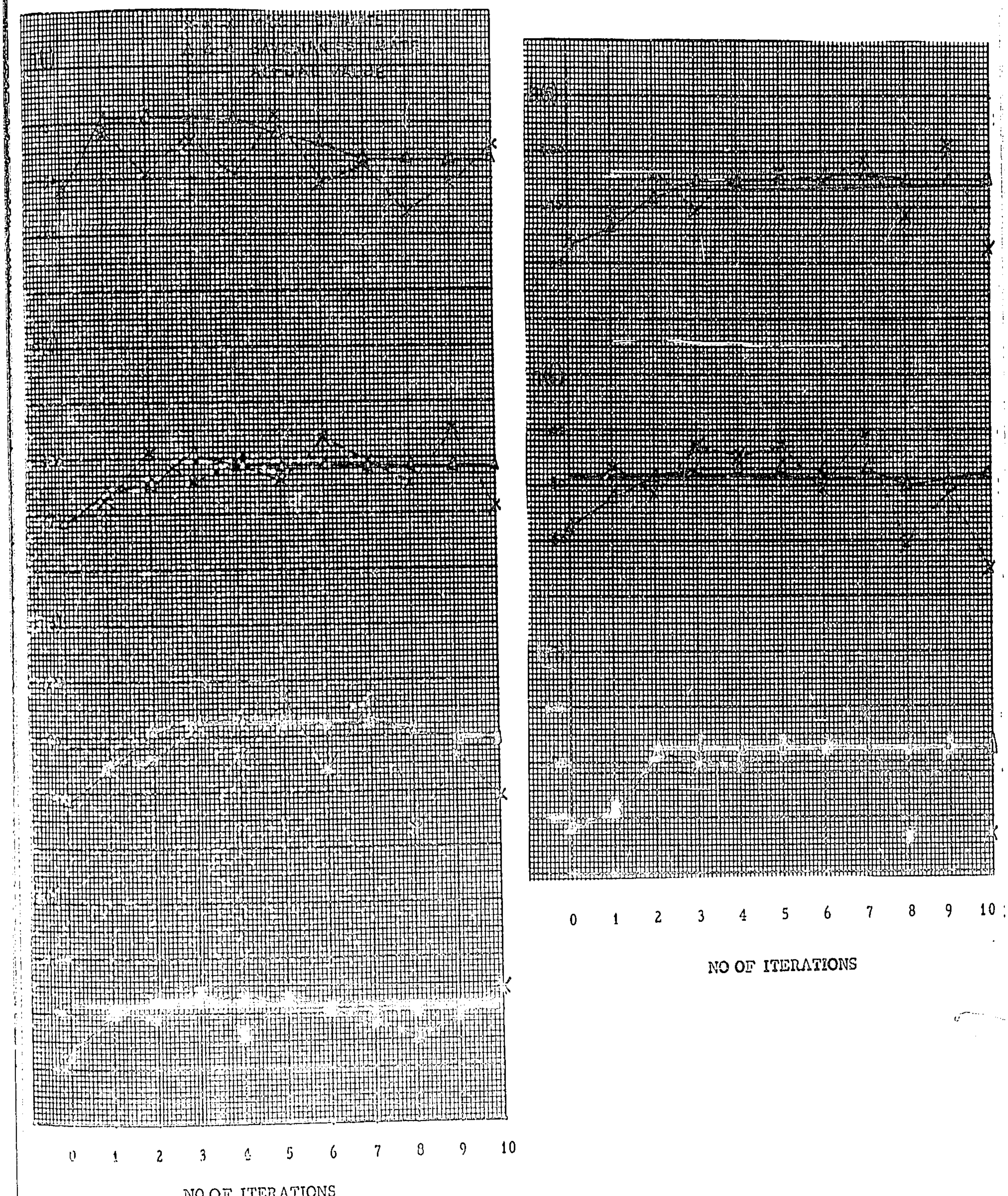


Fig 4-2-A Impulse Response of a highly-damped second order system

Maximum Likelihood and Bayesian estimates with standard <u>a priori</u> knowledge and Noise Level.

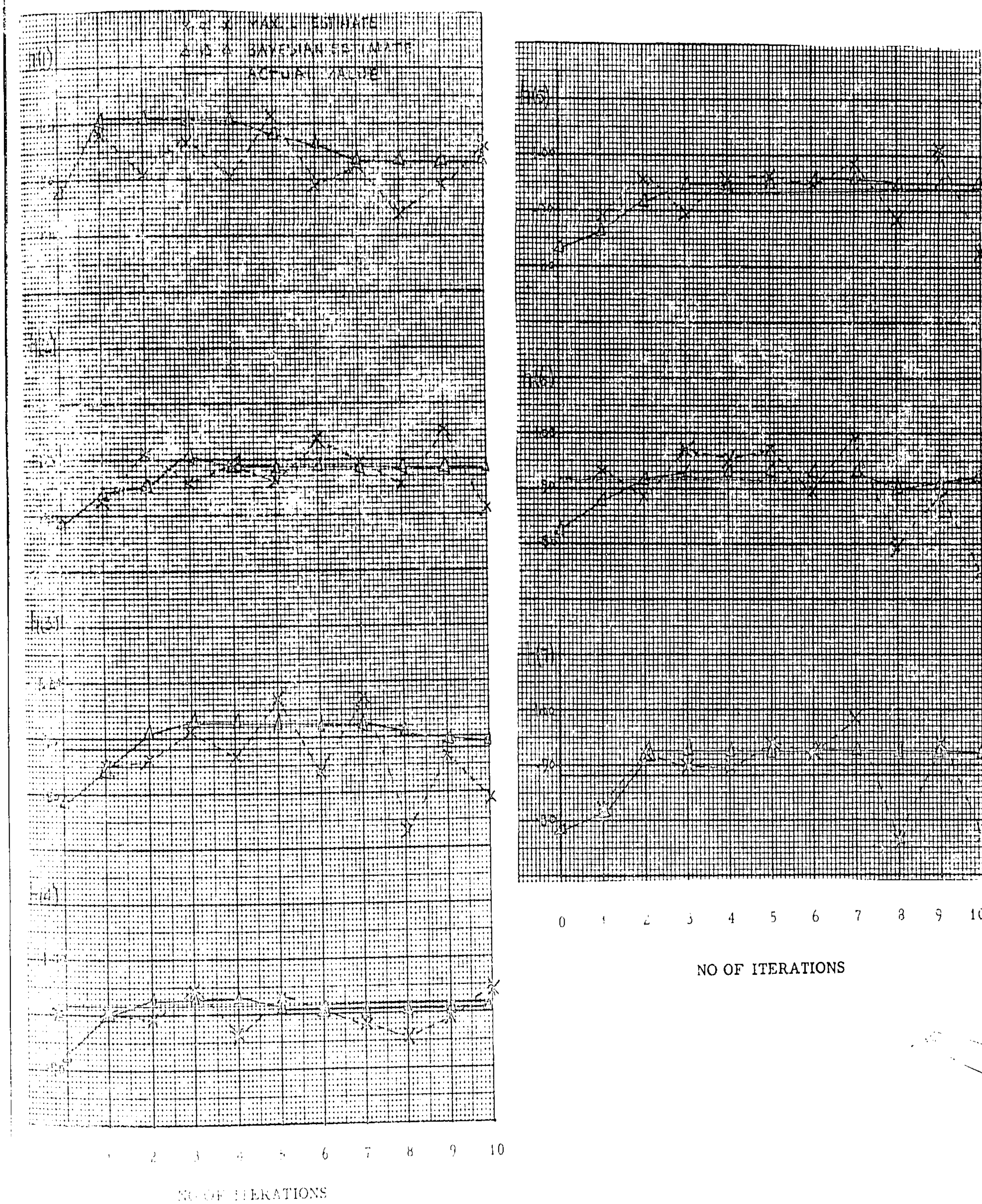
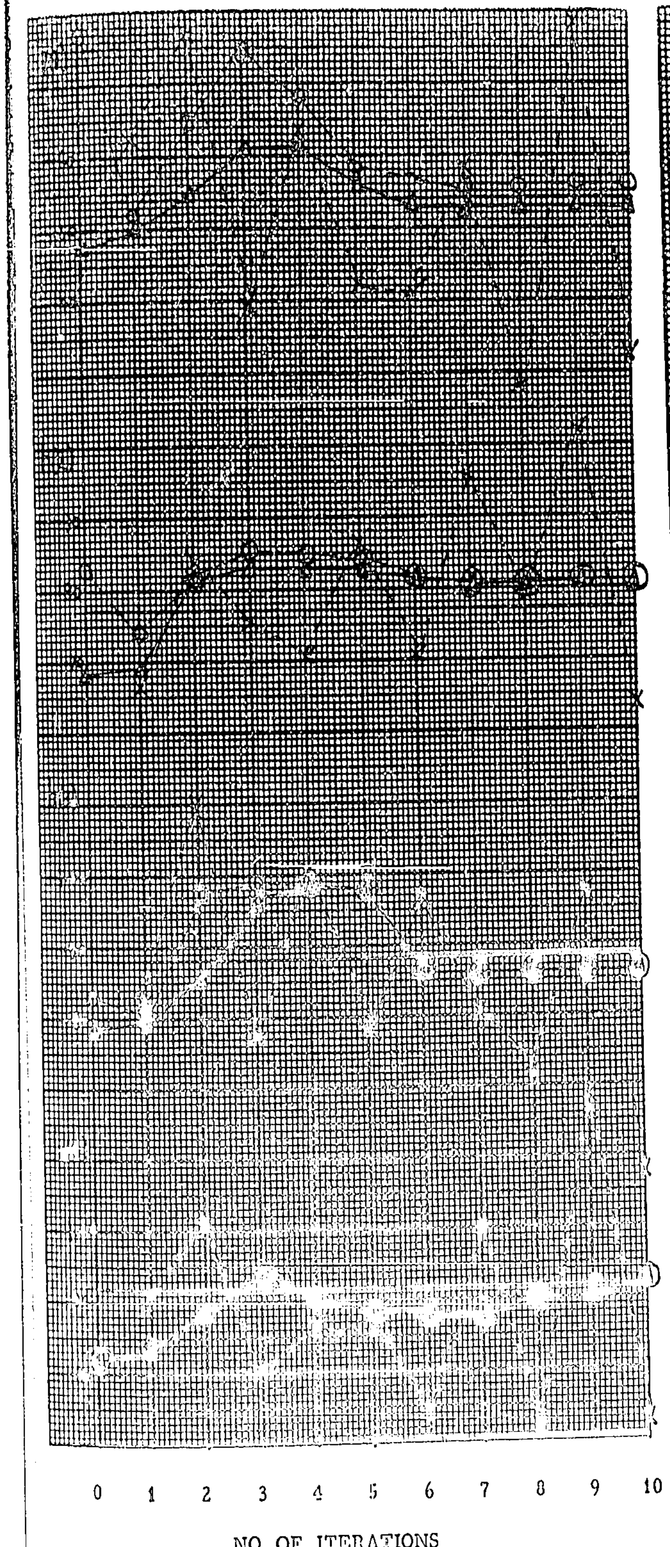


