MASTER

Parameter estimation of linear processes using self-adjusting models nonlinear-in-the-parameters

Nicola, V.F.

Award date:
1979

Group Measurement and Control
Department of Electrical Engineering
EINDHOVEN UNIVERSITY OF TECHNOLOGY
Eindhoven,The Netherlands

PARAMETER ESTIMATION OF LINEAR PROCESSES USING SELF-ADJUSTING MODELS NONLINEAR-IN-THE-PARAMETERS

by V.F. Nicola

Submitted in partial fulfillment of the requirement for the degree of Ir. (M.Sc.) at the Eindhoven University of Technology. This work was carried out from December 1977 until December 1978 in the professional group measurement and control under the supervision of Prof.Dr.Ir. P. Eykhoff and Ir. H.H. van de Ven.

Abstract

The parameters of a rational time-discrete transfer function of a linear process are to be estimated using an identical model nonlinear-in-the-parameters. The model parameters are adjusted iteratively through the minimization of a least squares error criterion function. Different iterative minimization methods are discussed and compared.
The input signal for estimation is analysed and found to have a major effect on the shape of the error criterion in the parameter space.
The choice of a suitable input signal is simple and practical; this will result in a desirable shape of the error criterion function avoiding us a constrained minimization problem and leading to better convergence properties. The conjugate gradient method is adopted for the minimization of the error criterion function and found to be much superior to the steepest descent method.

Acknowledgement

I wish to express my sinceer appreciation for the special attention and patience shown by Prof. Dr. Ir. P. Eykhoff. I am very thankful to Ir. H. H. van de Ven for many valuable advices and proper guidance.
A pronounced effect on many parts of this work has been due to fruitful and enjoing discussions with Ir. A. A. H. Damen to whom I am particularly indebted. Assistance and useful discussions with colleagues and other members of the professional group measurement and control are acknowledged with pleasure. I am very grateful to Miss. Josefien Braam for her great help in the preparation of the report. The patience and the moral support of my parents is deeply appreciated.

Contents

1. Introduction:

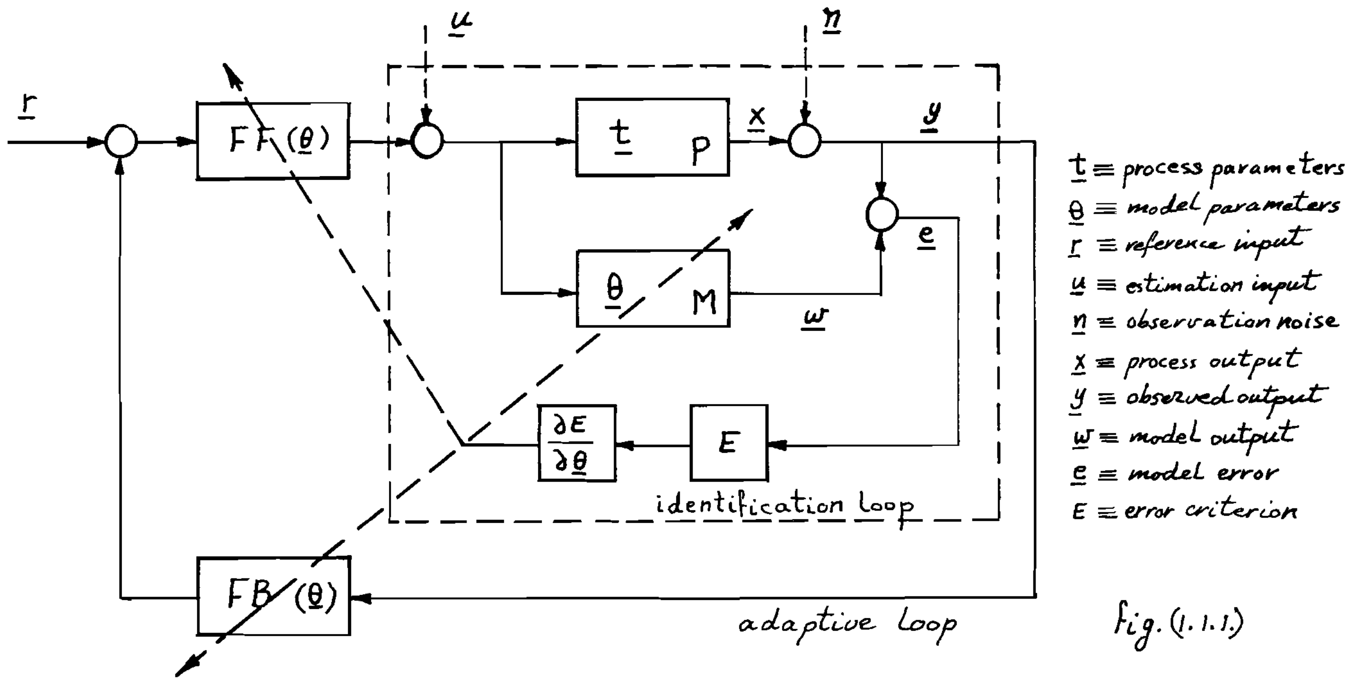


fig. (1.1.1)

Parameter estimation techniques have shown to be of importance for modern applications in a variety of fields.The vast development and the fine circuits integration in the fields of computers and digital signal processing made these techniques available for practical applications.

One such application is in an adaptive loop where the controller parameters should follow the (relatively slow) changes in the process parameters.This can be achieved by building an estimation model which is subjected to the same input signal (may be superimposed to the normal operating input to the process) as the process.Stationary process and model outputs are compared and the error together with its sensitivity w.r.t. the model parameters are used to adjust the model parameters as to approach the process parameters. This scheme is shown in figure (1.1.1.).In this report,we are concerned with the parameter estimation problem where the process is assumed to be in a stable operating condition .If a digital computer is used as a tool for the estimation of the process parameters,then the input/output observation samples are used to obtain an estimate for a digitally simulated model.The equivalent analog parameters can be obtained utilizing a suitable transformation technique.

If the model is linear-in-the-parameters,an explicit relation for the estimate in terms of the observations is obtainable.

When the model is nonlinear-in-the-parameters,the estimate is approached iteratively using "hill climbing" techniques on a suitably chosen error criterion.

The case of models linear-in-the-parameters has been extensively discussed in the literature.Therefore,in this report,we devote our attention to the case of models nonlinear-in-the-parameters as it represents an important class of processes.We have chosen an identical model of a linear time-discrete transfer function as an example of a model nonlinear-in-the-parameters.A comparison with generalized model linear-in-the-parameters is of interest and will be presented.As a prior knowledge about the observation noise is assumed not to be available,we adopt a least squares error criterion;the properties of the estimates will be discussed.The frequency contents of the input signal is expected to influence the estimation problem;therefore,a detailed analysis of the input signal spectrum and its effects on the error criterion function is of much importance and will be investigated.The choice of the iterative minimization method will undoubtfully have a considerable effect on the convergence of the model adjustment process;consequently,we devote part of this report to the presentation and comparision of different minimization methods.

Some of the ideas and results of investigations will be used for the implementation of the model adjustment estimation routines.These routines will be experimented on parameter estimation of a second order linear time-discrete process and conlusions will be derived.

2. Statistical Estimation From Observations:

The parameter estimation problem is recongnized to be the determination of an estimate $\underline{\hat{\theta}}$ for the true process parameters $\underline{t}$. The estimate may be either an explicit or an implicit function of the observations $\underline{y}$ taken from the process. If these observations are stochastic in nature (or contaminated with stochastic disturbances), then the estimate will be stochastic as well and the estimation problem can, best, be viewed as a statistical problem. Depending on the available prior knowledge about the statistical distribution of the parameters $\underline{t}$ and the observations $\underline{y}$ we obtain different classes of estimates as will be shown.

2.1. Some important properties of estimates:

The statistical distribution in the form of a probability density function provides complete information about the statistical properties of the estimate. This probability density function is dependent on the length of the observation interval (or equivalently, dependent on the number of observation samples k used for estimation), see fig. 2.1.1.

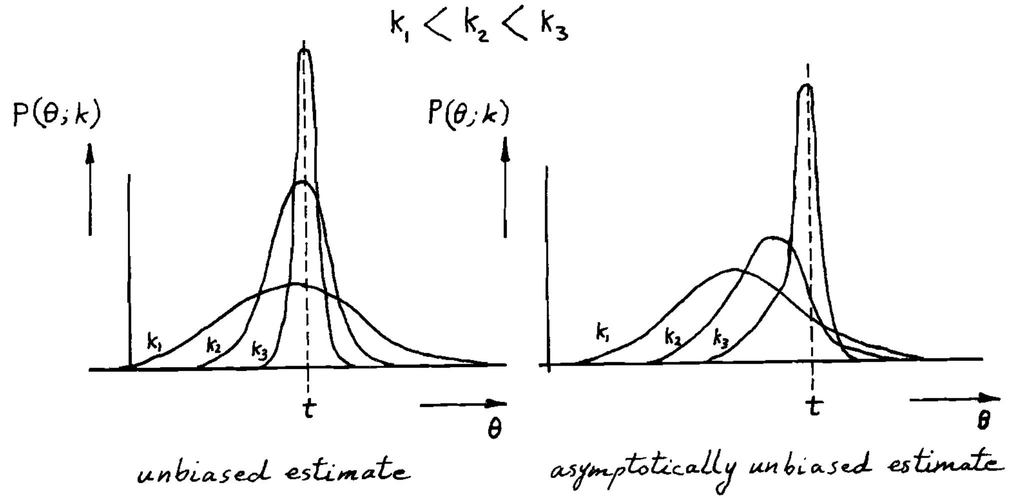


fig. 2.1.1.

For reasons of convenience we will consider only the most significant practical properties of the distribution; namely, its mean $\mathcal{E}\{\hat{\underline{\theta}}\}$ and its covariance cov $\{\hat{\underline{\theta}}\}$;they define a Gaussian distribution completely.
In the following we define some important properties of estimates,

- Bias: is defined to be the expected deviation from the true parameter value $\underline{t}$

$$\mathcal{E}\{\hat{\underline{\theta}}-\underline{t}\} = \mathcal{E}\{\hat{\underline{\theta}}\} - \underline{t}$$

The estimate is unbiased if for each k

$$\mathcal{E}\{\hat{\underline{\theta}}\} = \underline{t}$$

The estimate is asymptotically unbiased if

$$\lim_{k\to\infty} \mathcal{E}\{\hat{\underline{\theta}}\} = \underline{t}$$

- Consistency: the estimate is consistent if

$$\lim_{k\to\infty} P(|\hat{\underline{\theta}}-\underline{t}| > \epsilon) = 0$$

with ϵ arbitrarily small, i.e. $\hat{\underline{\theta}}$ converges in probability to the true value $\underline{t}$. For unbiased (or asymptotically unbiased) estimate, the condition

$$\lim_{k\to\infty} \operatorname{cov}\{\hat{\underline{\theta}}\} = 0\,I \quad , (I \text{ is the identity matrix})$$

implies the consistency of the estimate.

Note that the condition

$$\lim_{k\to\infty} \hat{\underline{\theta}} = \underline{t}$$

implies the asymptotic unbiasedness and the consistency of the estimate $\hat{\underline{\theta}}$.

- Efficiency: the estimate is efficient if it has the minimum theoretically attainable covariance

$$\operatorname{cov}\{\hat{\underline{\theta}}\} \leqslant \operatorname{cov}\{\hat{\underline{\gamma}}\} \longrightarrow \operatorname{Det.}[\operatorname{cov}\{\hat{\underline{\gamma}}\} - \operatorname{cov}\{\hat{\underline{\theta}}\}] \geqslant 0$$

where $\hat{\underline{\gamma}}$ is any other estimate, provided that $\mathcal{E}\{\hat{\underline{\theta}}\} = \mathcal{E}\{\hat{\underline{\gamma}}\}$.

The estimate is asymptotically efficient if

$$\lim_{k\to\infty} \operatorname{cov}\{\hat{\underline{\theta}}\} \leqslant \operatorname{cov}\{\hat{\underline{\gamma}}\}$$

- Sufficiency: the estimate is sufficient if for all other estimates $\hat{\underline{Y}}$ it is true that $p(\hat{\underline{Y}} \mid \hat{\underline{\theta}})$ is independent of $\underline{t}$.

- Normality: the estimate is normal if the probability distribution $p(\hat{\underline{\theta}};\underline{t})$ is Gaussian (normal distribution).
  The estimate is asymptotically normal if $p(\hat{\underline{\theta}};\underline{t})$ approaches Gaussian distribution for $k \to \infty$.

In the following we discuss the derivation of different classes of estimates according to the available prior knowledge. We start with the case of most available prior knowledge and further discuss successively the cases of less available prior knowledge.

2.2. Bayesian Estimation: (BE)

We are given the following.
The prior probability distribution of the parameters $p(\underline{t})$, and the conditional probability distribution of the observations $p(\underline{y} \mid \underline{t})$
Following Bayes' rule we can write

$$P(\underline{y} \mid \underline{t})\; P(\underline{t}) = P(\underline{y}, \underline{t}) = P(\underline{t} \mid \underline{y})\; P(\underline{y})$$

Consequently, a posterior probability distribution of the parameters $\underline{t}$ can be written as follows

$$P(\underline{t} \mid \underline{y}) = \frac{P(\underline{y} \mid \underline{t})\; P(\underline{t})}{P(\underline{y})}$$

Since

$$P(\underline{y}) = \int\!\!\int_{-\infty}^{+\infty} \cdots\cdots \int P(\underline{t};\underline{y})\; d\underline{t}$$

is not a function of $\underline{t}$, we may write $p(\underline{t} \mid \underline{y})$ as a function of $\underline{t}$

$$P(\underline{t} \mid \underline{y}) = L(\underline{t};\underline{y})\; P(\underline{t})$$

where $L(\underline{t};\underline{y})$ is the likelihood of $\underline{t}$ given $\underline{y}$ and defined to be

$$L(\underline{t};\underline{y}) \propto P(\underline{y} \mid \underline{t})$$

A Baysian estimate is derived by maximizing the posterior probability distribution $p(\underline{t} \mid \underline{y})$ w.r.t. $\underline{t}$. By equating the derivative to zero, then

$$\frac{\partial}{\partial \underline{t}} \{ P(\underline{t} \mid \underline{y}) \} \Big|_{\underline{t} = \hat{\underline{\theta}}} = \frac{\partial}{\partial \underline{t}} \{ L(\underline{t}; \underline{y}) \, P(\underline{t}) \} \Big|_{\underline{t} = \hat{\underline{\theta}}} = \underline{0} \longrightarrow \hat{\underline{\theta}}_{BE}$$

from which the Baysian estimate (BE) follows.

2.3. Maximum Likelihood Estimation: (MLE)

We are given the conditional probability distribution $p(\underline{y} \mid \underline{t})$ with the absence of the prior knowledge about the parameters. The MLE follows from the BE by assuming a uniform distribution over the interval under consideration. Hence

$$P(\underline{t} \mid \underline{y}) = L(\underline{t}; \underline{y}) \, P(\underline{t}) \propto L(\underline{t}; \underline{y})$$

since $\qquad P(\underline{t}) = \text{constant}$

and the MLE is derived by maximizing the likelihood $L(\underline{t};\underline{y})$ w.r.t. $\underline{t}$, or laternatively, maximizing $\ln L(\underline{t};\underline{y})$, (since the logarithmic function is monotonic).

Taking the derivative w.r.t. $\underline{t}$ and equating to zero at $\underline{t} = \hat{\underline{\theta}}$; then

$$\frac{\partial}{\partial \underline{t}} \{ \ln L(\underline{t}; \underline{y}) \} \Big|_{\underline{t} = \hat{\underline{\theta}}} = \underline{0} \longrightarrow \hat{\underline{\theta}}_{MLE}$$

from which the Maximum Likelihood Estimate (MLE) follows.

Here we mention in brief some interesting properties of the MLE:

- Asymptotic normality: the probability distribution $p(\hat{\underline{\theta}}_{MLE};\underline{t})$ approaches a normal distribution for $k \to \infty$.
- Asymptotic efficiency: minimum variance (or best accuracy) for $k \to \infty$

  $$\text{Det.} [\, cov\{\hat{\underline{\gamma}}\} - cov\{\hat{\underline{\theta}}_{MLE}\} \,] \geqslant 0$$

  where $\hat{\underline{\gamma}}$ is any other estimate, provided that $\mathcal{E}\{\hat{\underline{\gamma}}\} = \mathcal{E}\{\hat{\underline{\theta}}_{MLE}\}$
- Asymptotic unbiasedness: $\lim_{k \to \infty} \mathcal{E}\{\hat{\underline{\theta}}_{MLE}\} = \underline{t}$

  i.e. the bias in the estimate tends to zero with increasing number of observations.
- Consistency: $\qquad \lim_{k \to \infty} P(|\hat{\underline{\theta}}_{MLE} - \underline{t}| > \epsilon) = 0$

  with ϵ arbitrarily small. Since the estimate is asymptotically unbiased then consistency implies

$$\lim_{k \to \infty} \operatorname{cov} \{ \hat{\underline{\theta}}_{MLE} \} = 0 \; I$$

- Invariance: if $\hat{\underline{\theta}}_{MLE}$ is a MLE of $\underline{t}$, then $g(\hat{\underline{\theta}}_{MLE})$ is a MLE of $g(\underline{t})$.