Fig [illegible] Impulse Response of a highly-damped second order system

Maximum Likelihood and Bayesian estimates with standard *a priori* knowledge and Noise Level.



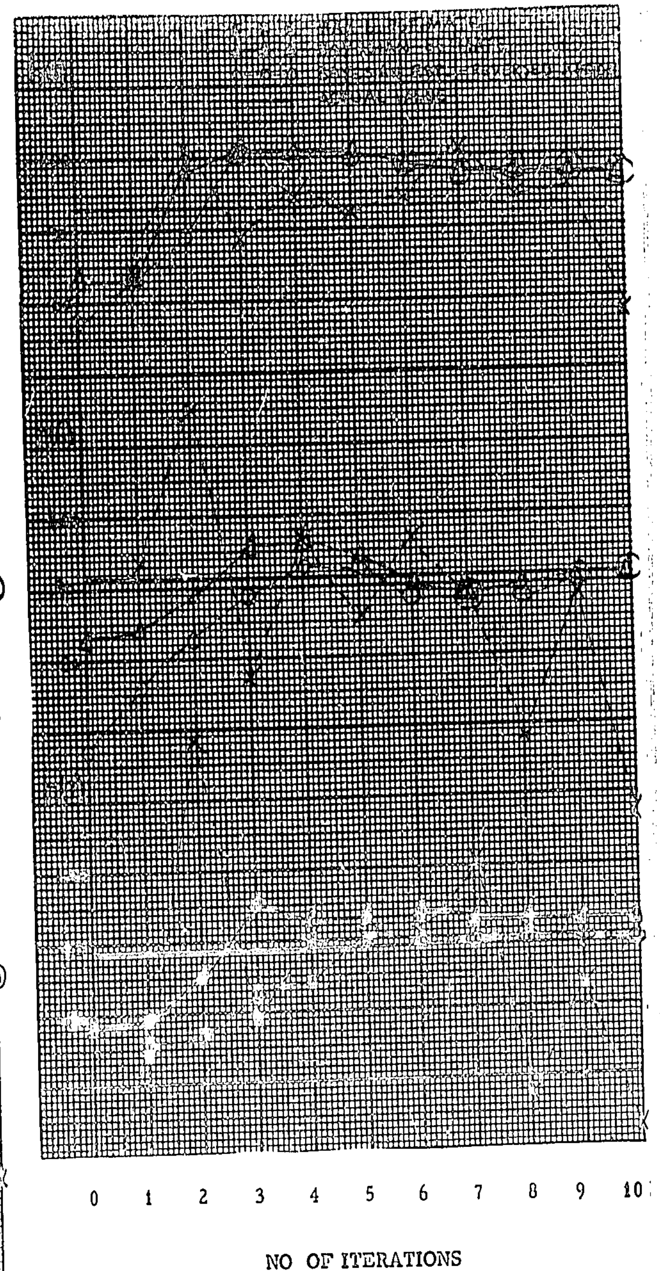


Fig 4-2-E Impulse Response of a highly-damped second order system

Maximum Likelihood and Bayesian estimates with standard a priori knowledge and double Noise Level.

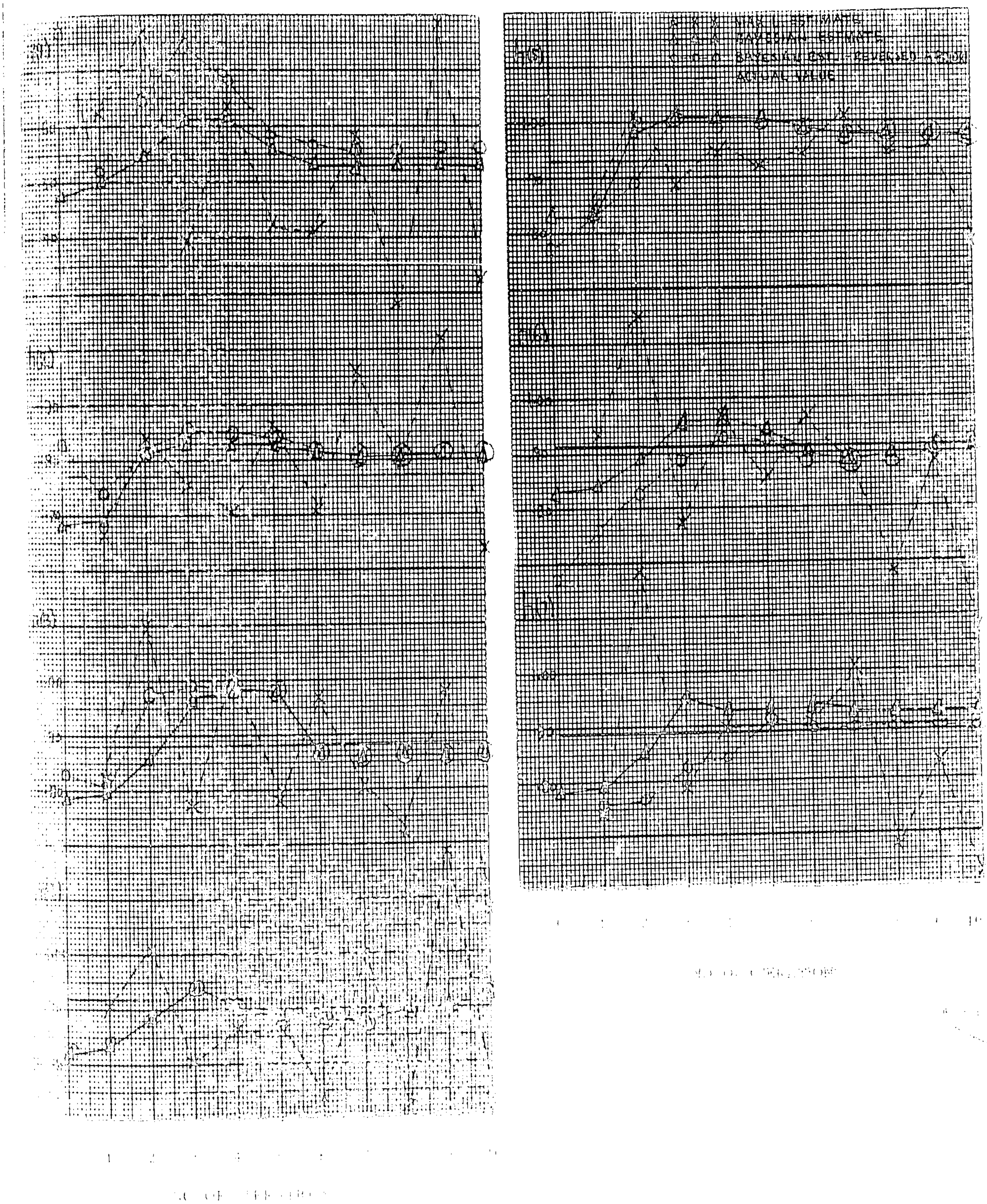


Fig. [illegible]

[illegible] Bayesian [illegible] a priori knowledge and double Noise Level.

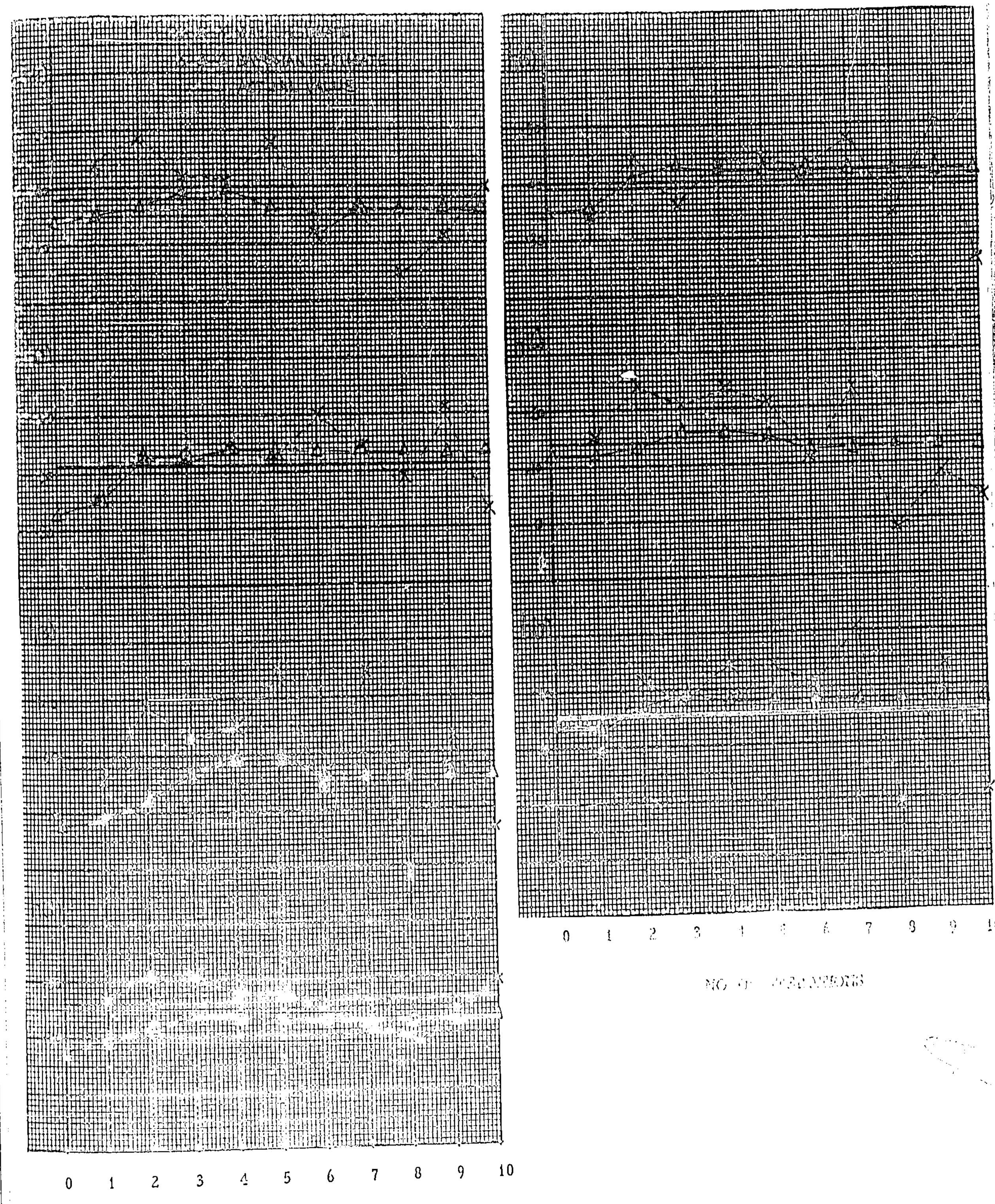
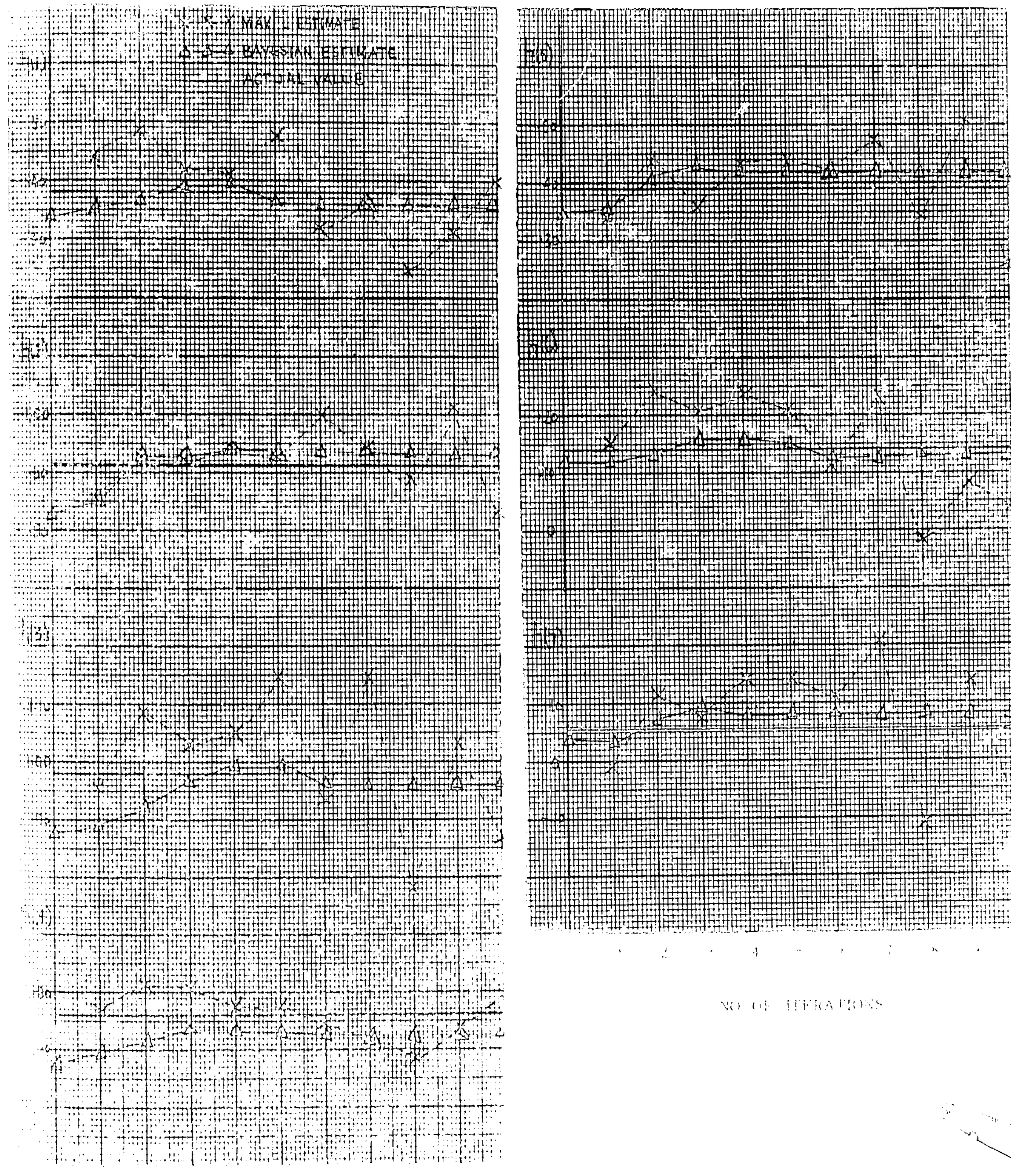