2.4. Markov Estimation:(ME)(or Weighted Least Squares Estimation)

Let the observations be related to the parameters $\underline{t}$ as follows

$$\underline{y} = \underline{f}(\underline{t}) + \underline{n}$$

where $\underline{f}$ $(\underline{t})$ is the undisturbed process output and $\underline{n}$ is a stochastic disturbance with a given covariance (N).

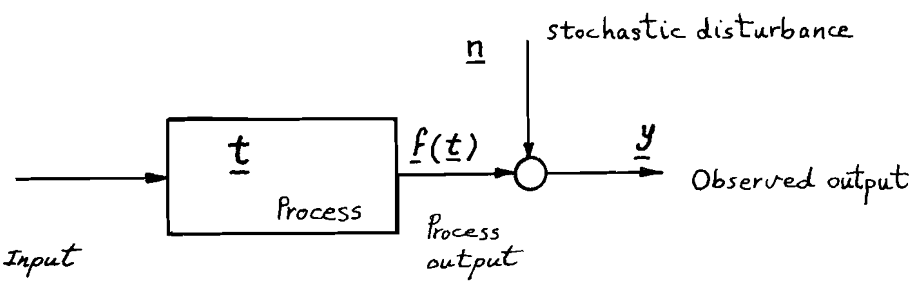


Fig. 2.4.1.

The Markov estimate follows from MLE by assuming that $\underline{n}$ has a Gaussian distribution with zero mean, thus

$$\mathcal{E}\{\underline{n}\} = \underline{0} \quad \text{and} \quad \operatorname{cov}\{\underline{n}\} = \mathcal{E}\{\underline{n}\,\underline{n}^T\} = N$$

then probability distribution $p(\underline{n})$

$$P(\underline{n}) = \frac{1}{((2\pi)^k |N|)^{1/2}} \exp\left[-\frac{1}{2}(\underline{n}^T N^{-1} \underline{n})\right]$$

where k is the number of components in the vector $\underline{n}$.
The conditional probability distribution $p(\underline{y} \mid \underline{t})$ now can be written

$$P(\underline{y} \mid \underline{t}) = \frac{1}{((2\pi)^k |N|)^{1/2}} \exp\left[-\frac{1}{2}(\underline{y}-\underline{f}(\underline{t}))^T N^{-1}(\underline{y}-\underline{f}(\underline{t}))\right]$$

taking the logarithm, then

$$\ln P(\underline{y} \mid \underline{t}) = -\frac{1}{2}\ln[(2\pi)^k |N|] - \frac{1}{2}(\underline{y}-\underline{f}(\underline{t}))^T N^{-1}(\underline{y}-\underline{f}(\underline{t}))$$

and the ME is derived by minimizing the weighted sum of squares given by the expression

$$(\underline{y} - \underline{f}(\underline{t}))^T N^{-1} (\underline{y} - \underline{f}(\underline{t}))$$

w.r.t. $\underline{t}$ (This follows from MLE by maximizing $p(\underline{y} \mid \underline{t})$ w.r.t. $\underline{t}$).

Taking the derivative of the last expression w.r.t. $\underline{t}$ and equating to zero at $\underline{t} = \hat{\underline{\theta}}$, then

$$\left(\frac{\partial}{\partial \underline{t}} \{\underline{f}^T(\underline{t})\} N^{-1} \underline{y} - \frac{\partial}{\partial \underline{t}} \{\underline{f}^T(\underline{t})\} N^{-1} \underline{f}(\underline{t}) \right) \Big|_{\underline{t} = \hat{\underline{\theta}}} = \underline{0} \longrightarrow \hat{\underline{\theta}}_{ME}$$

from which the Markov Estimate (ME) follows.

Considerable simplification is gained if $\underline{f}(\underline{t})$ is linear function of $\underline{t}$. Let

$$\underline{f}(\underline{t}) = \Omega \, \underline{t}$$

an explicit solution for $\hat{\underline{\theta}}$ can readily be written

$$\hat{\underline{\theta}} = (\Omega^T N^{-1} \Omega)^{-1} \Omega^T N^{-1} \underline{y}$$

This estimate is linear w.r.t. the observation vector $\underline{y}$, its properties can be derived in a similar manner as for the least squares estimate which will be considered in the next section.

If $\underline{f}(\underline{t})$ is a nonlinear function of $\underline{t}$, then the estimate follows by solving a set of nonlinear equations; this is usually done using iterative methods or alternatively by using "hill climbing" techniques to find the minimum of the weighted sum of squares. The properties of this implicit solution are discussed in section (3.2.).

2.5. Least Squares Estimation: (LSE)

In the case when we have no prior knowledge about the stochastic disturbance $\underline{n}$, a Least Squares Estimate follows from MLE by assuming that $\underline{n}$ is "white" Gaussian noise with zero mean

i.e. $\mathcal{E}\{\underline{n}\} = \underline{0}$, and

$$\text{cov}\{\underline{n}\} = \mathcal{E}\{\underline{n}\,\underline{n}^T\} = \sigma^2 I$$

The conditional probability distribution $p(\underline{y} \mid \underline{t})$ can be written

$$p(\underline{y} \mid \underline{t}) = \frac{1}{((2\pi)^k \sigma^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}\left(\underline{y} - \underline{f}(\underline{t})\right)^T\left(\underline{y} - \underline{f}(\underline{t})\right)\right]$$

taking the logarithm, then

$$\ln p(\underline{y} \mid \underline{t}) = -\frac{1}{2}\ln\left[(2\pi)^k \sigma^2\right] - \frac{1}{2\sigma^2}\left(\underline{y} - \underline{f}(\underline{t})\right)^T\left(\underline{y} - \underline{f}(\underline{t})\right)$$

and the LSE is derived by minimizing the sum of squares given by the expression

$$\left(\underline{y} - \underline{f}(\underline{t})\right)^T\left(\underline{y} - \underline{f}(\underline{t})\right)$$

w.r.t. $\underline{t}$ (This follows from MLE by maximizing $p(\underline{y} \mid \underline{t})$ w.r.t. $\underline{t}$) Taking the derivative of the last expression w.r.t. $\underline{t}$ and equating to zero at $\underline{t} = \hat{\underline{\theta}}$.

$$\left(\frac{\partial}{\partial \underline{t}}\{\underline{f}^T(\underline{t})\}\,\underline{y} - \frac{\partial}{\partial \underline{t}}\{\underline{f}^T(\underline{t})\}\,\underline{f}(\underline{t})\right)\Big|_{\underline{t}=\hat{\underline{\theta}}} \longrightarrow \hat{\underline{\theta}}_{LSE}$$

from which the LSE follows.

2.5.1. The linear case:

If $\underline{f}(\underline{t})$ is a linear function of $\underline{t}$, an explicit solution for $\hat{\underline{\theta}}$ can be derived, let

$$\underline{f}(\underline{t}) = \Omega\,\underline{t}$$

then $\hat{\underline{\theta}}$ is given by

$$\hat{\underline{\theta}} = (\Omega^T \Omega)^{-1} \Omega^T \underline{y}$$

This is a linear estimate w.r.t. $\underline{y}$. We will consider some of its properties. The observations $\underline{y}$ are related to the parameters $\underline{t}$ by the following

$$\underline{y} = \Omega\,\underline{t} + \underline{n} \qquad \text{, then}$$

$$\hat{\underline{\theta}} = (\Omega^T \Omega)^{-1} \Omega^T (\Omega\,\underline{t} + \underline{n})$$

$$= \underline{t} + (\Omega^T \Omega)^{-1} \Omega^T \underline{n}$$

If the elements of the matrix Ω and the vector $\underline{n}$ are uncorrelated, then

$$\mathcal{E}\{\hat{\underline{\theta}}\} = \mathcal{E}\{\underline{t}\} + (\Omega^T \Omega)^{-1} \Omega^T\, \mathcal{E}\{\underline{n}\}$$

and if our assumption that $\mathcal{E}\{\underline{n}\}=\underline{o}$ is true, then

$$\mathcal{E}\{\underline{\hat{\theta}}\} = \underline{t}$$

and the estimate is unbiased. An expression for the covariance $\text{cov}\{\underline{\hat{\theta}}\}$ can be derived as follows

$$\text{cov}\{\underline{\hat{\theta}}\} = \mathcal{E}\{(\underline{\hat{\theta}}-\underline{t})(\underline{\hat{\theta}}-\underline{t})^T\}$$

$$= \mathcal{E}\{(\Omega^T\Omega)^{-1}\Omega^T\,\underline{n}\,\underline{n}^T\,\Omega\,(\Omega^T\Omega)^{-1}\}$$

If the elements of Ω and $\underline{n}$ are uncorrelated, then

$$\text{cov}\{\underline{\hat{\theta}}\} = (\Omega^T\Omega)^{-1}\Omega^T N\,\Omega\,(\Omega^T\Omega)^{-1}$$

and if our assumption that $N = \sigma^2 I$ is true, then

$$\text{cov}\{\underline{\hat{\theta}}\} = \sigma^2(\Omega^T\Omega)^{-1}$$

Note that $\lim_{k\to\infty} \text{cov}\{\underline{\hat{\theta}}\} = \text{plim}_{k\to\infty}\left[\frac{\sigma^2}{k}\left(\frac{\Omega^T\Omega}{k}\right)^{-1}\right] = 0\,I$

since $\text{plim}_{k\to\infty}\left[\left(\frac{\Omega^T\Omega}{k}\right)^{-1}\right]$ is finite

This implies consistency if the estimate is unbiased, and we conclude that the LSE possess the properties of the MLE if the assumptions that

$$\mathcal{E}\{\underline{n}\} = \underline{o} \quad \text{and} \quad \mathcal{E}\{\underline{n}\,\underline{n}^T\} = \sigma^2 I$$

are true. (This should be obvious since the LSE is derived form MLE based on these assumptions). The case when the elements of Ω and $\underline{n}$ are correlated is discussed in some detail in section (3.1).

2.5.2. The nonlinear case:

If $\underline{f}(\underline{t})$ is a nonlinear function of $\underline{t}$, a solution for $\underline{\hat{\theta}}$ may not be obtainable by explicit analytical relation and iterative methods are used to solve the set of nonlinear equations in $\underline{\hat{\theta}}$. Another approach to the solution is by using "hill climbing" techniques on the error function $E = (\underline{y} - \underline{f}(\underline{t}))^T(\underline{y} - \underline{f}(\underline{t}))$ to find the value of $\underline{\hat{\theta}}$ at which E is minimum.

The solution $\hat{\underline{\theta}}$ is implicitly expressed in the relation

$$\frac{\partial}{\partial \underline{t}} \left\{ \underline{f}^T(\underline{t}) \right\} \cdot \underline{f}(\underline{t}) \Big|_{\underline{t} = \hat{\underline{\theta}}} = \frac{\partial}{\partial \underline{t}} \left\{ \underline{f}^T(\underline{t}) \right\} \left(\underline{f}(\underline{t}) + \underline{n} \right) \Big|_{\underline{t} = \hat{\underline{\theta}}}$$

where $\underline{y}$ is replaced by $(\underline{f}(\underline{t}) + \underline{n})$. The implicit presence of the term

$$\frac{\partial}{\partial \underline{t}} \left\{ \underline{f}^T(\underline{t}) \right\}$$

in the above relation will lead to a bias in the estimate $\hat{\underline{\theta}}$. It will be shown in section (3.2.) that if the assumption that $\varepsilon\{\underline{n}\} = o$ is true, then the estimate $\hat{\underline{\theta}}$ is consistent.

Analysis and proof of the consistency of nonlinear least squares estimates is found in [9].

3. Models of linear time-discrete transfer functions

3.1 Generalized models linear - in - the - parameters:

The process is given by

$$y(k) = \frac{B(z^{-1})}{A(z^{-1})} \{u(k)\} + n(k)$$

where u(k), y(k) and n(k) are the input, output and noise samples at instant k, respectively.

It is assumed that $\quad \mathcal{E}\{n(i)\} = 0$

$$A(z^{-1}) = 1 + a_1 z^{-1} + \cdots\cdots + a_n z^{-n}$$

$$B(z^{-1}) = b_0 + b_1 z^{-1} + \cdots\cdots + b_m z^{-m} \quad , \quad n \geq m$$

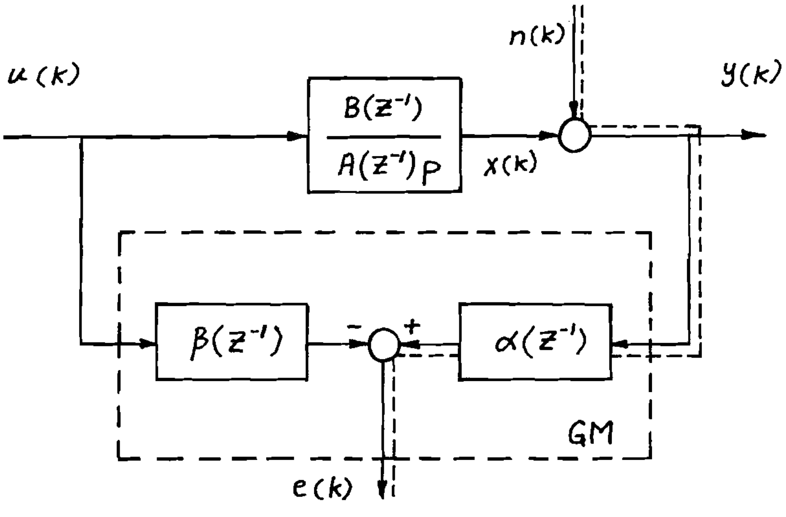


fig. 3.1.1.

The model error e(k) can be chosen to be linear in the model parameters, in such a case

$$e(k) = \alpha(z^{-1})\{y(k)\} - \beta(z^{-1})\{u(k)\}$$

This is called a generalized error linear - in - the - parameters (see fig. 3.1.1.); this type of error presents a considerable simplification on the model implementation and the derivation of the error sensitivity w.r.t. the model parameters. Furthermore, for least squares estimations the following expression

$$E = \underline{e}^T \underline{e} = (\underline{y} - \Omega_{u,y}\,\underline{\theta})^T (\underline{y} - \Omega_{u,y}\,\underline{\theta}) = \sum_{i=n+1}^{n+k} \left(y(i) - \Omega^T_{u,y}(i)\,\underline{\theta} \right)^2$$

where $\underline{e} = \underline{y} - \Omega_{u,y}\,\underline{\theta}$

is to be minimized , *where*

$$\underline{\theta}^T = [\beta_0 , \beta_1 , \ldots\ldots , \beta_m , -\alpha_1 , -\alpha_2 , \ldots\ldots , -\alpha_n] \quad , \text{ and}$$

$$\Omega^T_{u,y}(i) = [u(i), u(i-1), \ldots, u(i-m) , y(i-1), y(i-2), \ldots, y(i-n)]$$

E is a quadratic function in the parameters. The optimal parameters $\hat{\underline{\theta}}$ which minimizes E can be obtained explicitly by the relation

$$\hat{\underline{\theta}} = (\Omega^T_{u,y}\,\Omega_{u,y})^{-1}\,\Omega^T_{u,y}\,\underline{y}$$

or implicitly using "hill climbing" techniques which may be suited for application on a quadratic error criterion. On the other hand, a disadvantage of this representation is that the estimated parameters suffers asymptotic biasedness if the output observations are contaminated with noise; this is due to the fact that the noise effect in the residuals e(i) is correlated with the previous output observations,
y(i-1), y(i-2), ... y(i-n) (i.e. $\underline{e}$ is correlated with the observation matrix $\Omega_{u,y}$). The following analysis will help us to follow the noise effect on the asymptotic bias of the estimate.
The residuals are given by

$$\underline{e} = \underline{y} - \Omega_{u,y}\,\underline{\theta} \quad , \text{ and}$$

$$\underline{y} = \underline{x} + \underline{n} = \Omega_{u,x}\,\underline{t} + \underline{n}$$

where $\underline{t}$ is the true parameter vector

$$\underline{t}^T = [b_0 , b_1 , \ldots\ldots , b_m , -a_1 , -a_2 , \ldots\ldots , -a_n] \quad , \text{then}$$

$$\underline{e} = \Omega_{u,x}\,\underline{t} + \underline{n} - \Omega_{u,x}\,\underline{\theta} - \Omega_{o,n}\,\underline{\theta}$$

$$= \Omega_{u,x}\,\Delta\underline{\theta} + \underline{e}_{n\theta} \quad , \text{ where}$$

$$\Delta\underline{\theta} = \underline{t} - \underline{\theta} \quad , \text{ and}$$

$$\underline{e}_{n\theta} = \underline{n} - \Omega_{o,n}\,\underline{\theta}$$

The first term ($\Omega_{u,x}\ \underline{\Delta\theta}$) is the error due to the difference between the model and the process parameters, the second term ($\underline{e}_{n\theta}$) is the error due to the noise path through part of the generalized model. Now, in terms of the observation matrix we can write

$$\underline{e} = \Omega_{u,y}\ \Delta\underline{\theta} + \underline{e}_{nt} \quad , \text{ where}$$

$$\underline{e}_{nt} = \underline{n} - \Omega_{o,n}\ \underline{t}$$

least squares estimate for $\underline{\Delta\theta}$ when obtained by minimizing $\underline{e}^T\underline{e}$ will lead to

$$\Delta\hat{\underline{\theta}} = -\left(\Omega^T_{u,y}\ \Omega_{u,y}\right)^{-1} \Omega_{u,y}\ \underline{e}_{nt}$$

For asymptotically unbiased estimate for $\underline{t}$, then $\lim_{k\to\infty} \mathcal{E}\{\Delta\hat{\underline{\theta}}\}$ must be equal to zero

$$\lim_{k\to\infty} \mathcal{E}\{\Delta\hat{\underline{\theta}}\} = \operatorname*{plim}_{k\to\infty}\left[\Delta\hat{\underline{\theta}}\right]$$

$$= \operatorname*{plim}_{k\to\infty}\left[\left(\Omega^T_{u,y}\Omega_{u,y}\right)^{-1}\Omega^T_{u,y}\ \underline{e}_{nt}\right]$$

$$= \operatorname*{plim}_{k\to\infty}\left[\frac{\Omega^T_{u,y}\ \Omega_{u,y}}{k}\right]^{-1} \times \operatorname*{plim}_{k\to\infty}\left[\frac{\Omega^T_{u,y}\ e_{nt}}{k}\right]$$

$$= Q^{-1} \times \begin{bmatrix} \varphi_{en}(1) \\ \varphi_{en}(2) \\ \vdots \\ \varphi_{en}(n) \end{bmatrix}$$

$$Q = \operatorname*{plim}_{k\to\infty}\left[\frac{\Omega^T_{u,y}\ \Omega_{u,y}}{k}\right] \quad , \text{ and}$$

$$\varphi_{en}(i) = \varphi_{e_{nt}\,n}(i)$$

is the cross correlation between the noise in the residuals $\underline{e}$ and the noise in the observations $\underline{y}$.