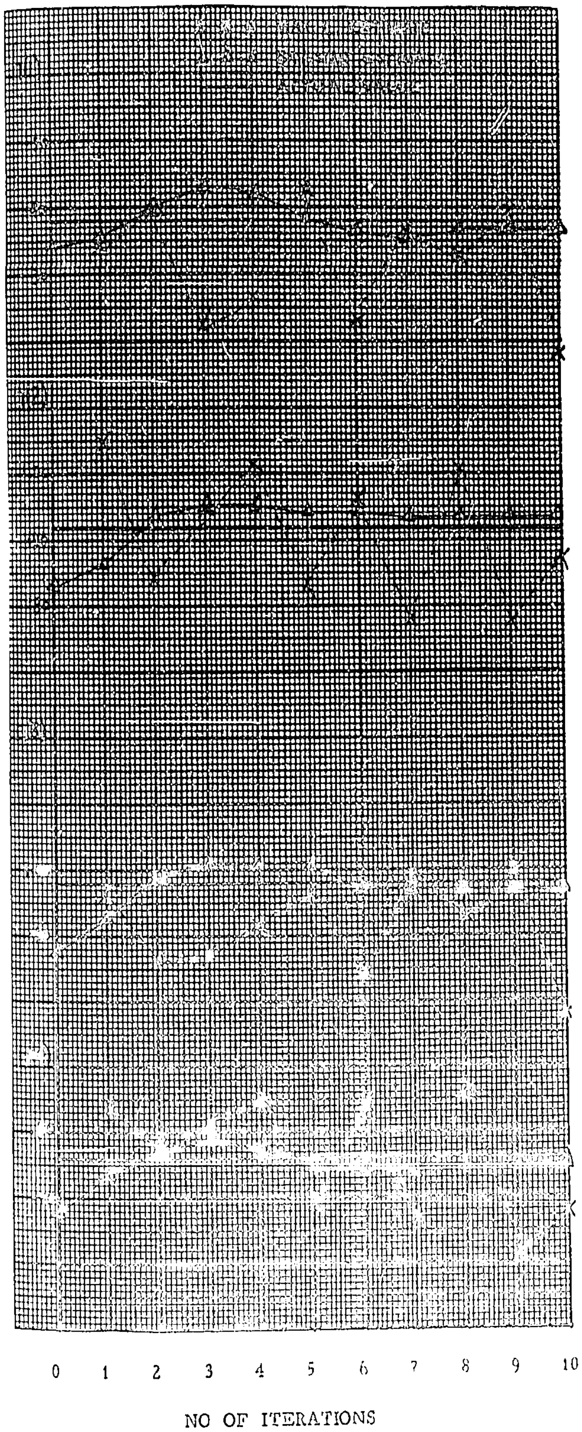
Fig 4-3-A Impulse Response of a critically-damped second order system.

Maximum Likelihood and Bayesian estimates with standard a priori knowledge and Noise Level.



... critically ...

... likelihood and Bayesian estimates with standard a prior. knowledge and Noise Level



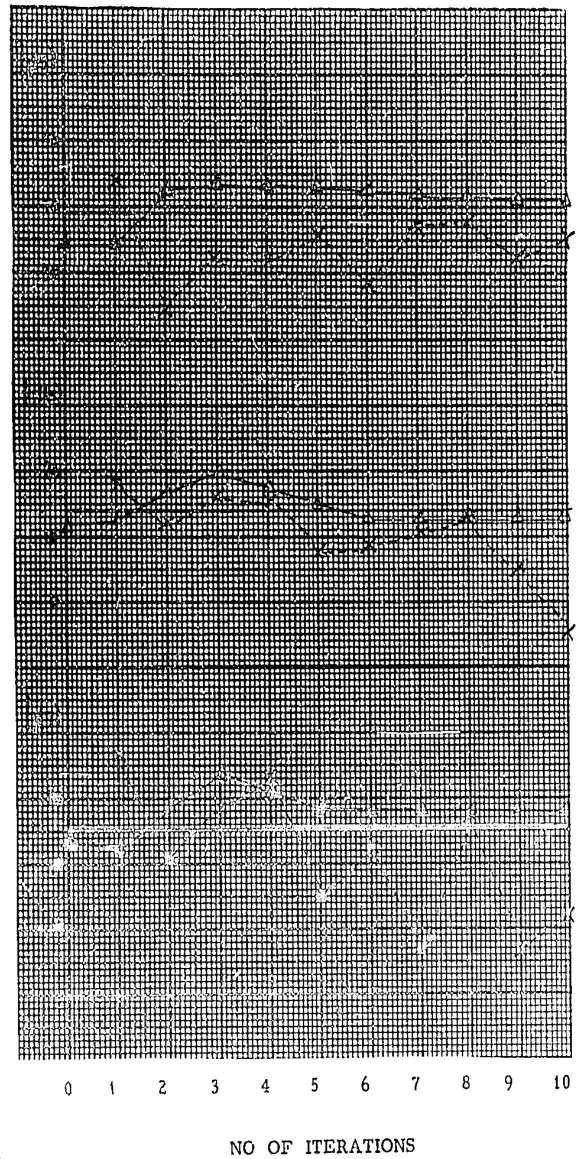


Fig. 4-3-B Impulse Response of a Critically-damped second order system.

Maximum Likelihood and Bayesian estimates with standard a priori knowledge and double Noise Level.

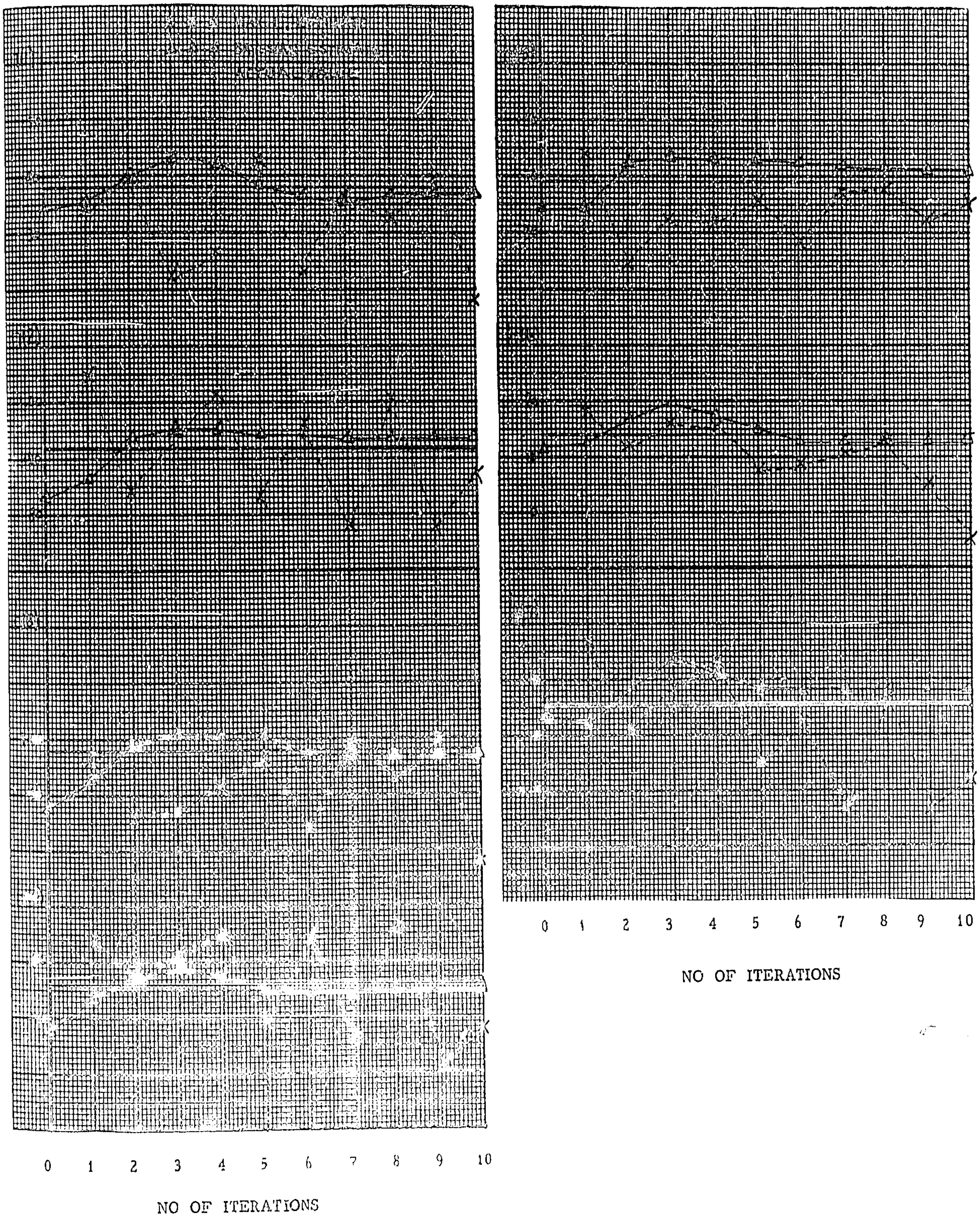
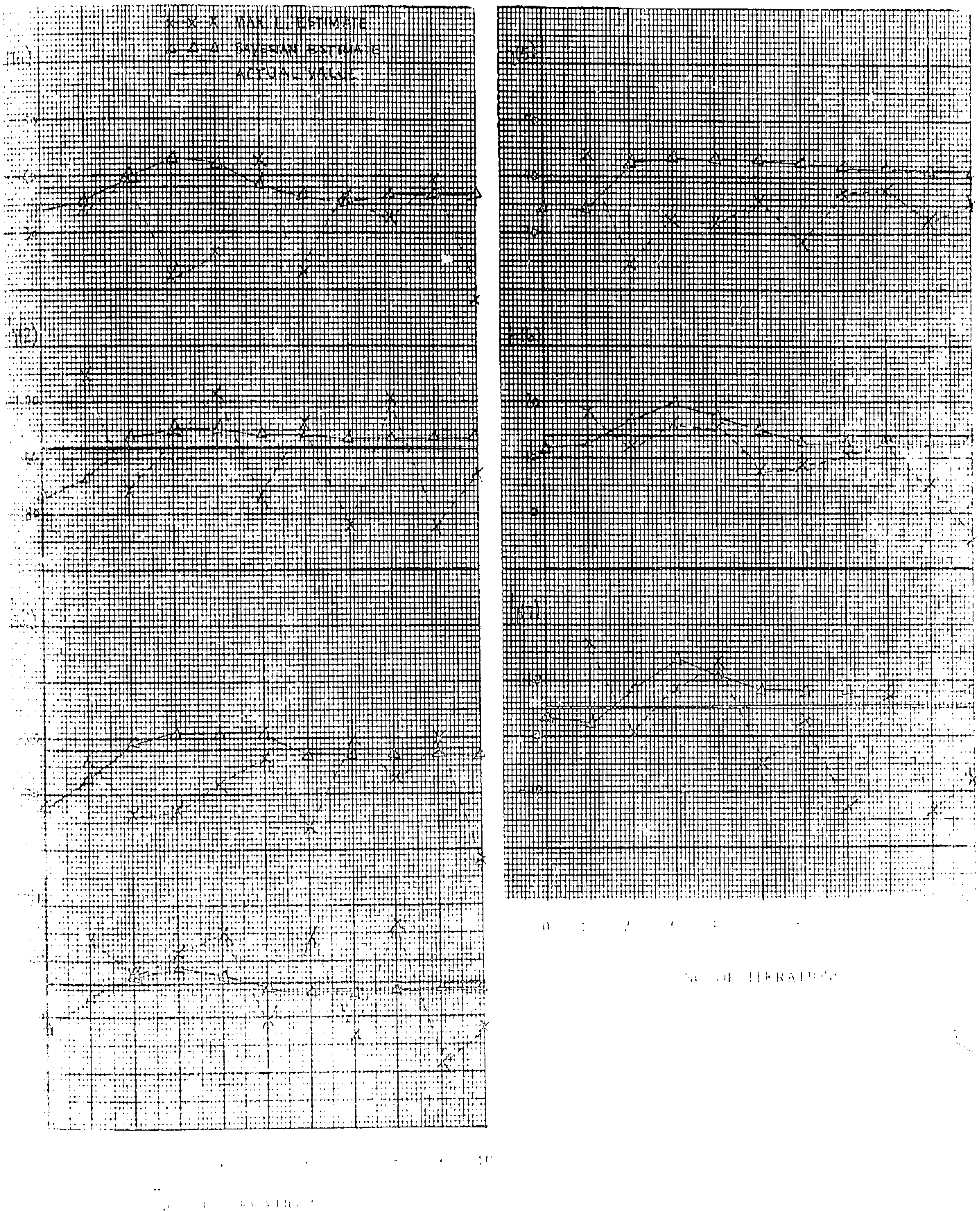