It can be seen that the asymptotic bias can be eleminated if the vector

$$\begin{bmatrix} \varphi_{en}(1) \\ \vdots \\ \varphi_{en}(n) \end{bmatrix} = \underline{0}$$

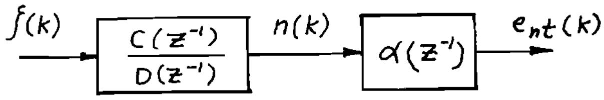


fig. 3.1.2.

i.e. there is no correlation between the samples of the noise residuals $e_{nt}(i)$ and the previous noise samples in the output observations $n(i-1)$, $n(i-2)$, ..., $n(i-n)$.

This condition can be fulfilled only of the noise residuals $e_{nt}(i)$ are uncorrelated i.e. for "white" noise residuals.

If the noise sequence n(k) has some known correlation, e.g. derived from "white" noise sequence through the filter $\frac{C(z^{-1})}{D(z^{-1})}$ as in fig. 3.1.2., we can follow an estimation scheme as in fig. 3.1.3., in which the input and output signals of the process are filtered before they are used to generate the generalized error, parameters are estimated iteratively and used to update the filter parameters. When convergence is achieved filter parameters will be equal to noise and model parameters and the additive noise in the generalized error is effectively "white"; this implies the asymptotic unbiasedness of the estimates for the proposed scheme. However, no proof is available yet for the convergence of this scheme.

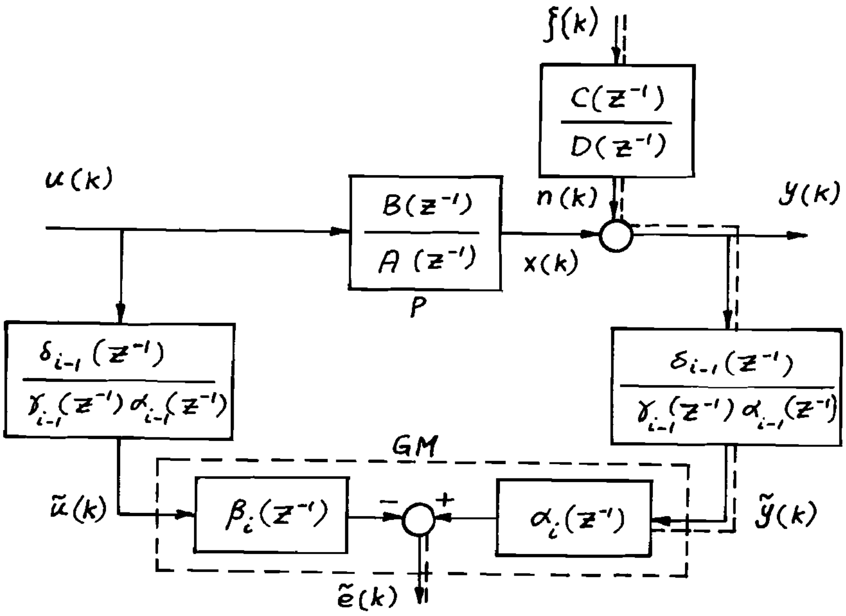


fig. 3.1.3.

The following proofs the consistency of the asymptotically unbiased estimates

$$cov\{\Delta\underline{\hat{\theta}}\} = \mathcal{E}\{\Delta\underline{\hat{\theta}}\cdot\Delta\underline{\hat{\theta}}^T\}$$

$$= \mathcal{E}\{(\Omega_{u,y}^T\Omega_{u,y})^{-1}\Omega_{u,y}^T\underline{e}_{nt}\underline{e}_{nt}^T\Omega_{u,y}(\Omega_{u,y}^T\Omega_{u,y})^{-1}\}$$

$$\lim_{k\to\infty} cov\{\Delta\underline{\hat{\theta}}\} = \operatorname*{plim}_{k\to\infty}[\Delta\underline{\hat{\theta}}\;\Delta\underline{\hat{\theta}}^T]$$

$$= \left(k^{-1}\operatorname*{plim}_{k\to\infty}\left[\frac{\Omega_{u,y}^T\Omega_{u,y}}{k}\right]^{-1}\right)\times\left(k\operatorname*{plim}_{k\to\infty}\left[\frac{\Omega_{u,y}^T\underline{e}_{nt}\underline{e}_{nt}^T\Omega_{u,y}}{k}\right]\right)\times$$

$$\left(k^{-1}\operatorname*{plim}_{k\to\infty}\left[\frac{\Omega_{u,y}^T\Omega_{u,y}}{k}\right]^{-1}\right) = \lim_{k\to\infty}\frac{1}{k}(Q^{-1}NQ^{-1}) = 0\,I$$

where

$$Q = \operatorname*{plim}_{k\to\infty}\left[\frac{\Omega_{u,y}^T\Omega_{u,y}}{k}\right]$$

and

$$N = \operatorname*{plim}_{k\to\infty}\left[\frac{\Omega_{u,y}^T\underline{e}_{nt}\underline{e}_{nt}^T\Omega_{u,y}}{k}\right]$$

I is the identity matrix.

An alternative way for the previous analysis can be done using the following form for the residuals

$$\underline{e} = \underline{u} - \Omega_{u,y}\,\underline{\theta}$$

where

$$\underline{\theta}^T = [-\beta_1,\ldots\ldots,-\beta_m,\alpha_0,\alpha_1,\ldots\ldots,\alpha_n]$$

are the model parameters. $\underline{u}$ is the undisturbed input given by

$$\underline{u} = \Omega_{u,x}\,\underline{t} = (\Omega_{u,y} - \Omega_{o,n})\,\underline{t}$$

where

$$\underline{t}^T = [-b_1,\ldots\ldots,-b_m,a_0,a_1,\ldots\ldots,a_n]$$

are the process parameters, and

$$\Omega_{u,y}^T(i) = [u(i-1),\ldots\ldots,u(i-m),y(i),\ldots\ldots,y(i-n)]$$

least squares estimation of $\underline{\theta}$ obtained by minimizing $(\underline{e}^T.\underline{e})$ and is given by

$$\begin{aligned}\underline{\hat{\theta}} &= (\Omega_{u,y}^{T}\,\Omega_{u,y})^{-1}\,\Omega_{u,y}^{T}\,\underline{u}\\ &= (\Omega_{u,y}^{T}\,\Omega_{u,y})^{-1}\,\Omega_{u,y}^{T}(\Omega_{u,y}-\Omega_{o,n})\,\underline{t}\\ &= \underline{t}-(\Omega_{u,y}^{T}\,\Omega_{u,y})^{-1}\,\Omega_{u,y}^{T}\,\Omega_{o,n}\,\underline{t}\end{aligned}$$

The second term in the last expression causes a bias in the estimate of $\underline{\theta}$

$$\Delta\,\underline{\hat{\theta}} = (\Omega_{u,y}^{T}\,\Omega_{u,y})^{-1}\,\Omega_{u,y}^{T}\,\underline{e}_{nt}$$

where $\underline{e}_{nt} = \Omega_{o,n}\,\underline{t}$ is the noise in the residuals $\underline{e}$

$$\begin{aligned}\lim_{k\to\infty}\mathcal{E}\{\Delta\underline{\hat{\theta}}\} &= \operatorname*{plim}_{k\to\infty}[\Delta\,\underline{\hat{\theta}}]\\ &= \operatorname*{plim}_{k\to\infty}\left[(\Omega_{u,y}^{T}\,\Omega_{u,y})^{-1}\,\Omega_{u,y}^{T}\,\underline{e}_{nt}\right]\\ &= \left(\operatorname*{plim}_{k\to\infty}\left[\frac{\Omega_{u,y}^{T}\,\Omega_{u,y}}{k}\right]^{-1}\right)\times\left(\operatorname*{plim}_{k\to\infty}\left[\frac{\Omega_{u,y}^{T}\,\underline{e}_{nt}}{k}\right]\right)\\ &= Q^{-1}\times\begin{bmatrix}\psi_{en}(1)\\ \vdots\\ \psi_{en}(n)\end{bmatrix}\end{aligned}$$

where $$Q = \operatorname*{plim}_{k\to\infty}\left[\frac{\Omega_{u,y}^{T}\,\Omega_{u,y}}{k}\right]$$

and $\psi_{en}(i) = \psi_{e_{nt}n}(i)$ is the cross correlation between the noise in the residuals $\underline{e}$ and the noise in the observations $\underline{y}$. This is the same result as obtained before; for bias free estimate the noise in the residuals should be uncorrelated (i.e. "white" noise residuals). The consistency of the estimate follow in the same way as derived before.

<u>Parameter sensitivity model</u>: see fig. 3.1.4.

The generalized error is given by

$$\begin{aligned}\tilde{e}(k) &= (1+\alpha_1 z^{-1}+\cdots+\alpha_n z^{-n})\{\tilde{y}(k)\}-(\beta_0+\beta_1 z^{-1}+\cdots+\beta_m z^{-m})\{\tilde{u}(k)\}\\ &= \tilde{y}(k)+\alpha_1\tilde{y}(k-1)+\cdots+\alpha_n\tilde{y}(k-n)-\beta_0\tilde{u}(k)-\beta_1\tilde{u}(k-1)-\cdots-\beta_m\tilde{u}(k-m)\end{aligned}$$

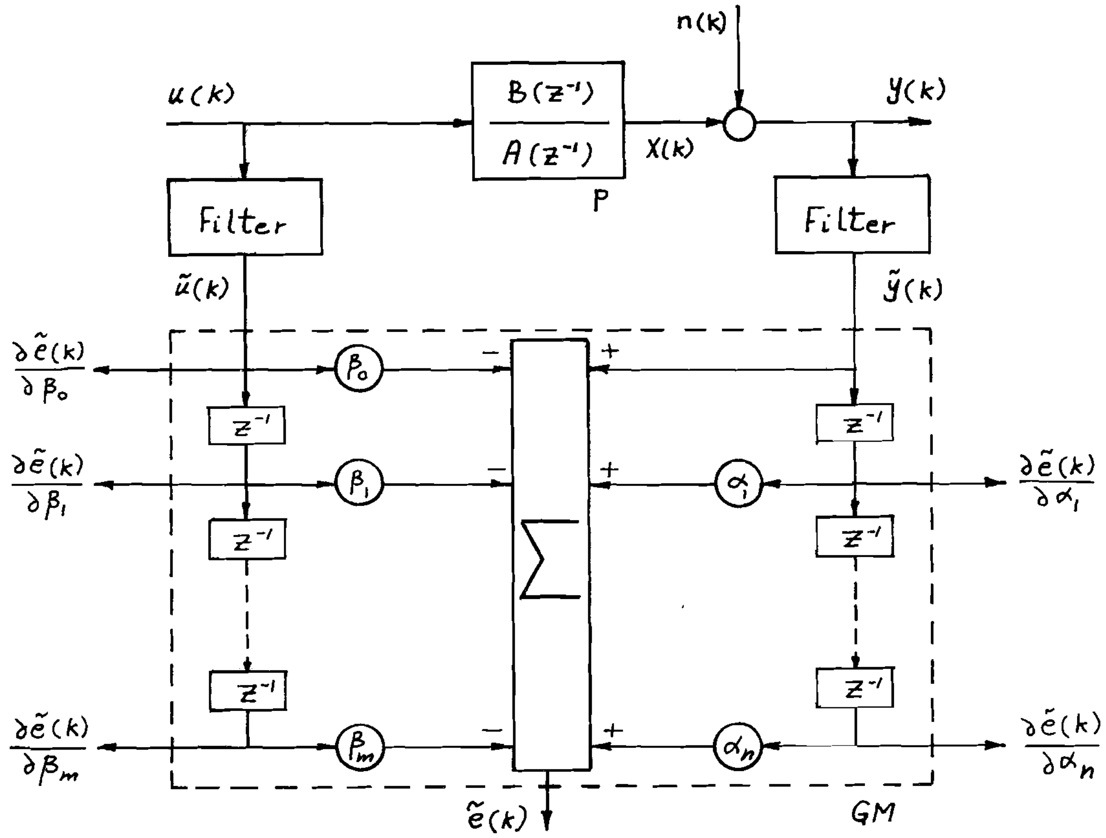


fig. 3.1.4.

It can be easily implemented on a digital computer. The error sensitivity w.r.t. the parameters is given by

$$\frac{\partial}{\partial \alpha_i}\{\tilde{e}(k)\} = y(k-i) \quad , \quad i = 1, \ldots\ldots, n$$

$$\frac{\partial}{\partial \beta_j}\{\tilde{e}(k)\} = u(k-j) \quad , \quad j = 0, \ldots\ldots, m$$

Those sensitivities are already explicitly present in the model as shown in fig. 3.1.4.; this is a considerable simplicity gained by using the generalized model representation for a process.

3.2. Models, nonlinear-in-the-parameters:

Some advantages concerning the properties of the estimates can be gained when the model error is given by

$$e(k) = y(k) - \frac{\beta(z^{-1})}{\alpha(z^{-1})}\{u(k)\}$$

The model error e(k) is nonlinear in some of the model parameters, namely $\alpha_1, \alpha_2, \ldots, \alpha_n$, but it is linear in the parameters $\beta_o, \beta_1, \ldots, \beta_m$. For least squares estimation, the expression

$$E_N = \sum_{k=1}^{N} e^2(k) = \underline{e}^T \underline{e}$$

is to be minimized, where

$$\begin{aligned} \underline{e} &= [e(1), \ldots\ldots\ldots, e(N)]^T \\ &= \underline{y} - \underline{w} \qquad = \underline{y} - U \underline{f}(\underline{\beta}, \underline{\alpha}) \\ &= \underline{y} - U \underline{f}(\underline{\theta}) \end{aligned}$$

where the model output $\underline{w}$ is represented as a linear function of the input observation matrix U, and $\underline{f}$ is a vector whose elements are functions linear-in-the-parameters $\underline{\beta}$ but nonlinear-in-the-parameters $\underline{\alpha}$, $\underline{\theta}^T = [\underline{\beta}, \underline{\alpha}]$. The model error can also be written as follows

$$\begin{aligned} \underline{e} &= U \underline{f}(\underline{t}) + \underline{n} - U \underline{f}(\underline{\theta}) \\ &= U \, \Delta \underline{f}(\underline{\theta}) + \underline{n} \end{aligned}$$

where

$$U \Delta \underline{f}(\underline{\theta}) = U (\underline{f}(\underline{t}) - \underline{f}(\underline{\theta}))$$

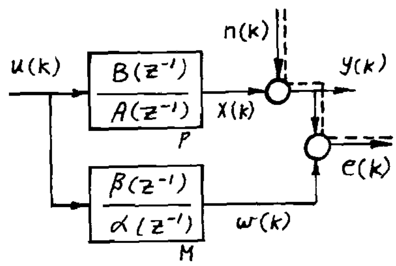


Fig. 3.2.1.

is the error due to the missmatch of the model parameters and the process parameters $\underline{t}$, and $\underline{n}$ is the noise error in the process output. For a finite number of observations used in the estimation of the parameters then

$$E_N = \underline{e}^T \underline{e} = \Delta \underline{f}^T U^T U \Delta \underline{f} + 2 \Delta \underline{f}^T U^T \underline{n} + \underline{n}^T \underline{n}$$

For a minimum of E_N w.r.t. $\underline{\theta}$

$$\frac{\partial}{\partial \underline{\theta}} \{E_N\} = 2 \frac{\partial}{\partial \underline{\theta}} \{\Delta \underline{f}^T\} U^T U \Delta \underline{f} + 2 \frac{\partial}{\partial \underline{\theta}} \{\Delta \underline{f}^T\} U^T \underline{n} = \underline{o}$$

$$\frac{\partial}{\partial \underline{\theta}} \{\Delta \underline{f}^T\} = \frac{\partial}{\partial \underline{\theta}} \{ \underline{f}^T(\underline{\theta}) \} = D^T \quad \text{, hence}$$

$$D^T U^T U \underline{f}(\hat{\underline{\theta}}) = D^T U^T U \underline{f}(\underline{t}) + D^T U^T \underline{n}$$

$\hat{\underline{\theta}}$ can be written explicitly in the following

$$\hat{\underline{\theta}} = \underline{f}^{-1} \{ \underline{f}(\underline{t}) + (D^T U^T U)^{-1} D^T U^T \underline{n} \}$$

where $\underline{f}^{-1}$ is the inverse nonlinear function vector of $\underline{f}$ (e.g. if $f(x)=\frac{1}{x}$, then $f^{-1}(x)$ is such that $f^{-1}\{f(x)\} = x$, i.e. $f^{-1}(x) = \frac{1}{x}$ in this special case).
Note that $\underline{f}^{-1}$ is of different dimension than $\underline{f}$.
$\hat{\underline{\theta}}$ may be expanded in Taylor series as follows

$$\hat{\underline{\theta}} = \underline{f}^{-1}\{\underline{f}(\underline{t})\} + \underline{O}\{\underline{f}(\underline{t}), (D^T U^T U)^{-1} D^T U^T \underline{n}\}$$

where $\underline{O}$ is a nonlinear function vector operating on $\underline{f}(t)$ and the noise term

$$\mathcal{E}\{\hat{\underline{\theta}}\} = \underline{t} + \mathcal{E}\{\underline{O}\{\underline{f}(\underline{t}), (D^T U^T U)^{-1} D^T U^T \underline{n}\}\}$$

It is clear that even if the noise sequence $\underline{n}$ is "white" and of zero mean, the estimate will be biased; this is due to the fact that the nonlinear operations performed on the noise will cause undesirable changes in its characteristics. Simple example will help to illustrate the effect of noise in nonlinear parameter estimation,consider the one dimensional case of a simple nonlinear process with no dynamics, given by

$$y(i) = \frac{1}{b} u(i) + n(i)$$

where n(i) is measurement noise independent of the input u(i) and of zero mean $\mathcal{E}\{n(i)\} = o$. A model is given by

$$w(i) = \frac{1}{\beta} u(i)$$

and the model error is given by

$$e(i) = y(i) - w(i) = \left(\frac{1}{b} - \frac{1}{\beta}\right) u(i) + n(i)$$

a least squares estimate is obtained by minimizing a sum of squares given by

$$E_N = \sum_{i=1}^{N} e^2(i)$$

$$= \sum_{i=1}^{N}\left(\left(\frac{1}{b} - \frac{1}{\beta}\right)^2 \cdot u^2(i) + 2\left(\frac{1}{b} - \frac{1}{\beta}\right)\cdot u(i).n(i) + n^2(i)\right)$$

$$\frac{\partial}{\partial\beta}\{E_N\} = o \quad , \text{ for minimum } E_N$$

$$\frac{\partial}{\partial\beta}\{E_N\} = \sum_{i=1}^{N}\left(2\,u^2(i)\left(\frac{1}{b}-\frac{1}{\beta}\right)\cdot\frac{1}{\beta^2} + 2\,u(i)\,n(i)\,\frac{1}{\beta^2}\right)$$

$$= \frac{2}{\beta^2}\sum_{i=1}^{N}\left(u^2(i)\left(\frac{1}{b}-\frac{1}{\beta}\right) + u(i)\,n(i)\right) = 0 \longrightarrow \hat{\beta}$$

$$\frac{1}{\hat{\beta}} = \frac{1}{b} + \frac{\sum_{i=1}^{N} u(i)\,n(i)}{\sum_{i=1}^{N} u^2(i)} = \frac{1}{b} + \delta(\underline{u},\underline{n})$$

where

$$\delta(\underline{u},\underline{n}) = \sum_{i=1}^{N} u(i)\,n(i) \Big/ \sum_{i=1}^{N} u^2(i)$$

is a linear function of $\underline{n}$, then

$$\hat{\beta} = b \big/ \left(1 + b\,\delta(\underline{u},\underline{n})\right)$$

$$= b\left(1 - b\,\delta(\underline{u},\underline{n}) + b^2\,\delta^2(\underline{u},\underline{n}) - b^3\,\delta^3(\underline{u},\underline{n}) + \cdots\cdots\right)$$

$$\mathcal{E}\{\hat{\beta}\} = b - b^2\,\mathcal{E}\{\delta(\underline{u},\underline{n})\} + b^3\,\mathcal{E}\{\delta^2(\underline{u},\underline{n})\} - b^4\,\mathcal{E}\{\delta^3(\underline{u},\underline{n})\} + \cdots$$

Since $\mathcal{E}\{\delta(\underline{u},\underline{n})\} = 0$ because is linear in $\underline{n}$, then

$$\mathcal{E}\{\hat{\beta}\} = b + b^3\,\mathcal{E}\{\delta^2(\underline{u},\underline{n})\} - b^4\,\mathcal{E}\{\delta^3(\underline{u},\underline{n})\} + \cdots$$

and the estimates are biased, since $\mathcal{E}\{\delta^P(\underline{u},\underline{n})\} \neq 0$ for $P > 1$

Now consider the asymptotic case when the number of observations N grows unlimited (Here $\lim_{N\to\infty}$ implies that N grows finitely large)

$$E = \lim_{N\to\infty} E_N = \lim_{N\to\infty} \sum_{i=1}^{N} e^2(i)$$

$$= \lim_{N\to\infty} \sum_{i=1}^{N}\left(\left(\frac{1}{b}-\frac{1}{\beta}\right)^2 u^2(i) + 2\left(\frac{1}{b}-\frac{1}{\beta}\right)u(i)\,n(i) + n^2(i)\right)$$

Since

$$\lim_{N\to\infty}\sum_{i=1}^{N} n(i) = \lim_{N\to\infty} N\,\mathcal{E}\{n(i)\} = 0$$

we can write

$$E = \sum_{i=1}^{N}\left(\frac{1}{b}-\frac{1}{\beta}\right)^2 u^2(i) + N\,\mathcal{E}\{n^2(i)\}$$

Now, minimization of E will lead to the estimate $\hat{\beta}=b$, which is unbiased with no variance.

In the multidimensional case, let $\hat{\underline{\theta}}_N$ be the estimate resulting from the minimization of E_N and consider the asymptotic case with large number of observation

$$E = \lim_{N\to\infty} E_N$$
$$= \Delta\underline{f}^T U^T U \Delta\underline{f} + 2N\Delta\underline{f}^T U^T \mathcal{E}\{\underline{n}\} + N\mathcal{E}\{\underline{n}^T\underline{n}\}$$

if $\mathcal{E}\{\underline{n}\} = 0$, then

$$E = \Delta\underline{f}^T U^T U \Delta\underline{f} + N\mathcal{E}\{\underline{n}^T\underline{n}\}$$

and minimization of E will lead to the estimate $\hat{\underline{\theta}} = \underline{t}$ which is unbiased with zero covariance matrix we conclude that if $\hat{\underline{\theta}}_N$ is the estimate based on N observations then

$$\lim_{N\to\infty} \hat{\underline{\theta}}_N = \underline{t}$$

which implies the consistency of the nonlinear least squares estimator. Detailed proofs and analysis of the asymptotic properties of nonlinear least squares estimator can be found in paper [9] .

In comparison with the generalized linear model, the nonlinear model delivers asymptotically consistent estimate with neither the need of filtering the input and output observations, nor the need of extending the estimation problem to include the noise filter parameters.

Parameter sensitivity model:

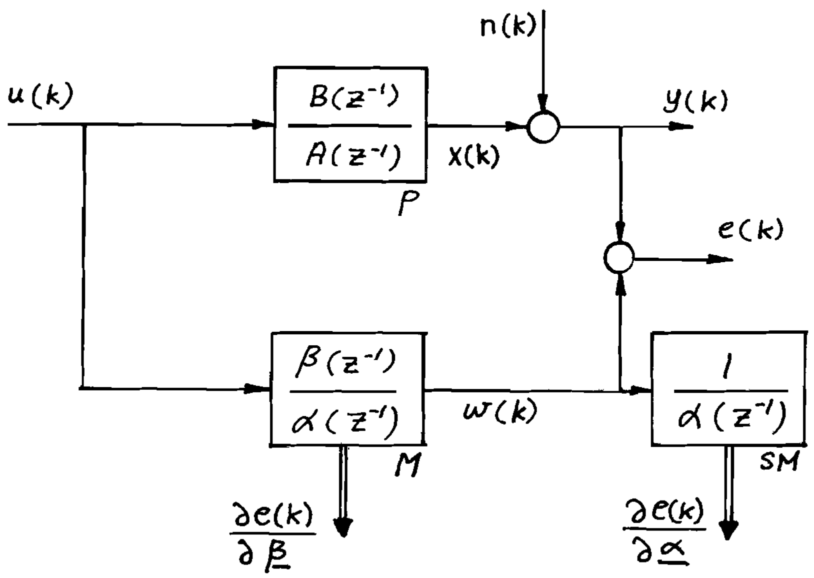


fig. 3.2.2.

As the model is given by its time-discrete transfer function

$$H_M(z) = \frac{w(z)}{u(z)} = \frac{\beta_0 + \beta_1 z^{-1} + \cdots\cdots + \beta_m z^{-m}}{\alpha_0 + \alpha_1 z^{-1} + \cdots\cdots + \alpha_n z^{-n}} = \frac{\beta(z^{-1})}{\alpha(z^{-1})}$$

simulation on a digital computer can easily be implemented. The corresponding difference equation is written as follows

$$(\alpha_0 + \alpha_1 z^{-1} + \cdots\cdots + \alpha_n z^{-n})\{w(k)\} = (\beta_0 + \beta_1 z^{-1} + \cdots\cdots + \beta_m z^{-m})\{u(k)\}$$

w(k) can be calculated from u(k) and previous values of u(i) and w(i) (i < k)

$$w(k) = \frac{1}{\alpha_0}\left(\beta_0 u(k) + \beta_1 u(k-1) + \cdots + \beta_m u(k-m) - \alpha_1 w(k-1) - \cdots\cdots - \alpha_n w(k-n)\right)$$

we can immediately write for the error e(k)

$$e(k) = y(k) - w(k)$$

which is used to determine the value of the error criterion function for certain parameter values

$$E = \sum_{k=k_1}^{k_2} e^2(k)$$

The parameter sensitivity functions are given by

$$\frac{\partial}{\partial \beta_i}\{E\} = 2\sum_{k=k_1}^{k_2} e(k) \cdot \frac{\partial e(k)}{\partial \beta_i}, \quad i = 0, 1, \ldots\ldots, m$$

$$\frac{\partial}{\partial \alpha_j}\{E\} = 2\sum_{k=k_1}^{k_2} e(k) \cdot \frac{\partial e(k)}{\partial \alpha_j}, \quad j = 0, 1, \ldots\ldots, n$$

where the values $\frac{\partial e(k)}{\partial \beta_i}$ and $\frac{\partial e(k)}{\partial \alpha_j}$ need to be calculated

$$\frac{\partial}{\partial \beta_i}\{e(z)\} = \frac{\partial}{\partial \beta_i}\{y(z) - w(z)\} = -\frac{\partial}{\partial \beta_i}\{w(z)\}$$

$$= \frac{-z^{-i}\; u(z)}{\alpha_0 + \alpha_1 z^{-1} + \cdots\cdots + \alpha_n z^{-n}}$$

If we define

$$em(z) = \frac{u(z)}{\alpha_0 + \alpha_1 z^{-1} + \cdots\cdots\cdots + \alpha_n z^{-n}}, \quad \text{then}$$

$$\frac{\partial}{\partial \beta_i}\{e(z)\} = -z^{-i}\, em(z), \quad i = 0, 1, \ldots\ldots, m$$

$$\frac{\partial}{\partial \alpha_j}\{e(z)\} = \frac{\partial}{\partial \alpha_j}\{y(z)-w(z)\} = -\frac{\partial}{\partial \alpha_j}\{w(z)\}$$

$$= \frac{-u(z)}{(\alpha_0+\alpha_1 z^{-1}+\cdots\cdots+\alpha_n z^{-n})^2}\left(-z^{-j}(\beta_0+\beta_1 z^{-1}+\cdots\cdots+\beta_m z^{-m})\right)$$

$$= \frac{z^{-j}\ u(z)\ .\ H_M(z)}{\alpha_0+\alpha_1 z^{-1}+\cdots\cdots+\alpha_n z^{-n}} = \frac{z^{-j}\ w(z)}{\alpha_0+\alpha_1 z^{-1}+\cdots\cdots+\alpha_n z^{-n}}$$

If we define $esm(z) = \dfrac{w(z)}{\alpha_0+\alpha_1 z^{-1}+\cdots\cdots+\alpha_n z^{-n}}$, then

$$\frac{\partial}{\partial \alpha_j}\{e(z)\} = z^{-j}\ esm(z) \qquad , \ j = 0, 1, \ldots\ldots, n$$

we, also, can write $w(z)$ in terms of em(z) as follows

$$w(z) = \frac{\beta_0+\beta_1 z^{-1}+\cdots\cdots+\beta_m z^{-m}}{\alpha_0+\alpha_1 z^{-1}+\cdots\cdots+\alpha_n z^{-n}}\ .\ u(z)$$

$$= (\beta_0+\beta_1 z^{-1}+\cdots\cdots+\beta_m z^{-m})\ .\ em(z)$$

The sequence em(k) can be generated from the model using the relation

$$em(k) = \frac{1}{\alpha_0}\left(u(k) - \alpha_1\ em(k-1) - \cdots\cdots - \alpha_n\ em(k-n)\right)$$

consequently we are able to calculate the following quantities

$$w(k) = \beta_0\ em(k)+\beta_1\ em(k-1)+\cdots\cdots+\beta_m\ em(k-m)$$

$$e(k) = y(k) - w(k)$$

$$\frac{\partial}{\partial \beta_i}\{e(k)\} = -\ em(k-i) \qquad , \ i = 0, 1, \ldots\ldots, m$$

The sequence esm(k) can be generated from the sensitivity model whose discrete-time transfer function is given by

$$H_{SM}(z) = \frac{1}{\alpha_0+\alpha_1 z^{-1}+\cdots\cdots+\alpha_n z^{-n}}$$

and whose input is the sequence $w(k)$

$$esm(k) = \frac{1}{\alpha_0}\left(w(k) - \alpha_1\ esm(k-1) - \cdots\cdots - \alpha_n\ esm(k-n)\right)$$

consequently we are able to calculate the following quantities

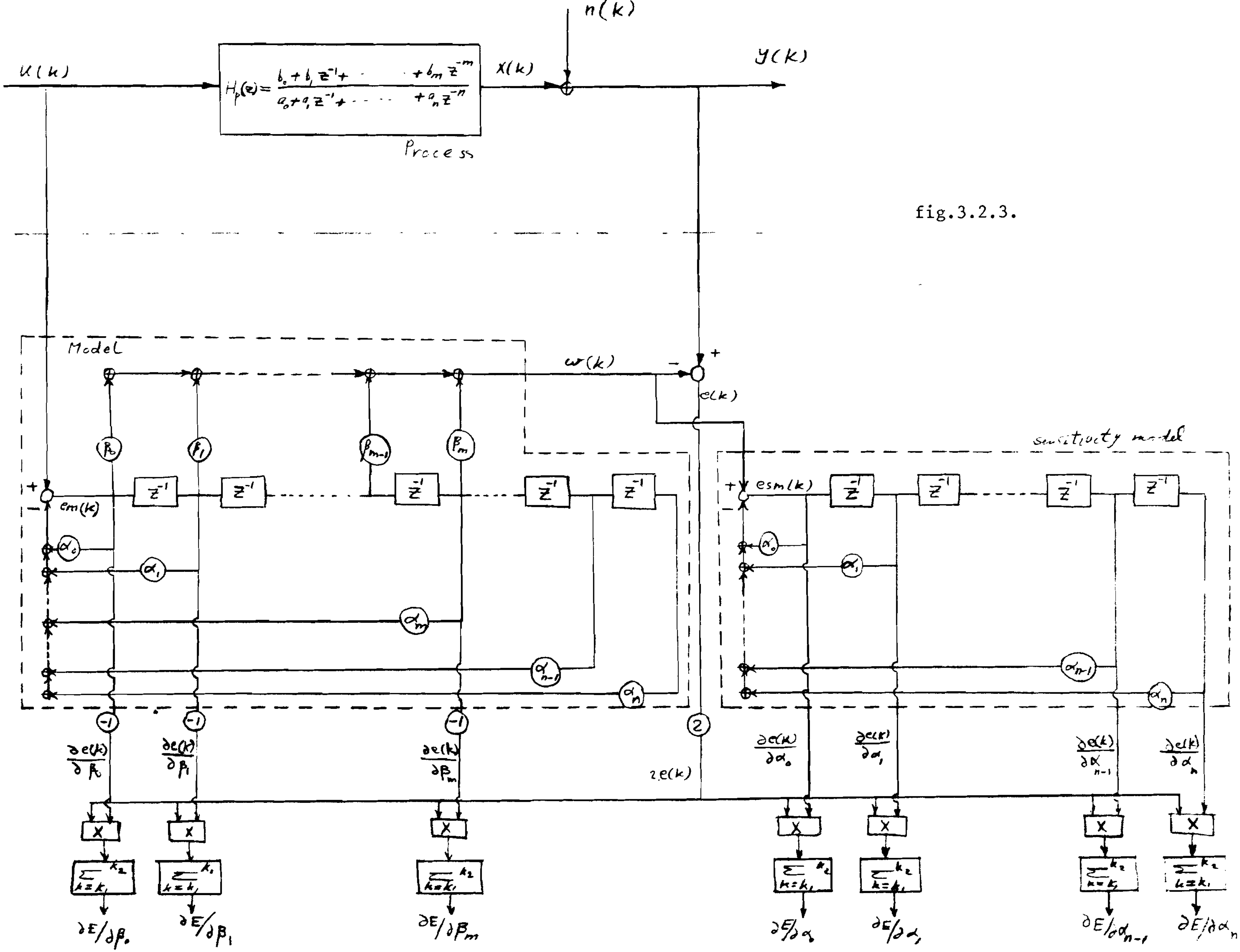

n(k)
u(k)
$H_p(z) = \frac{b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}}{a_0 + a_1 z^{-1} + \cdots + a_n z^{-n}}$
x(k)
y(k)
Process
Model
w(k)
e(k)
em(k)
esm(k)
sensitivity model
2e(k)
$\partial e(k)/\partial \beta_0$
$\partial e(k)/\partial \beta_1$
$\partial e(k)/\partial \beta_m$
$\partial e(k)/\partial \alpha_0$
$\partial e(k)/\partial \alpha_1$
$\partial e(k)/\partial \alpha_{n-1}$
$\partial e(k)/\partial \alpha_n$
$\partial E/\partial \beta_0$
$\partial E/\partial \beta_1$
$\partial E/\partial \beta_m$
$\partial E/\partial \alpha_0$
$\partial E/\partial \alpha_1$
$\partial E/\partial \alpha_{n-1}$
$\partial E/\partial \alpha_n$

fig.3.2.3.