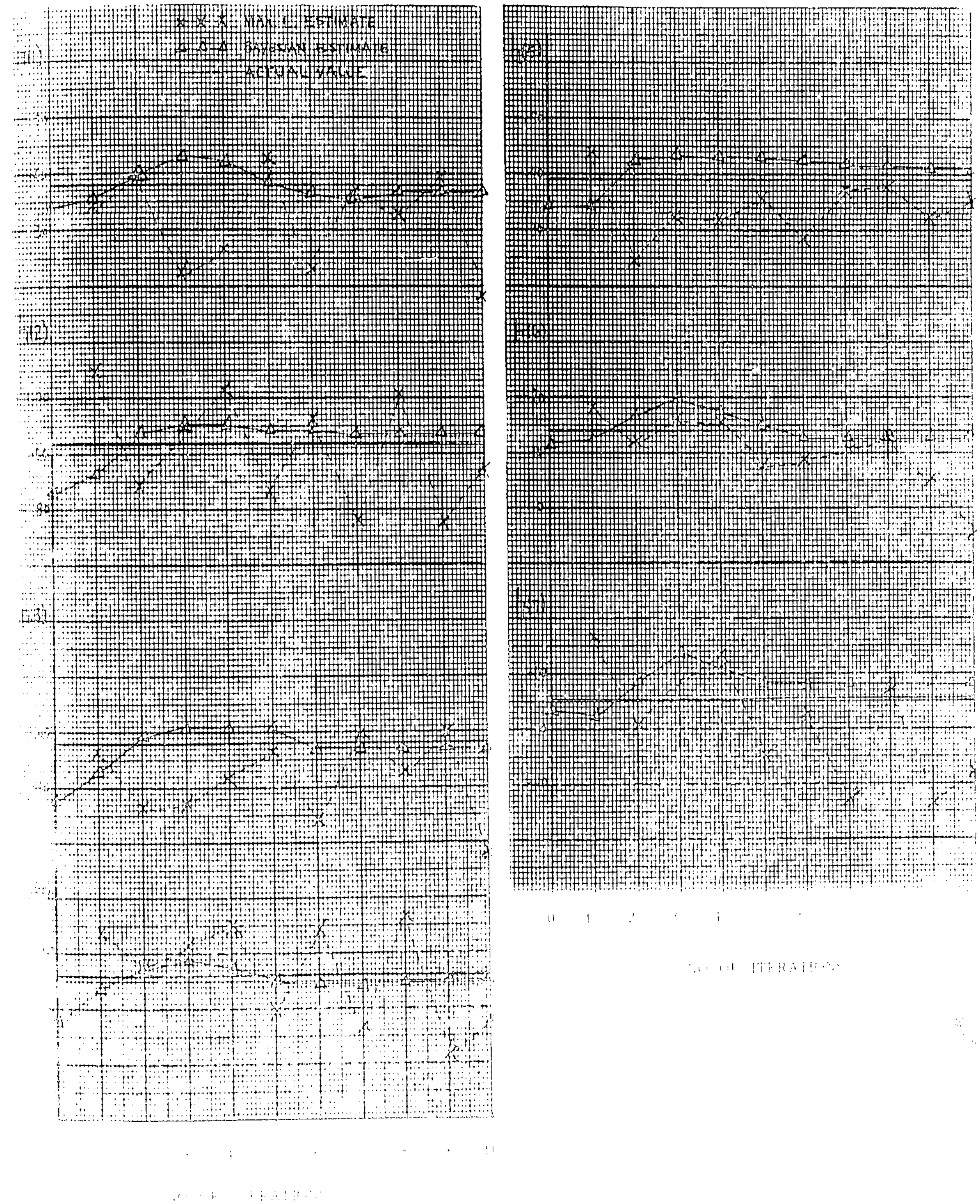
Fig. 4-3-B Impulse Response of a Critically-damped second order system.

Maximum Likelihood and Bayesian estimates with standard a priori knowledge and double Noise Level.



[illegible] mpulse Response of a Critically [illegible] amped second order system:

Maximum Likelihood and Bayesian estimates with standard a priori knowledge and double Noise Level



[illegible] response of a Critically [illegible] second order system.

[illegible] Likelihood and Bayesian estimates with standard *a priori* [illegible] and double Noise Level