$$\frac{\partial}{\partial \alpha_j}\{e(k)\} = esm(k-j) \qquad , \; j = 0, 1, \ldots\ldots, n$$

Now, the quantities $\frac{\partial}{\partial \beta_i}\{E\}$ and $\frac{\partial}{\partial \alpha_j}\{E\}$ can be evaluated and used to direct the adjustment of the unknown parameters.

fig. 3.2.3. A schematic for the simulation of the model and the sensitivity model for the determination of the parameters sensitivity functions: see page 25.

3.3 Model stability

3.3.1. Model stability in the z-plane

Adopting a second order process to be identified using model of the same structure,we can write the model pulse transfer function as follows

$$H(z) = \frac{\alpha_6 z^2 + \alpha_5 z + \alpha_4}{\alpha_3 z^2 + \alpha_2 z + \alpha_1}$$

The roots of the polynomial $(\alpha_3 z^2 + \alpha_2 z + \alpha_1)$ determine model stability If we write

$$H(z) = \frac{P_2(z)}{(z-r_1)(z-r_2)}$$

where r_1 and r_2 are the roots, given by

$$r_{1,2} = \frac{1}{2\alpha_3}\left(-\alpha_2 \pm j\sqrt{4\alpha_1\alpha_3 - \alpha_2^2}\right)$$

$$r_1 = |r_1|\, e^{j\phi_1} \quad , \text{and} \quad r_2 = |r_2|\, e^{j\phi_2}$$

where $|r_1| = |r_2|$ and $\phi_1 = -\phi_2$ for complex roots, and

$\phi_1 = \phi_2 = 0$ for real roots

by partial fractions $H(z) = \dfrac{C_1 z}{(z-r_1)} + \dfrac{C_2 z}{(z-r_2)}$

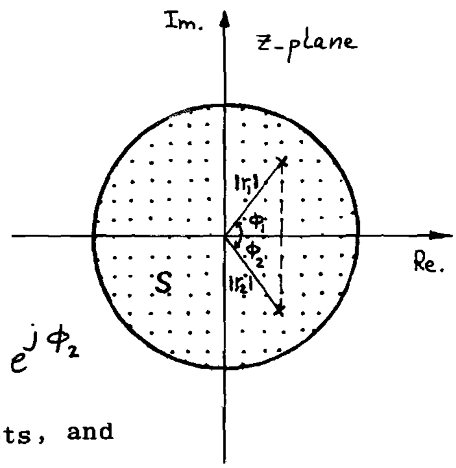


fig.3.3.1.

wich can be transformed in the time domain as follows

$$h(t) = C_1 e^{\frac{\ln r_1}{T_s}\cdot t} + C_2 e^{\frac{\ln r_2}{T_s}\cdot t} = C_1 e^{\gamma_1 t} + C_2 e^{\gamma_2 t}$$

where $\gamma = \dfrac{\ln r}{T_s}$, and T_s is the sampling period

where γ, in general,is complex.

The model is defined to be stable if the response is bounded for any bounded excitation, it can be proved that this condition is fulfilled only, if $Re(\gamma_1) < o$ and $R_1(\gamma_2) < o$

$$\gamma = \frac{1}{T_s} \ln(|r|\, e^{j\phi})$$

$$= \frac{1}{T_s} \ln|r| + \frac{j}{T_s}\phi$$

$$Re(\gamma) = \frac{1}{T_s} \ln|r|$$

which can be negative only if $|r| < 1$

hence, for the model to be stable, its poles in the complex z-plane must lie within the unit circle .

3.3.2. Model stability in the parameter space:

Determination of stability regions in the parameter space :

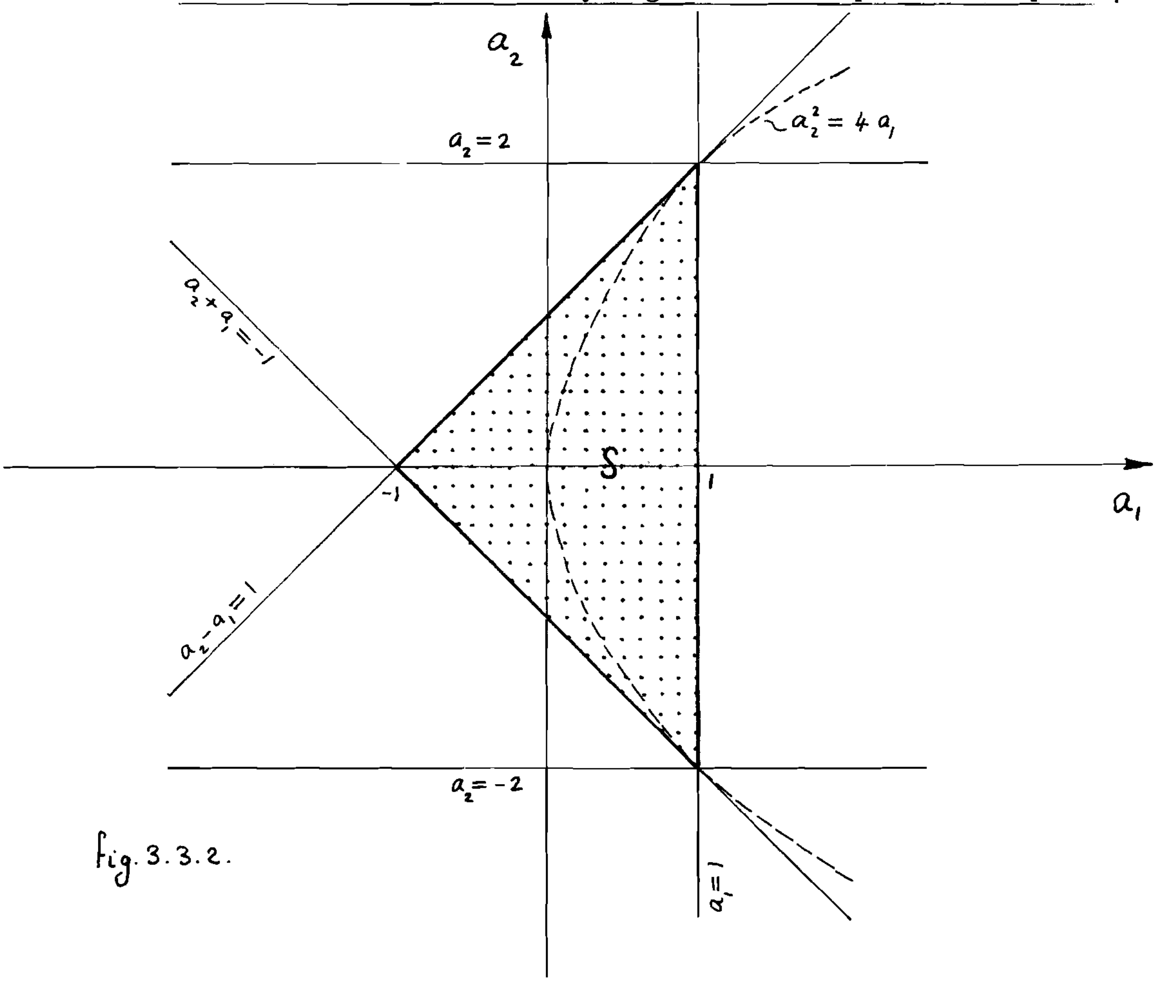


Fig. 3.3.2.

In the a_1-a_2 space: $(a_3=1)$

$$H(z) = z^2 + a_2 z + a_1 = o$$

$$z = \frac{-a_2 \pm \sqrt{a_2^2 - 4a_1}}{2}$$

for $(a_2^2 < 4a_1)$ $\longrightarrow$ Im. roots:

$$z = \frac{-a_2}{2} \pm \frac{j\sqrt{4a_1 - a_2^2}}{2}$$

$$|z|^2 = \frac{a_2^2}{4} + \frac{4a_1}{4} - \frac{a_2^2}{4} = a_1$$

for stability $(|z|^2 < 1)$ $\longrightarrow$ $\boxed{a_1 < 1}$

for $(a_2^2 > 4a_1)$ $\longrightarrow$ real roots:

$$z = \frac{-a_2}{2} \pm \frac{\sqrt{a_2^2 - 4a_1}}{2}$$

for stability $(-1 < z < 1)$, or

$$-2 < -a_2 \pm \sqrt{a_2^2 - 4a_1} < 2$$

$$(a_2 - 2) < \pm\sqrt{a_2^2 - 4a_1} < (a_2 + 2)$$

for $(a_2 - 2) < \pm\sqrt{a_2^2 - 4a_1}$ $\longrightarrow$ $(a_2 - 2) < -\sqrt{a_2^2 - 4a_1}$

$-(2 - a_2) < -\sqrt{a_2^2 - 4a_1}$ $\longrightarrow$ $(2 - a_2) > \sqrt{a_2^2 - 4a_1}$

i.e. $(2 - a_2) > 0$ $\longrightarrow$ $\boxed{a_2 < 2}$

$4 + a_2^2 - 4a_1 > a_2^2 - 4a_1$ $\longrightarrow$ $\boxed{a_2 - a_1 < 1}$

for $(a_2 + 2) > \pm\sqrt{a_2^2 - 4a_1}$ $\longrightarrow$ $(a_2 + 2) > \sqrt{a_2^2 - 4a_1}$

i.e. $(a_2 + a) > o$ $\longrightarrow$ $\boxed{a_2 > -2}$

$a_2^2 + 4 + 4a_2 > a_2^2 - 4a_1$ $\longrightarrow$ $\boxed{a_2 + a_1 > -1}$

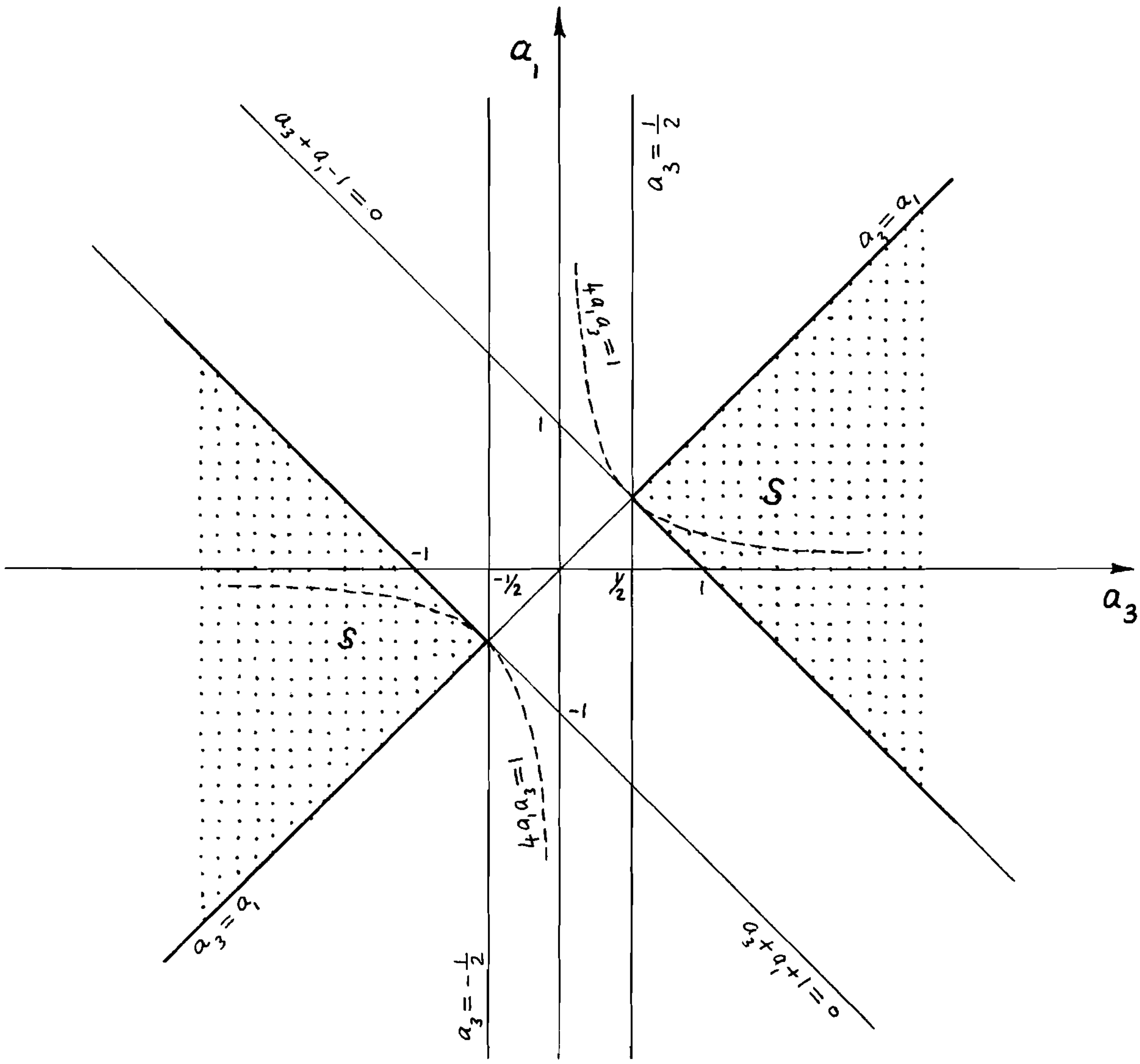


fig. 3.3.3.

<u>In the $a_1 - a_3$ space</u>: $(a_2=1)$

$$H(z) = a_3 z^2 + z + a_1 = 0$$

$$z = \frac{-1 \pm \sqrt{1 - 4a_1a_3}}{2a_3}$$

<u>for $(4a_1a_3 > 1)$ ⟶ Im. roots</u>:

$$z = \frac{(-1 \pm j\sqrt{4a_1a_3 - 1})}{2a_3}$$

$$|z|^2 = \frac{(1 + 4a_1a_3 - 1)}{4a_3^2} = \frac{a_1}{a_3}$$

for stability $(|z|^2 < 1)$ $\longrightarrow$ $\boxed{|a_1| < |a_3|}$

for $(4a_1a_3 < 1)$ $\longrightarrow$ real roots:

$$z = \frac{(-1 \pm \sqrt{1 - 4a_1a_3})}{2a_3}$$

for stability $(-1 < z < +1)$

$a_3 > 0$:

$$-2a_3 < -1 \pm \sqrt{1 - 4a_1a_3} < 2a_3$$

$$(1 - 2a_3) < \pm\sqrt{1 - 4a_1a_3} < (1 + 2a_3)$$

for

$$(1 - 2a_3) < \pm\sqrt{1 - 4a_1a_3} \longrightarrow (1 - 2a_3) < -\sqrt{1 - 4a_1a_3}$$

$$-(2a_3 - 1) < -\sqrt{1 - 4a_1a_3} \longrightarrow (2a_3 - 1) > \sqrt{1 - 4a_1a_3}$$

i.e.

$$(2a_3 - 1) > 0 \longrightarrow a_3 - \frac{1}{2} > 0$$

$$4a_3^2 + 1 - 4a_3 > 1 - 4a_1a_3 \longrightarrow a_3(a_3 + a - 1) > 0 \quad , a_3 > 0$$

$\boxed{a_3 > \frac{1}{2}}$

$\boxed{a_3 + a_1 - 1 > 0}$, $a_3 > 0$

and for $(1 + 2a_3) > \pm\sqrt{1 - 4a_1a_3} \longrightarrow (1 + 2a_3) > \sqrt{1 - 4a_1a_3}$

i.e. $(1 + 2a_3) > 0 \longrightarrow \frac{1}{2} + a_3 > 0$

$$4a_3 + 1 + 4a_3^2 > 1 - 4a_1a_3 \longrightarrow a_3(a_3 + a_1 + 1) > 0 \quad , a_3 > 0$$

$\boxed{a_3 > -\frac{1}{2}}$

$\boxed{a_3 + a_1 + 1 > 0}$ $a_3 > 0$

$a_3 < 0$:

$$-2a_3 > -1 \pm \sqrt{1 - 4a_1a_3} > 2a_3$$

$$(1 - 2a_3) > \pm\sqrt{1 - 4a_1a_3} > (1 + 2a_3)$$

Following the same analysis as above

$\boxed{a_3 < \frac{1}{2}}$ $\boxed{a_3 < -\frac{1}{2}}$

$\boxed{a_3 + a_1 - 1 < 0}$ $\boxed{a_3 + a_1 + 1 < 0}$, $a_3 < 0$

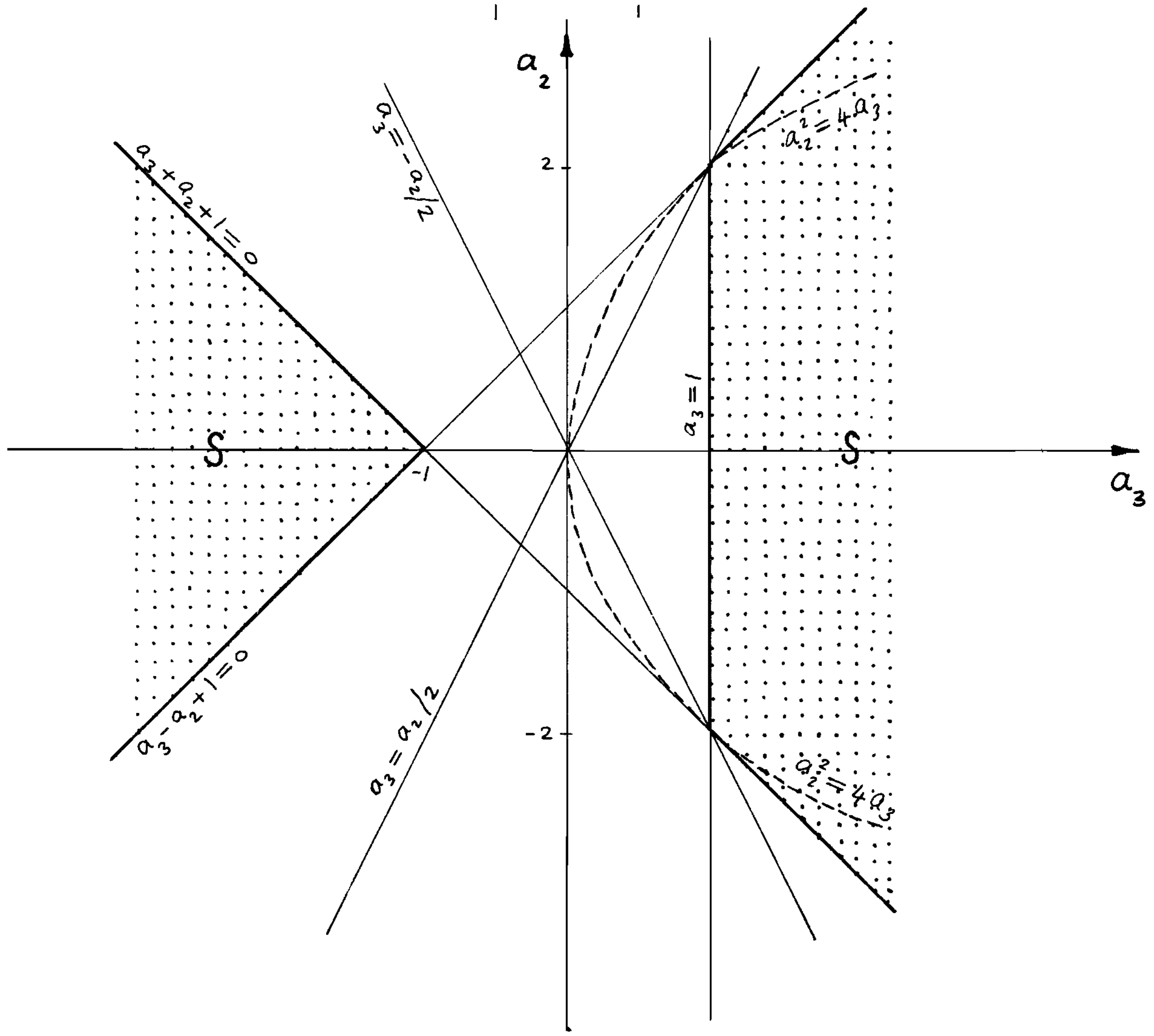


fig. 3.3.4.

<u>In the $a_2 - a_3$ space:</u> ($a_1=1$)

$$H(z) = a_3 z^2 + a_2 z + 1 = 0$$

$$z = \frac{-a_2 \pm \sqrt{a_2^2 - 4a_3}}{2a_3}$$

<u>for ($a_2^2 < 4a_3$) $\longrightarrow$ Im. roots:</u>

$$z = \frac{(-a_2 \pm j\sqrt{4a_3 - a_2^2})}{2a_3}$$

$$|z|^2 = \frac{(a_2^2 + 4a_3 - a_2^2)}{4a_3^2} = \frac{1}{a_3}$$

for stability ($|z|^2 < 1$) $\longrightarrow$ $\boxed{a_3 > 1}$

for ($a_2^2 > 4a_3$) $\longrightarrow$ real roots:

$$z = \frac{(-a_2 \pm \sqrt{a_2^2 - 4a_3})}{2a_3}$$

for stability ($-1 < z < 1$)

$a_3 > 0$: $-2a_3 < -a_2 \pm \sqrt{a_2^2 - 4a_3} < 2a_3$

$$(a_2 - 2a_3) < \pm\sqrt{a_2^2 - 4a_3} < (a_2 + 2a_3)$$

for $(a_2 - 2a_3) < \pm\sqrt{a_2^2 - 4a_3} \longrightarrow (a_2 - 2a_3) < -\sqrt{a_2^2 - 4a_3}$

$-(ea_3 - a_2) < -\sqrt{a_2^2 - 4a_3} \longrightarrow (2a_3 - a_2) > \sqrt{a_2^2 - 4a_3}$

i.e. $(2a_3 - a_2) > 0 \longrightarrow \boxed{a_3 > \frac{a_2}{2}}$

$4a_3^2 + a_2^2 - 4a_3a_2 > a_2^2 - 4a_3 \longrightarrow \boxed{a_3\,(a_3 - a_2 + 1) > 0}$, $a_3 > 0$

and, for $(a_2 + 2a_3) > \pm\sqrt{a_2^2 - 4a_3} \longrightarrow (a_2 + 2a_3) > \sqrt{a_2^2 - 4a_3}$

i.e. $(a_2 + 2a_3) > 0 \longrightarrow \boxed{a_3 > \frac{-a_2}{2}}$

$a_2^2 + 4a_3^2 + 4a_3a_2 > a_2^2 - 4a_3 \longrightarrow \boxed{a_3(a_3 + a_2 + 1) > 0}$, $a_3 > 0$

$a_3 < 0$: $-2a_3 > -a_2 \pm \sqrt{a_2^2 - 4a_3} > 2a_3$

$$(a_2 - 2a_3) > \pm\sqrt{a_2^2 - 4a_3} > (a_2 + 2a_3)$$

following the same analysis as above we get the following inequalities:

$\boxed{a_3 < \frac{a_2}{2}}$, $\boxed{a_3(a_3 - a_2 + 1) > 0}$, $a_3 < 0$ and $\boxed{a_3 < \frac{-a_2}{2}}$, $\boxed{a_3(a_3 + a_2 + 1) > 0}$, $a_3 < 0$

<u>In the $a_1 - a_2 - a_3$ space:</u>

Having determined the stability regions in 3 different parameter planes we can visualise the situation in the 3-dimensional parameter space $a_1 - a_2 - a_3$.

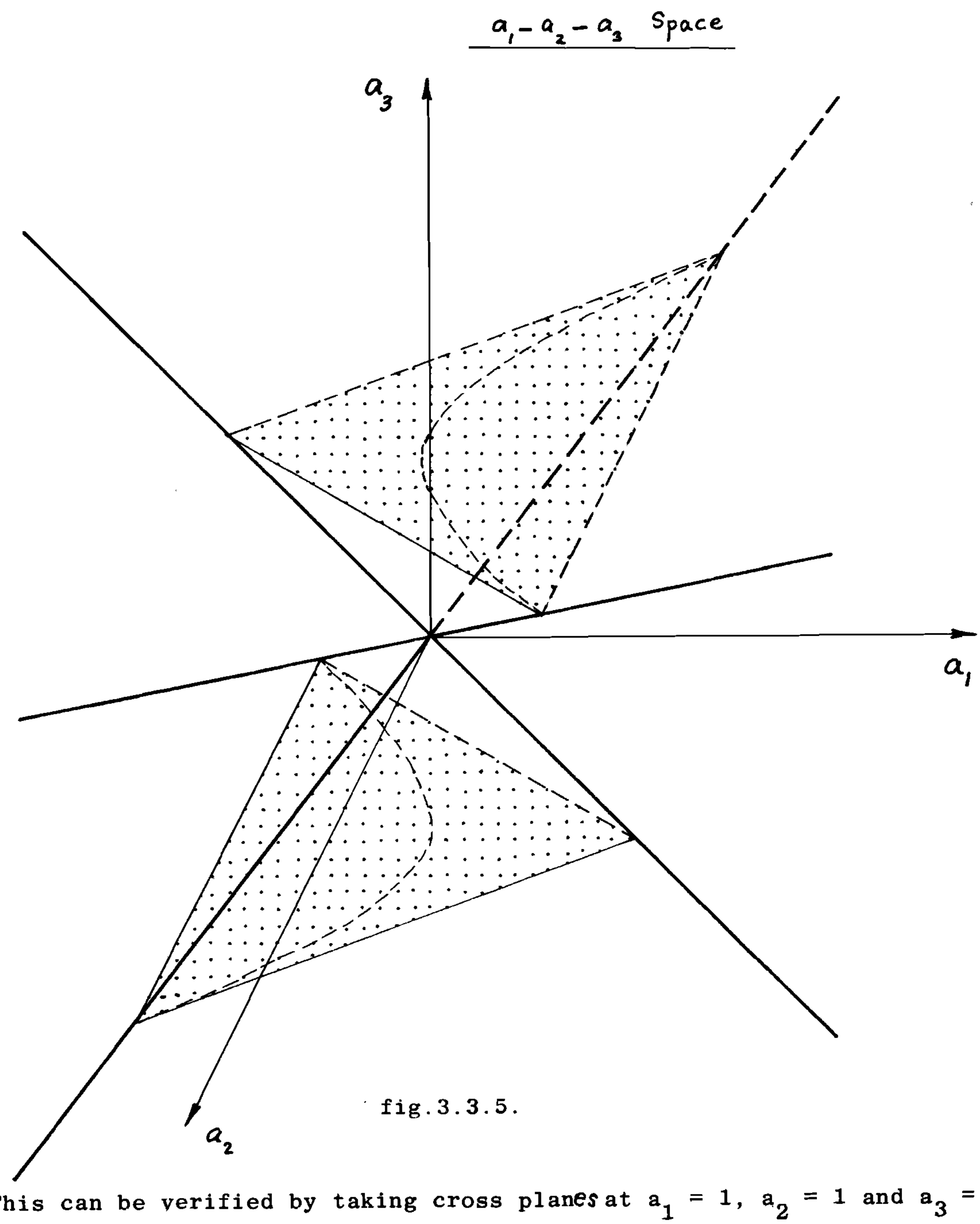


fig.3.3.5.

This can be verified by taking cross planes at $a_1 = 1$, $a_2 = 1$ and $a_3 = 1$ and comparing with planer stability regions derived previously.

3.3.3. Correspondence of stability regions in the z-plane and in the parameter space:

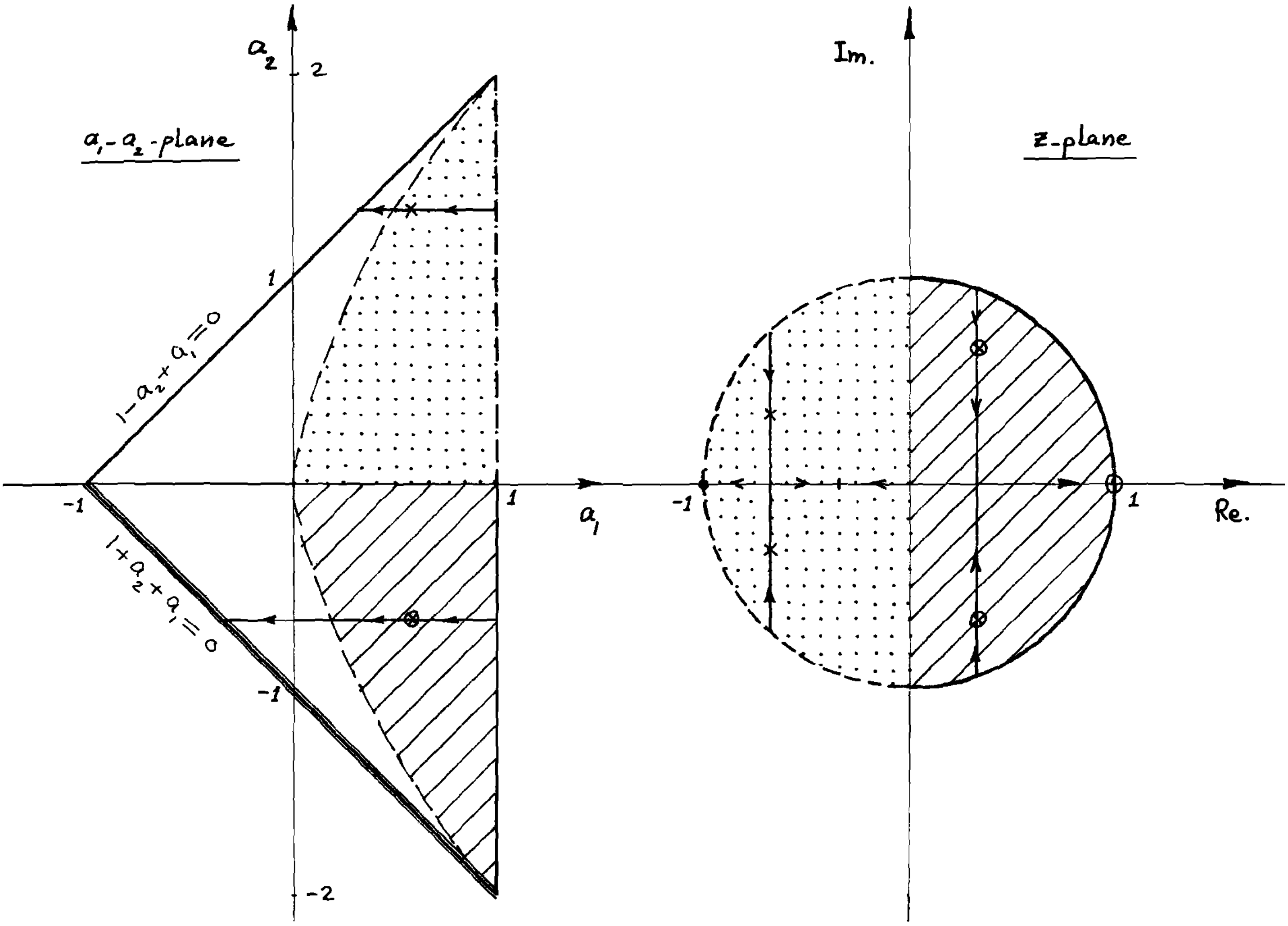


fig. 3.3.6.

It is of much importance for further discussions to study the correspondence of the stability regions in the z-plane and the parameter plane. For a second order model given by the discrete transfere function.

$$H(z) = \frac{P(z)}{z^2 + a_2 z + a_1}$$

where p(z) is a polynomial in z of order equal to or less than 2.

The equation

$$z^2 + a_2 z + a_1 = 0$$

determines the locations of the two model poles in the z-plane, they are given by

$$z_{1,2} = -\frac{a_2}{2} \pm \frac{\sqrt{a_2^2 - 4a_1}}{2}$$

Points in the stability region of the parameter plane satisfying the relation

$$a_2^2 < 4a_1$$

are transformed into complex conjugate poles in the z-plane; futher-more, for $a_2 > o$ the poles are in the left hand side of the unit circle

for $a_2 = o$ the poles are on the imagenary axis

for $a_2 < o$ the poles are in the right hand side of the unit circle

Points in the stability region of the parameter plane satisfying the relation

$$a_2^2 \geq 4a_1$$

are transformed into two real poles in z-plane; furthermore,

if the "=" sign is satisfied in the above relation the two poles coincide on the real axis of the z-plane

if the ">" sign is satisfied in the above relation the two poles depart from each other along the real axis of the z-plane.