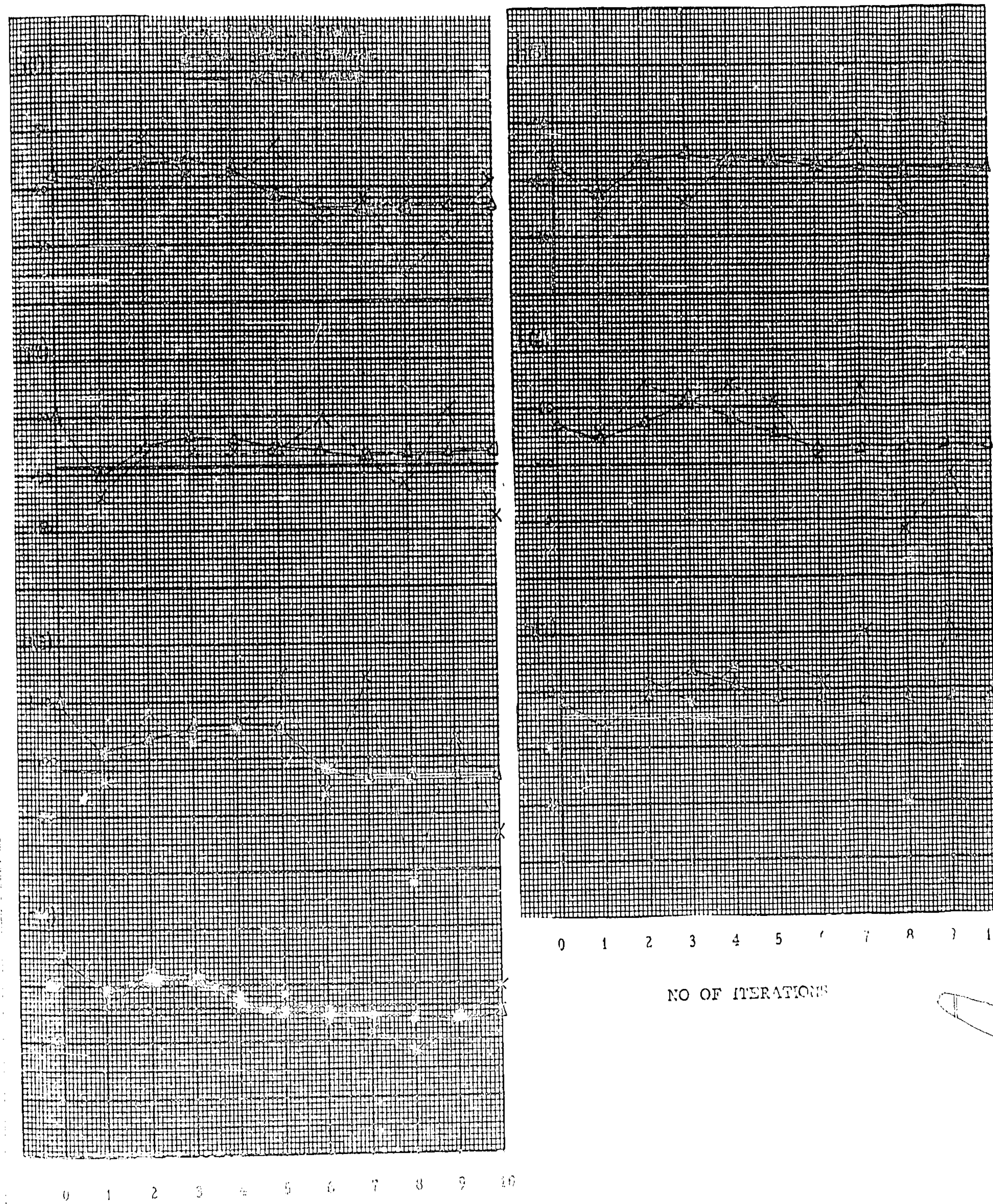


Fig 4-4 Impulse Response of a Critically-damped second order system.

Maximum Likelihood and Bayesian estimates with a priori knowledge 10% high and standard Noise Level.

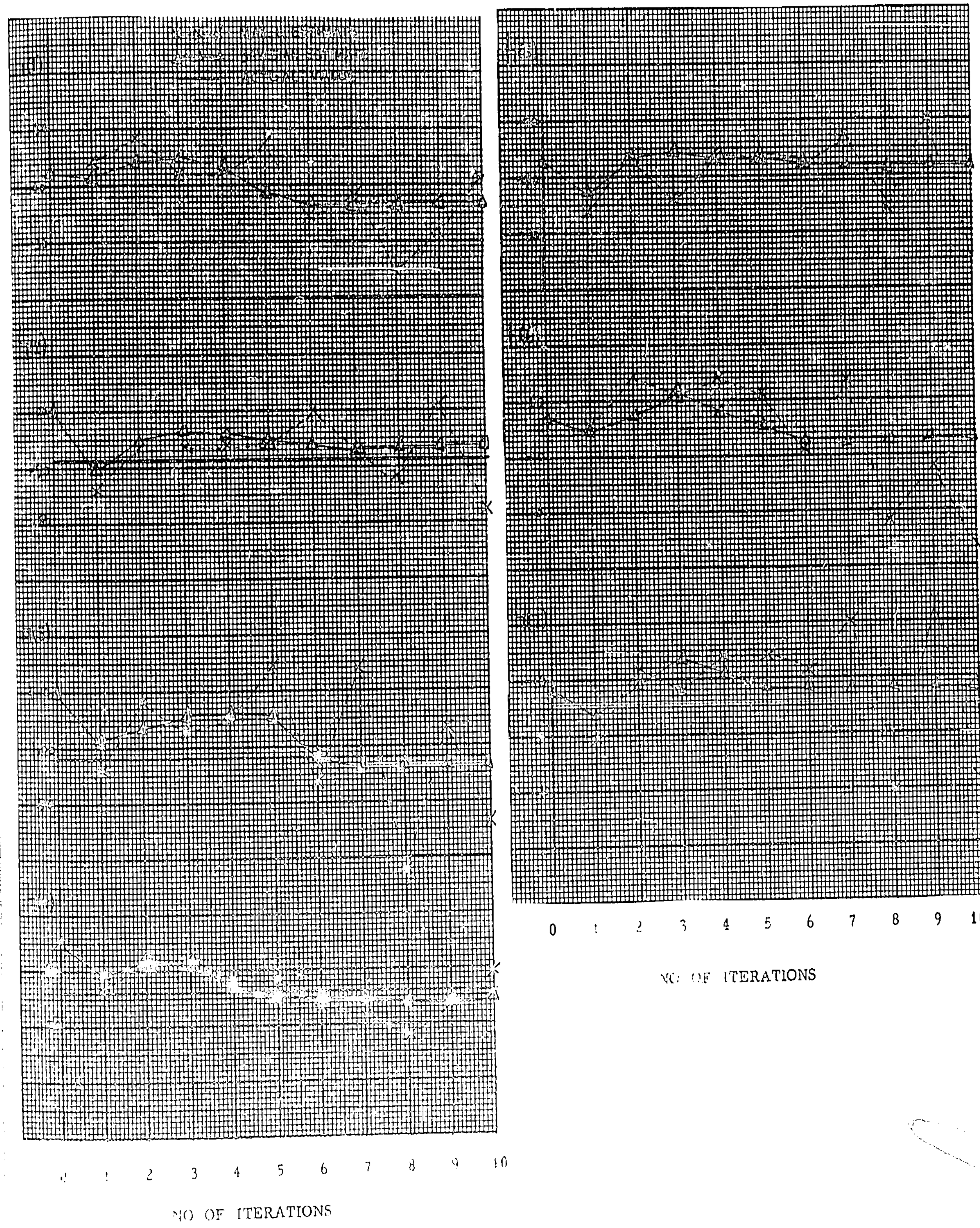
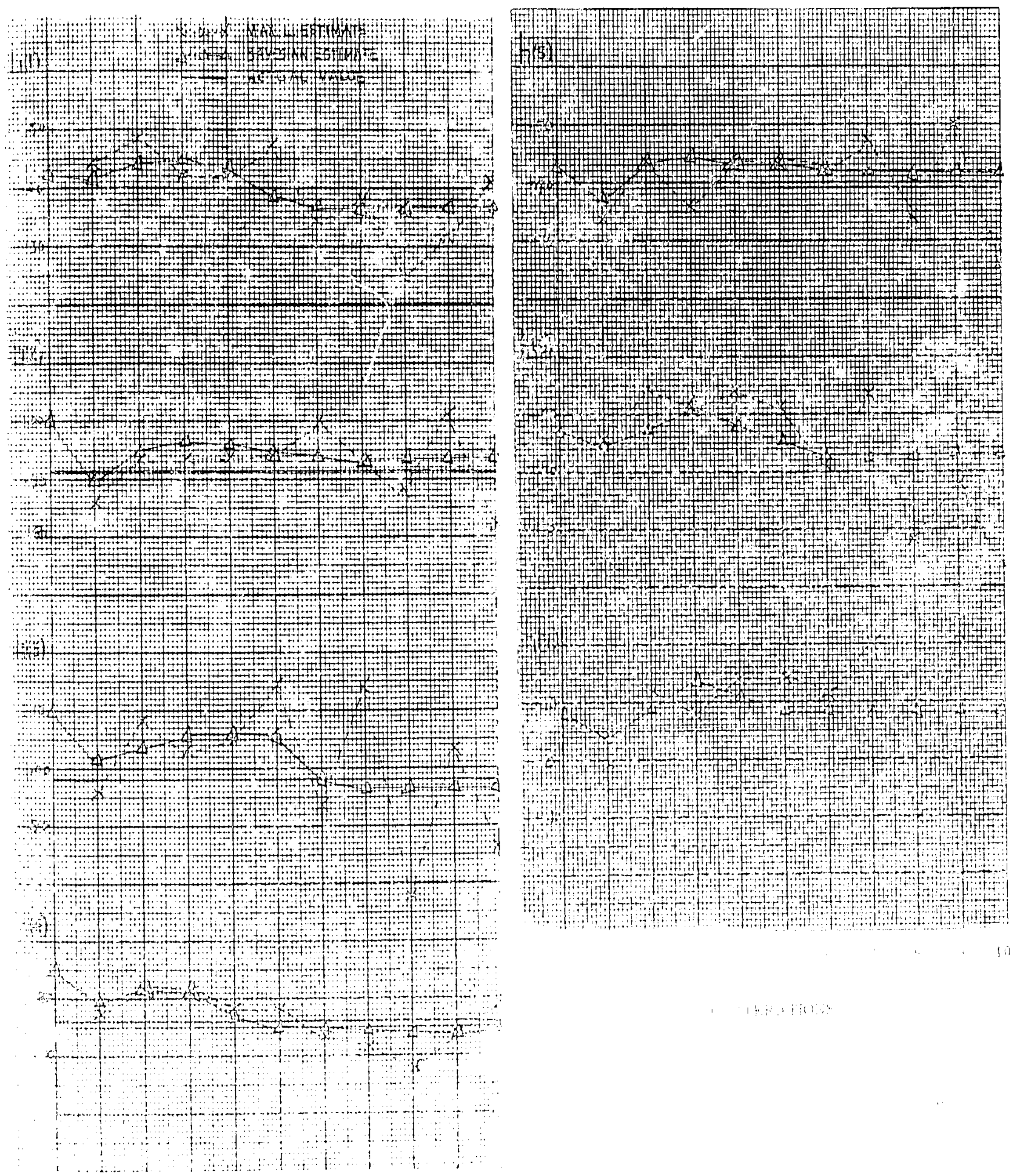


Fig 4-4 Impulse Response of a Critically-damped second order system.