Fig. 3.3.6. shows the roots of two points in the parameter plane departing from the vertical side $a_1 = 1$ and moving horizontally in the stability region untill the borders and their corresponding roots in the z-plane.

We note that each point on the side $a_1 = 1$ of the stability triangle in the parameter plane is transformed into complex conjugate pair of poles on the unit circle of the z-plane.

Points on the lower triangle side

$$1 + a_2 + a_1 = 0$$

are transformed into two real poles; one of them - at least - lies on the point z = 1 in the z-plane (this can be verified if we substitute z = 1 in the equation $z^2 + a_2 z + a_1 = 0$), points on the upper triangle side

$$1 - a_2 + a_1 = 0$$

are transformed into two real poles; one of them - at least - lies on the point z = -1 in the z-plane (this can be verified if we substitute z = -1 in the equation $z^2 + a_2 z + a_1 = 0$).

4. Input Signal

4.1. Binary Noise signal (maximum length sequences):

As white noise has well known desirable properties as an excitation for the process to be identified, we use the so called maximum length sequences as an input signal. This signal can be easily generated in continuous form using feedback shift registers as in fig. 4.1.1. or in sampled form by a simple simulation program on a computer. It has a power spectrum which approximates a white spectrum as good as desired, by increasing the number of shift registers n and/or decreasing the duration θ of each of the output states (i.e. increasing the clock frequency).
It is of help for further discussions to study in some detail the autocorrelation function $\Psi_{uu}(\tau)$ of the generated signal and its frequency spectrum. For this purpose we adopt the case where n = 3, the generated signal is periodic with a period $T = (2^n - 1)\,\theta$, ($T = 7\theta$ in our case). $N = (2^n - 1)$ is the maximum number of states contained in one period of the output signal using n shift registers, this maximum can be attained using suitable feedback connections of the shift registers. The duration of each of the output states is the same and equal to the clock-pulse period θ. A delayed form of the same periodic output signal can be obtained from any of the n-shift registers.
Shown in fig. (4.1.1.) is the signal generated through the transformation of the register states $1 \rightarrow +a$ and $o \rightarrow -b$. The values of a and b can be chosen to adjust the d.c. average of the test signal and its power content.

The autocorrelation function of the generated signal can be easily computed by multiplying the original signal with a shifted signal, for different shift values. Since the original signal is periodic, its autocorrelation function is also periodic with the same period as that of the signal. It has a triangular shape which approximates a dirac function for a large number of register elements (n) and/or high clock frequency (i.e. small θ); this is shown in fig. (4.2.3.).

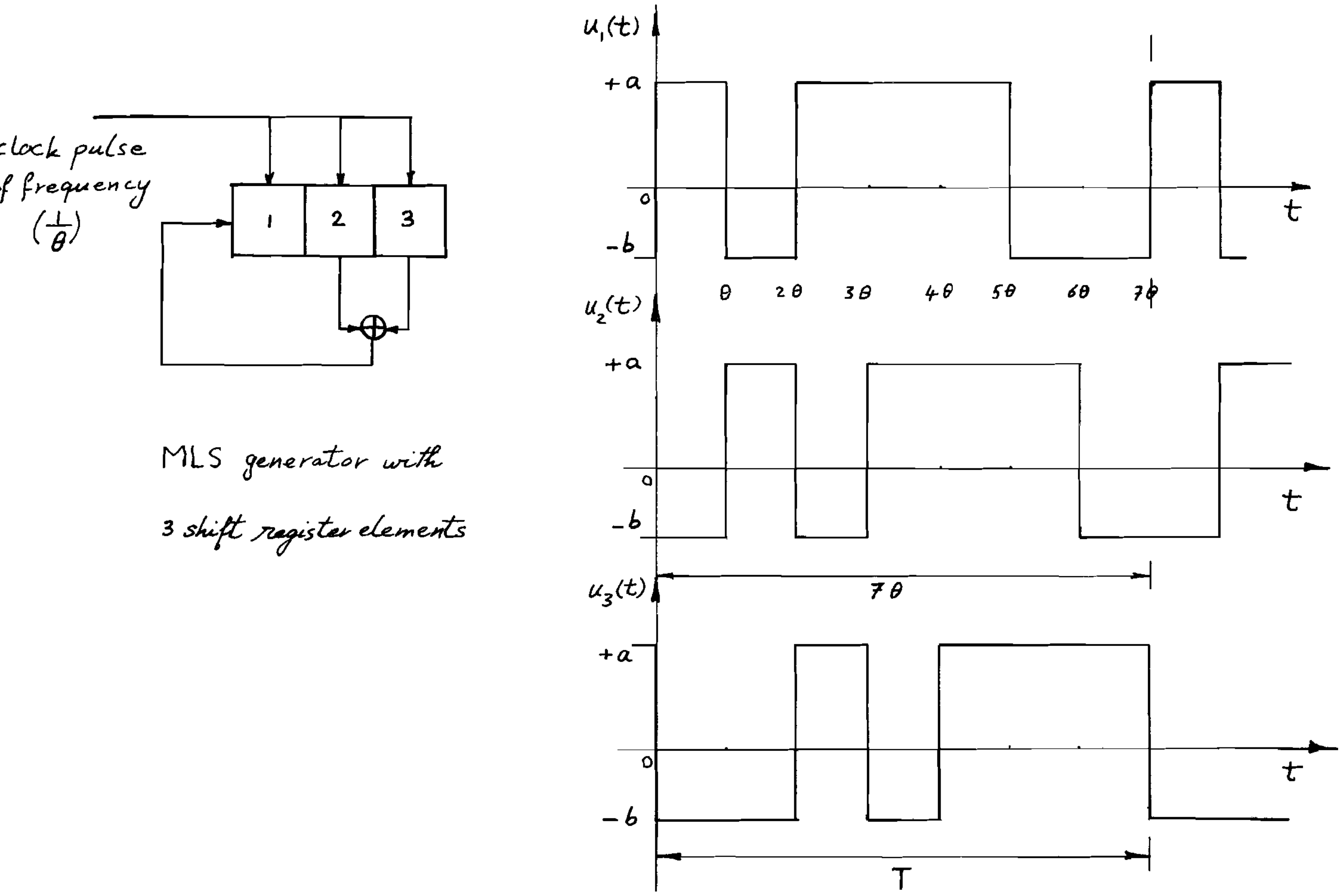


fig. (4.1.1.)

4.2. Effect of periodicity on the signal spectrum:

In the following theoretical analysis, it will be assumed that the input signal u(t) is periodic noncausal function of time. (A causal function f(t) must satisfy the condition , $f(t) = 0$ for $t < 0$)

i.e. $u(t) = u(t + kT)$, for $k = -\infty \rightarrow \infty$

where T is the signal period. This assumption is correct if the input signal is applied to the system under consideration until stationarity is achieved.

As the power spectral density and the correlation function of a periodic waveform are Fourier transform pair

$$\Phi_{uu}(\omega) = \mathcal{F}\,[\,\psi_{uu}(t)\,]$$

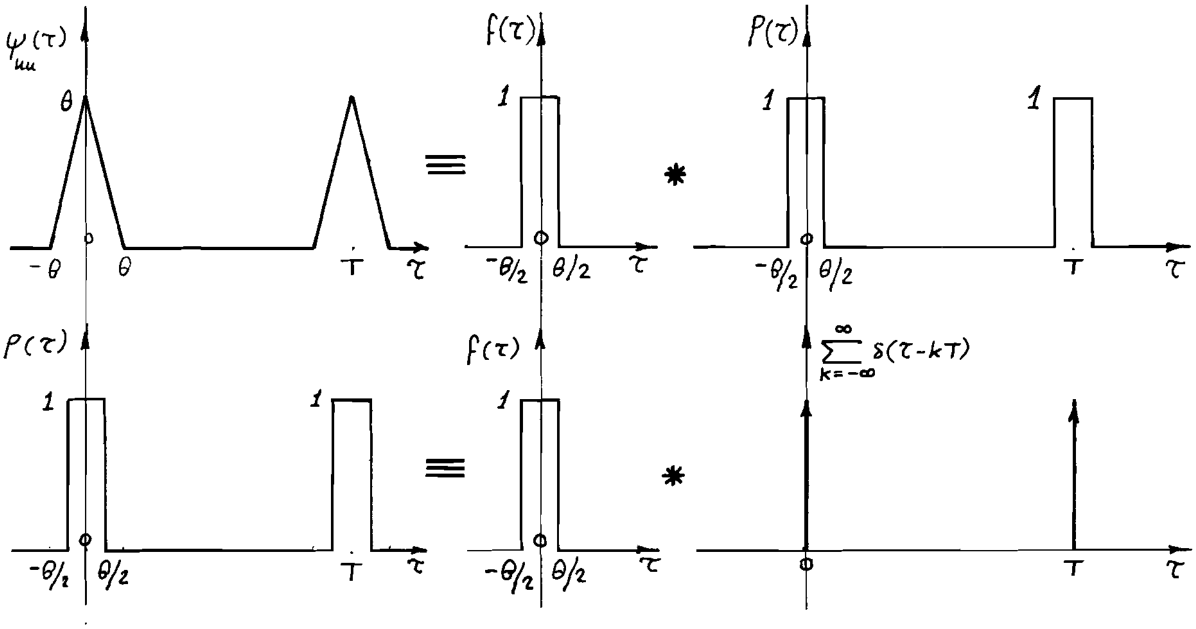


fig.(4.2.1.)

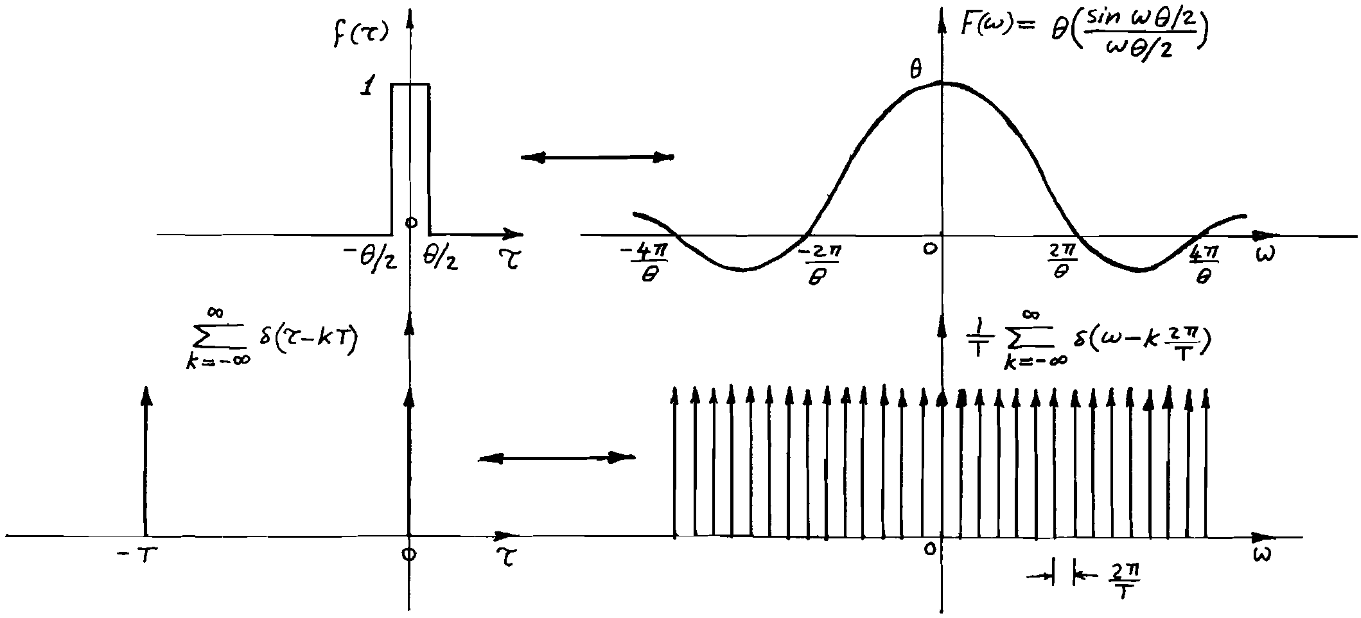


fig.(4.2.2.)

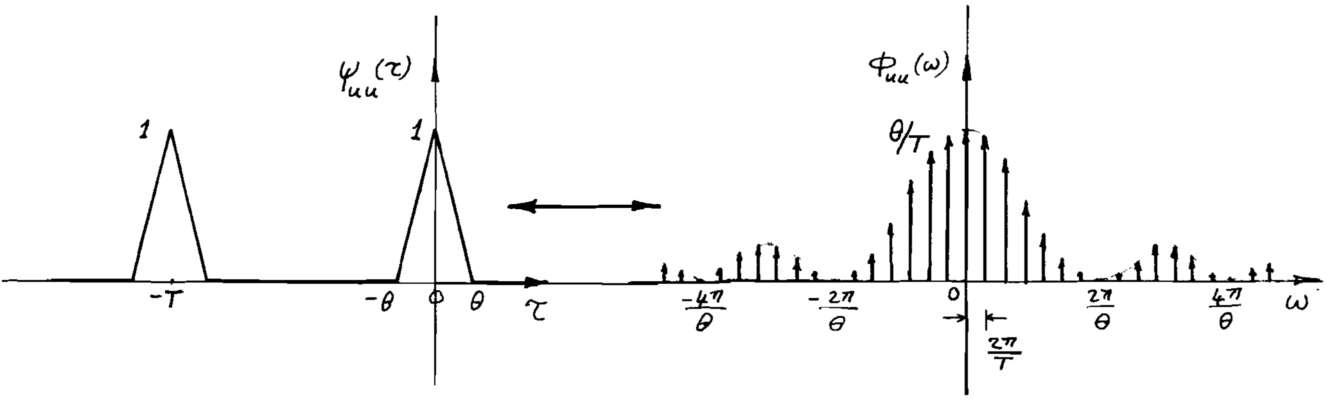


fig.(4.2.3.)

where $\phi_{uu}(\omega)$ is the power spectral density function of the signal u(t); it provides us with information about the signal power associated with a certain frequency ω.
One way to derive the power spectral density function $\phi_{uu}(\omega)$ is by recognizing that

$$\psi_{uu}(\tau) = \frac{1}{\theta} f(\tau) * f(\tau) * \sum_{k=-\infty}^{\infty} \delta(\tau - kT)$$

where (*) means convolution; this becomes multiplication in the frequency domain. See fig.(4.2.1.).

For the definition of f(τ) and its Fourier transform, see fig. (4.2.2.).
In the frequency domain we can write

$$\phi_{uu}(\omega) = \frac{1}{\theta} \times F(\omega) \times F(\omega) \times \frac{1}{T} \sum_{k=-\infty}^{\infty} \delta\left(\omega - \frac{2\pi}{T} k\right)$$

where

$$F(\omega) = \mathcal{F}\left[f(\tau)\right] = \theta \cdot \left(\frac{\sin\frac{\omega\theta}{2}}{\frac{\omega\theta}{2}}\right)$$

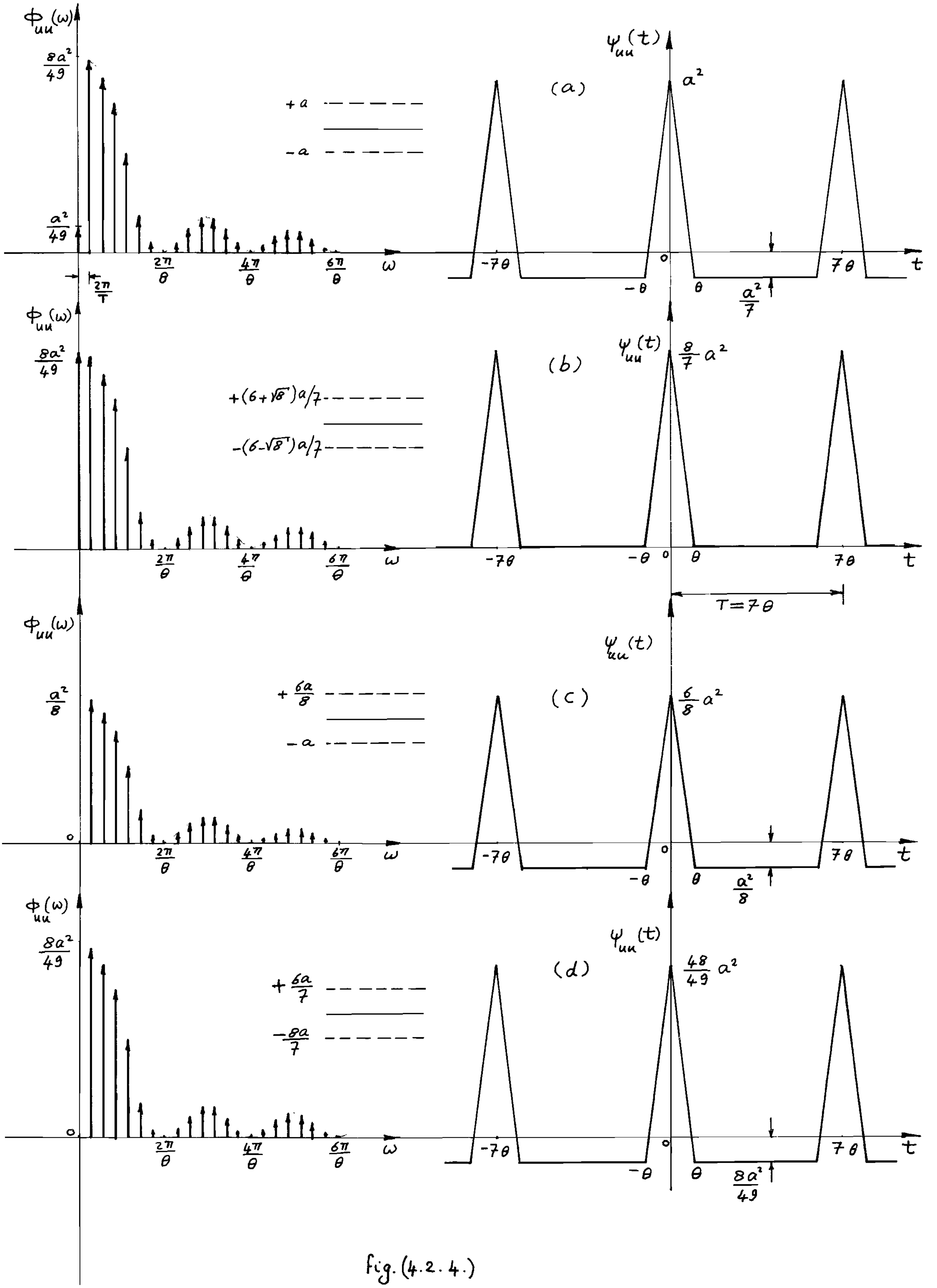
hence

$$\phi_{uu}(\omega) = \frac{\theta}{T} \cdot \left(\frac{\sin\frac{\omega\theta}{2}}{\frac{\omega\theta}{2}}\right)^2 \cdot \sum_{k=-\infty}^{\infty} \delta\left(\omega - k\frac{2\pi}{T}\right)$$