Maximum Likelihood and Bayesian estimates with a priori knowledge 10% high and standard Noise Level.


MAX. LI. ESTIMATE
BAYESIAN ESTIMATE

CHAPTER V

CONCLUSIONS AND RELATED PROBLEMS

Throughout this investigation, a method has been found which unifies and extends the works of Levin [1], Lindenlaub [2], and Hill et al [3]. The Bayesian approach to system identification has produced a much better result as compared to previously available methods. Some of these results were described in Chapter IV.

The role of apriori knowledge of the plant used in the Bayesian estimate is critical in the first few iterations. Once sufficient observations are taken, the effect of the original apriori knowledge diminishes very quickly. This is to be expected, since the Bayesian method uses each preceeding estimate as the most recent apriori knowledge of the plant. Another feature of the method is that it yields a reduction of system identification time T_I by a factor $1 - Q$.

The investigation carried out in this thesis represents only a small part of the identification problem and its relation to adaptive control systems. There remain many interesting and unsolved problems on system identification. Two of them will be indicated:

1. The problems of using the Bayesian method to identifying linear time varying systems warrant further studies.
2. Another interesting problem is to use the Bayesian method for on line identification within an adaptive control system.

REFERENCES

1. Levin, M. J. "Estimation of the Characteristics of Linear System in the Presence of Noise" Sc. D. Dissertation, Columbia University, New York, New York, 1959.

2. Lindenlaub, J. C. "Limits on the Identification Time for Linear Systems" PhD. Dissertation, Purdue University, Lafayette, Indiana, 1961.

3. Hill, J. D., and McMurtry, G. J. "An Application of Digital Computers to Linear System Identification" IEEE Trans. of Automatic Control, Vol. AC 9, Pages 536 - 538, Oct. 1964.

4. Hazlerigg, A. D. G. and Noton, A. R. M. "An Application of Cross-correlation equipment to Linear System Identification" Proc. of IEE, Vol. 112, Pages 2385 - 2400, Dec. 1965.

5. Anderson, G. W., Buland, R. N. and Cooper, G. R. "Uses of Cross-correlation in an Adaptive Control System" Proc. of the National Electronic Conference Vol. 15, Pages 34 - 45, Oct. 1959.

6. Kerr, R. B. and Surber, W. H. "Precision of Impulse Response Identification Based on Short, Normal Operating Records" IRE Trans. on Automatic Control, Vol. AC 6, Pages 173 - 182, May, 1961.

7. Parzen, E. "Extraction and Detection Problems and Reproducing Kernel Hilbert Spaces" J. of SIAM Control, Series A, Vol. 1, No. 1, Pages 43 - 46, 1962.

8. Kozin, C. H. and Cooper, G. R. "The Use of Prior estimate in Successive Point Estimation of a Random Signal" J. of SIAM Applied Mathematics, Vol. 14, No. 1, Pages 112 - 129, Jan. 1966.

9. Magill, D. T. "Optimal Adaptive Estimation of Sampled Stochastic Processes" TR No. 6302 - 3, Dec. 1963, Stanford Electronic Lab., Stanford University, Stanford, California.

10. Shen, D. W. C. and Rosenburg, A. E. "Regression Analysis and its Applications to the System Identification Problem" Joint Automatic Control Conference, University of Minnesota, Minneapolis, Minnesota. Paper No. 16 - 2, Pages 446 - 451, 1963.

11. Giese, C. and McGhee, R. B. "Estimation of Nonlinear System States and Parameters by Regression Methods" Joint Automatic Control Conference, Rensselaer Polytechnic Institute, Troy, New York Paper No. 6 - 3, Pages 46 - 52, 1965.

12. Cramer, H. "Mathematical Methods of Statistics" Princeton University Press, Princeton, New Jersey, 1951.

APPENDIX A

Basic Equations for Digital Computation

Program 1

Simulating an unknown plant,

$$Zn = \sum_{k=0}^{K} X_{n-k} h_k \qquad \text{A. 1}$$

$$Yn = Zn + Vn \qquad \text{A. 2}$$

where Vn is gaussian and white with zero mean and variance .01.

Do correlation

$$Xn = X_{n-2} \oplus X_{n-3} \qquad \text{A. 3}$$

$\oplus$: Modulo 2 addition

$$\hat{\phi}_{xx}(i) = \frac{1}{N+1} \sum_{n=o}^{N} X(n)\, X(n+i)$$

$$\hat{\phi}_{xy}(i) = \frac{1}{N+1} \sum_{n=o}^{N} X(n)\, Y(n+i) \qquad \text{A. 4}$$

Program 2

Pseudo Wiener Hopf Equation.

$$\hat{\underline{h}} = \hat{\phi}_{xx}^{-1}\, \hat{\phi}_{xy}(i) \qquad \text{A. 5}$$

Program 3

Bayesian Estimate

$$\underline{h}_k^* = I_k^{*-1} (\hat{h}_k \hat{I}_k + I^*_{k-1} h^*_{k-1}) \qquad \text{A. 6}$$

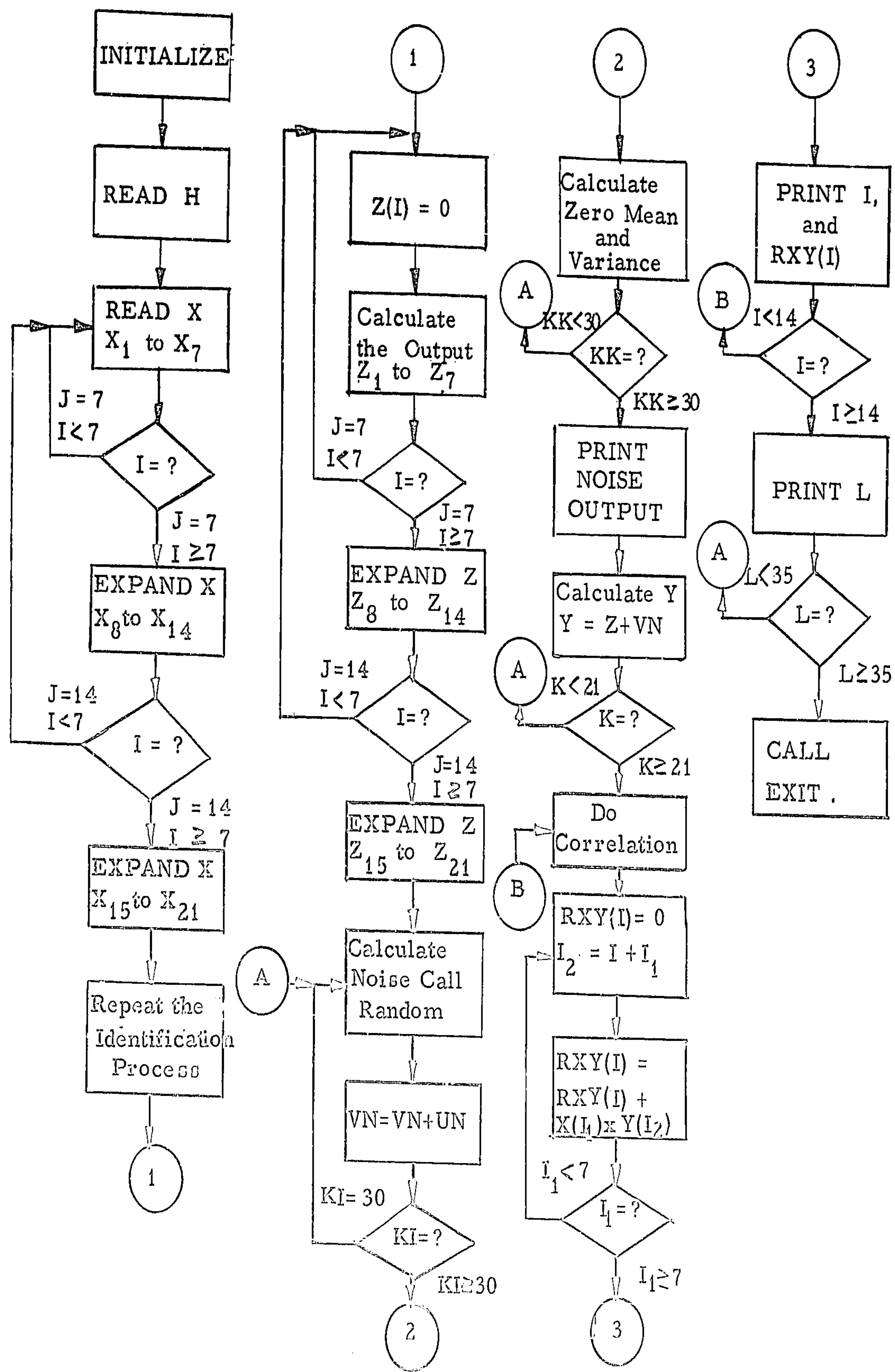
where

$$I_k^* = \hat{I}_k + I^*_{k-1}$$

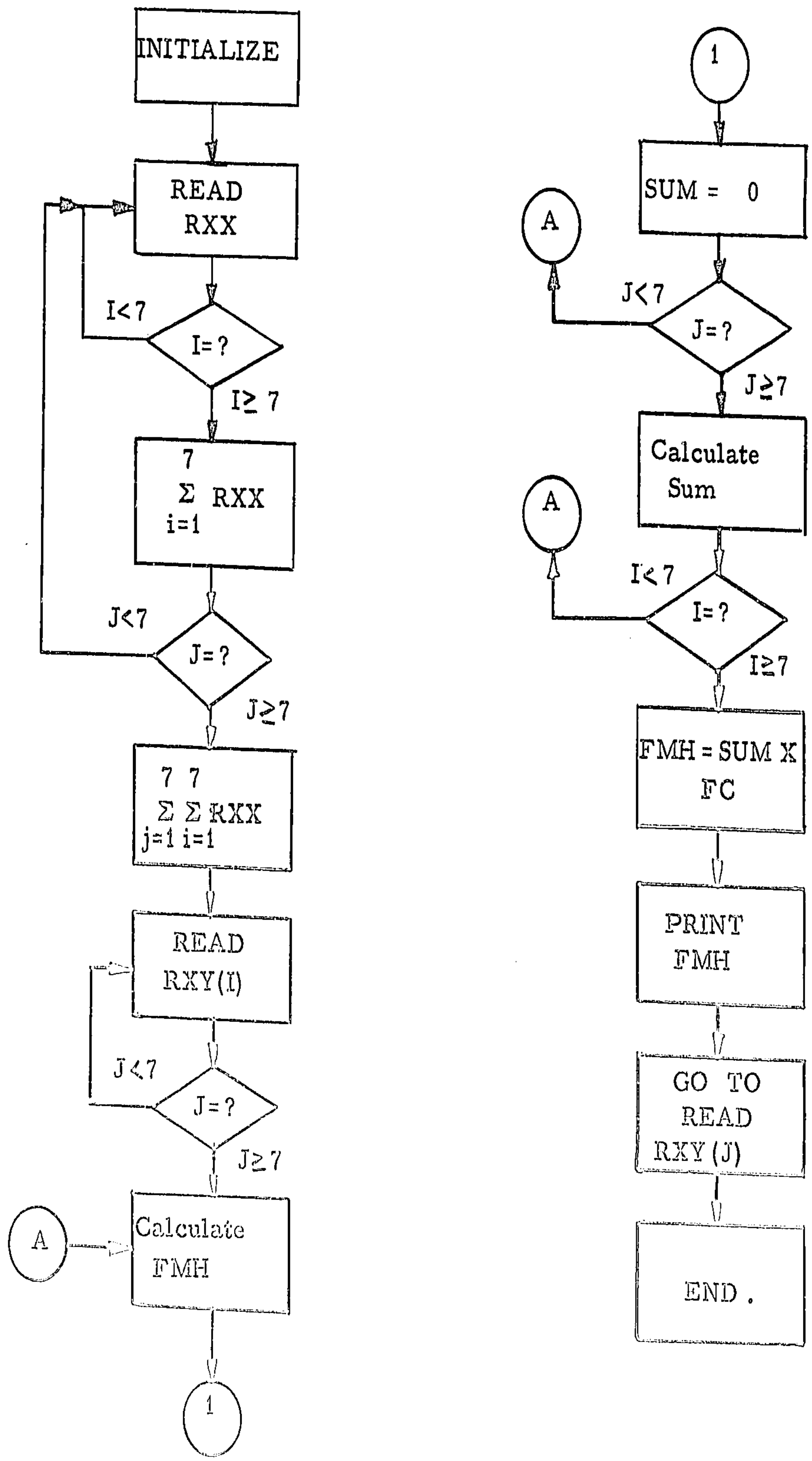
APPENDIX B

Flow Charts for Digital Computer Simulation

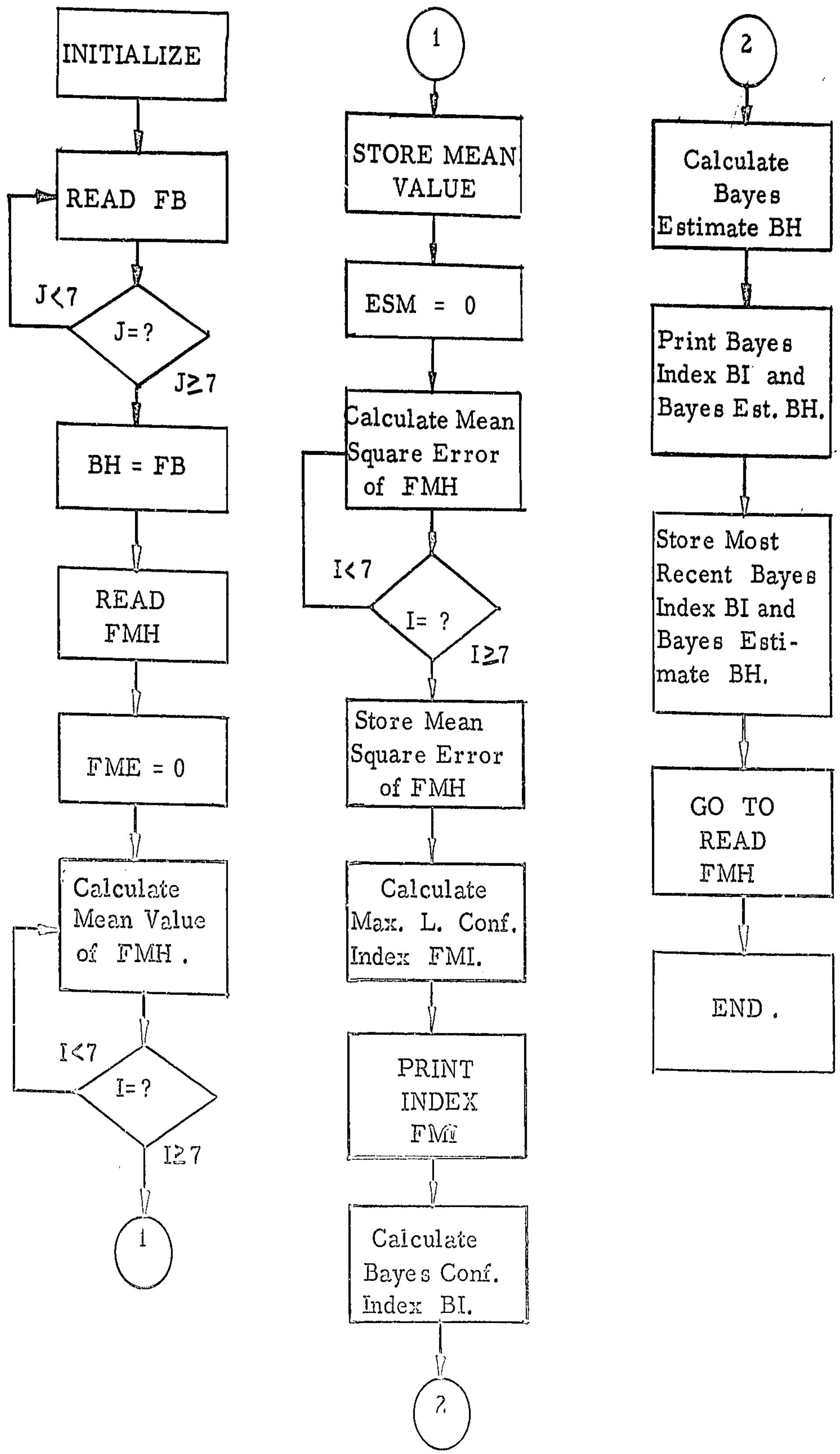
Flow Chart for Program 1.



Flow Chart for Program 2.



Flow Chart for Program 3.



VITAE

NAME:		Chun Loo
BORN:		Hon-How, Kung-Tung, China, May 18, 1937.
EDUCATION:		
	Primary:	La Tuque High School, La Tuque, P. Q,
	Secondary:	La Tuque High School, La Tuque, P. Q.
	University:	McGill University, Montreal, P. Q.
	Degree:	Bachelor of Engineering, (B. Eng)
	Option:	Electrical