Clearly, since $\phi_{uu}(\tau)$ is periodic with period T, its power spectral density $\phi_{uu}(\omega)$ is a series of impulses whose envelope is the Fourier transform of one period of $\phi_{uu}(\tau)$. The impulses are equidistant and are separated with a frequency period equals to $\frac{2\pi}{T}$. The autocorrelation function and its power spectral density are shown in fig. (4.2.3.).
In fig. (4.2.4.) we can see the effect of selecting different combinations of signal amplitude limits on its power spectrum.
Case (a) is a representation of equal amplitude signal limits (+a and -a)
Case (b) is of non-negative autocorrelation function
Case (c) and (d) are of zero d.c. signal component.

4.3. Effect of sampling on the signal spectrum:

As we simulate our model on digital computer we, necessarily, need to use as an input to our discrete model a sequence of samples taken from the continuous signal. It is of importance to study the power spectrum of the sampled sequence and the effect of the sampling rate on this spectrum.



(a)
(b)
(c)
(d)
+a
-a
+(6+√8)a/7
-(6-√8)a/7
+6a/8
-a
+6a/7
-8a/7
8a²/49
a²/49
a²/8
2π/T
2π/θ
4π/θ
6π/θ
ω
a²
8/7 a²
6/8 a²
48/49 a²
a²/7
a²/8
8a²/49
-7θ
-θ
θ
7θ
t
T=7θ

fig. (4.2.4.)

Since the input sequence is a numerical representation of the continuous signal at the sampling instants we, effectively, have impulse sampled signal. This will result in mathematical simplifications.
The spectrum of the sampled signal g (t) is related to the spectrum of the original signal g(t) by the well known relation

$$G^*(\omega) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} G(\omega - k\omega_s)$$

where $T_s = \frac{2\pi}{\omega_s}$ is the sampling period, $G^*(\omega)$ and $G(\omega)$ are the spectrum of the sampled and the original siganls respectively. The above relation states that the sampled signal spectrum is the periodic repetition of the original signal spectrum at every multiple of the sampling frequency ω_s (including zero) with a multiplication factor $\frac{1}{T_s}$, as shown in fig. (4.3.1.).

We consider the periodic binary noise signal discussed before, with n=3, when sampling with period $T_s = \theta/NSS$ (NSS samples for each signal state). The spectrum will be repeated with frequency $\omega_s = \frac{2\pi}{T_s} = \frac{2\pi}{\theta} \times NSS$. Shown in fig. (4.3.2.) is the sampled signal spectrum for different sampling rates (i.e. different values of NSS).
It is obvious that the spectrum is becoming more uniform for a lower number of values of NSS (i.e. for lower sampling frequency); this result extends for the case where NSS is a fraction (i.e. the sampling period $T_s > \theta$). This is an important conclusion, which implies that the signal spectrum envelope approaches a white spectrum for lower sampling rates.
In all cases the spectrum consists of a series of impulses of different amplitudes and separated by frequency $\frac{2\pi}{T}$, where T is the period of the binary noise signal. The area of each impulse represents the strength of the signal component at the corresponding frequency. Note that every impulse of finite (nonzero) area indicates the existence of a pole for the signal at the corresponding frequency.

4.4. The location of the input signal poles in the parameter space:

It is of interest for further discussions to locate these poles in the complex s-plane and in the complex z-plane as well as considering their corresponding locations in the parameter space.

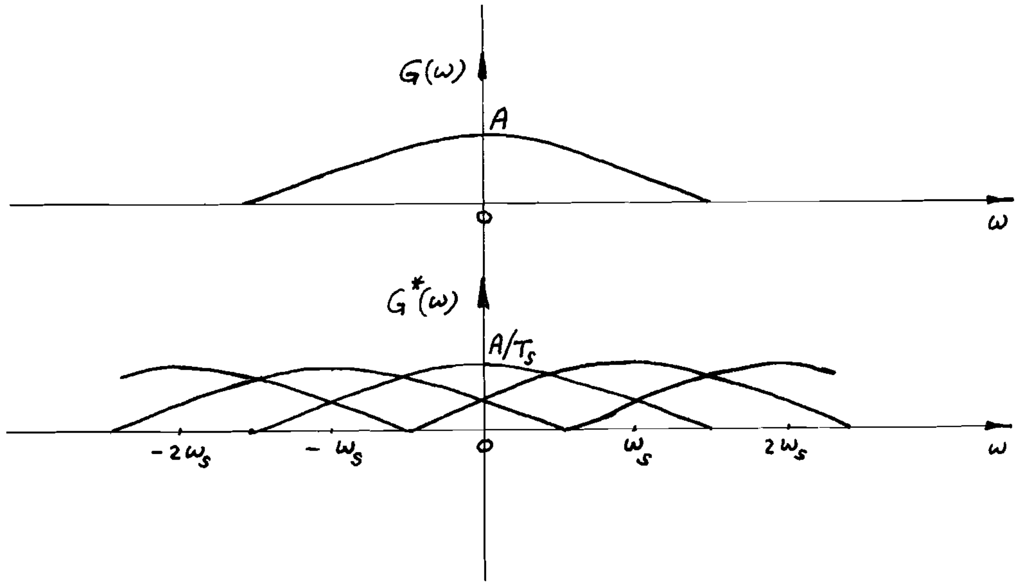


fig.(4.3.1.)

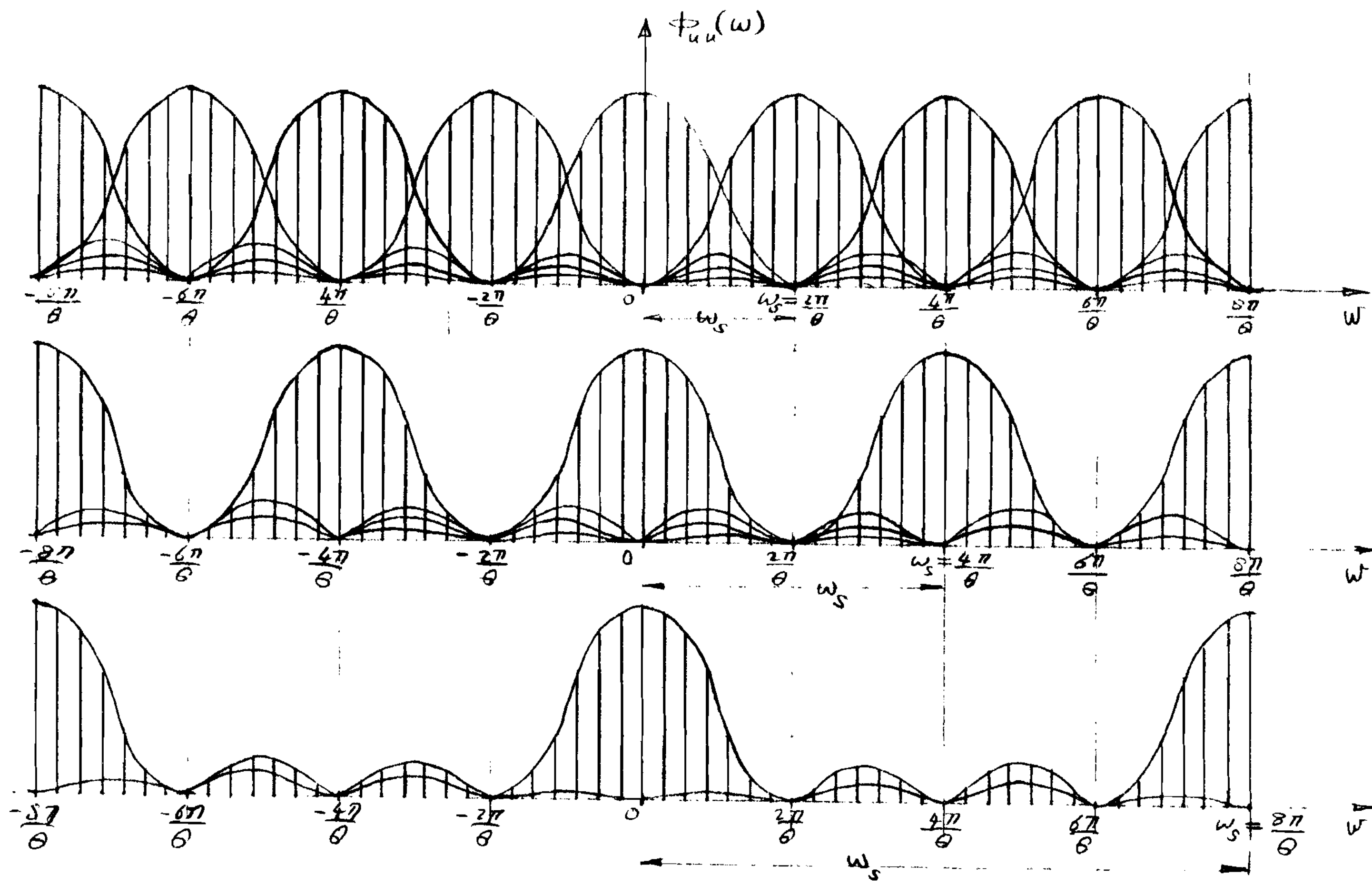


fig.(4.3.2.)

As the poles of the signal ly at equidistant points on the imaginary axis in the s-plane, we expect that they will be transformed equidistantly on the unit circle in the z-plane see fig. (4.4.1.). The relation $z = e^{ST_s} = e^{j\omega T_s}$ is used to transform the poles to the unit circle in the z-domain. It is clear that as we scan the primary strip on the imaginary axis of the s-plane $(\omega = \frac{-\pi}{T_s} \rightarrow \frac{\pi}{T_s})$ we move along the unit circle of the z-plane in anticlock wize direction $(\phi = \omega T_s = -\pi \rightarrow \pi)$. Similarly as we scan the secondary strip on the imagenary axis of the s-plane $(\omega = \frac{\pi}{T_s} \rightarrow \frac{3\pi}{T_s})$ we move along the unit circle of the z-plane $(\phi = \omega T_s = \pi \rightarrow 3\pi)$. The same applies for all other strips. As the pole distribution on every strip is the same it is enough to locate the poles of the primary strip on the unit circle. Poles of other strips will coincide with those of the primary strip.

As we have studied earlier the correspondence between the stability regions in the z-plane and the parameter space for the second order model, we can easily pose the poles location on the sides of the stability triangle in the $a_1 - a_2$ plane (corresponding to the unit circle in the z-plane), fig. (4.4.2.). The pole at z = 1 is projected on the lower side of the triangle; poles on the right hand side of the unit circle are projected on the lower part of the vertical side $(a_2 < o)$; poles on the left hand side of the unit circle are projected on the upper part of the vertical side $(a_2 > o)$. The relative spacing of the poles on the vertical side of the triangle is the same as the relative spacing of their projections on the real axis of the z-plane.

Fig. (4.4.1.) and (4.4.2.) show the location of the poles in the s-plane, z-plane and the $a_1 - a_2$ parameter plane respectively, for the case of sampled periodic binary noise signal generated using 3 register elements with sampling rate $T_s = \theta$. It is interesting to consider the case of higher sampling rate say $T_s = \frac{\theta}{4}$. The power spectrum in the primary strip in the frequency domain and the corresponding pole locations on the unit circle of the z-plane and on the triangle of stability in the parameter plane are shown in fig. (4.4.3.).

It is obvious that the concentration of poles corresponding to strong spectral components on the right hand side of the unit circle will lead to the concentration of singularities corresponding to strong spectral components on the lower part of the vertical side of the triangle.

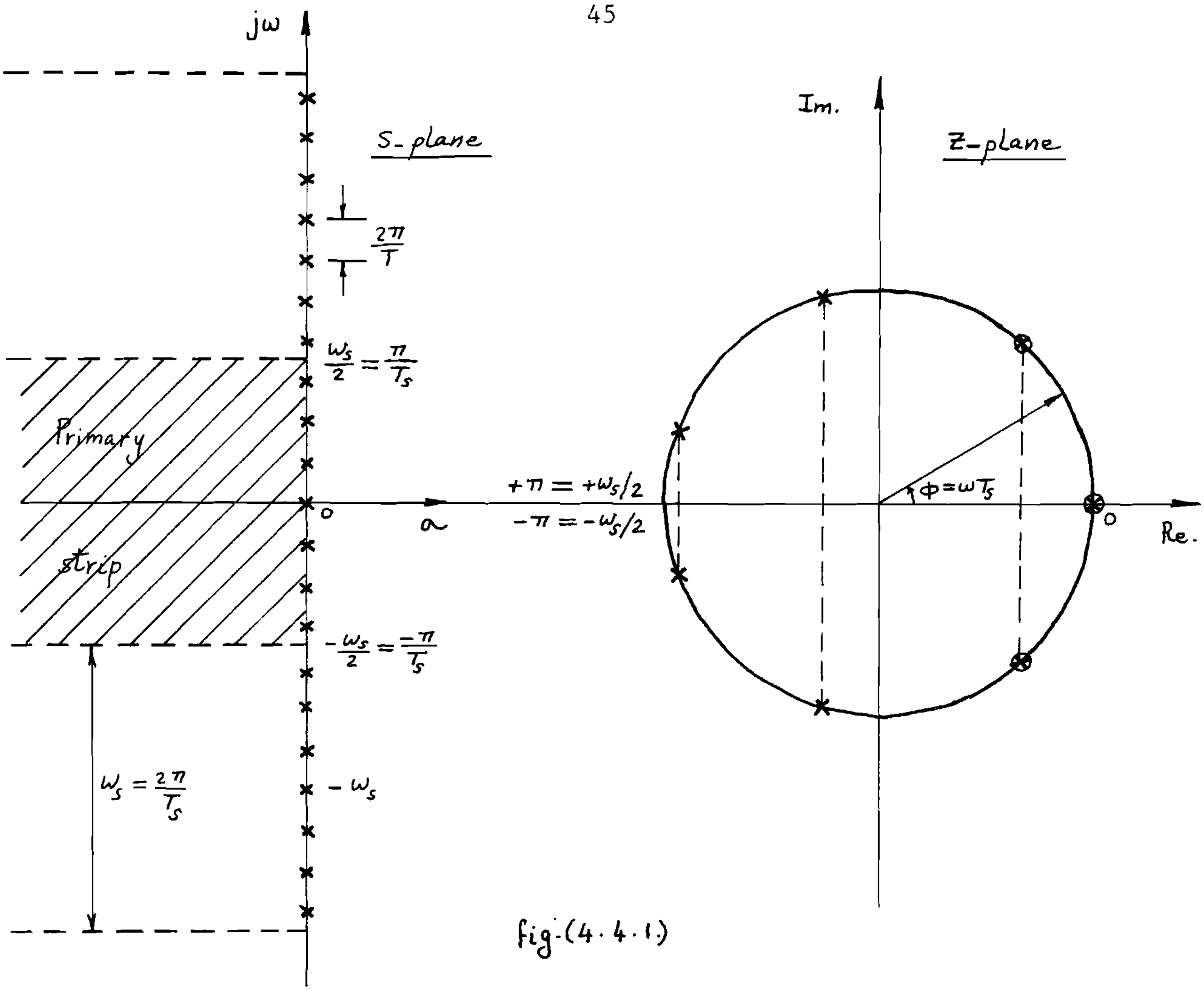


fig.(4.4.1.)

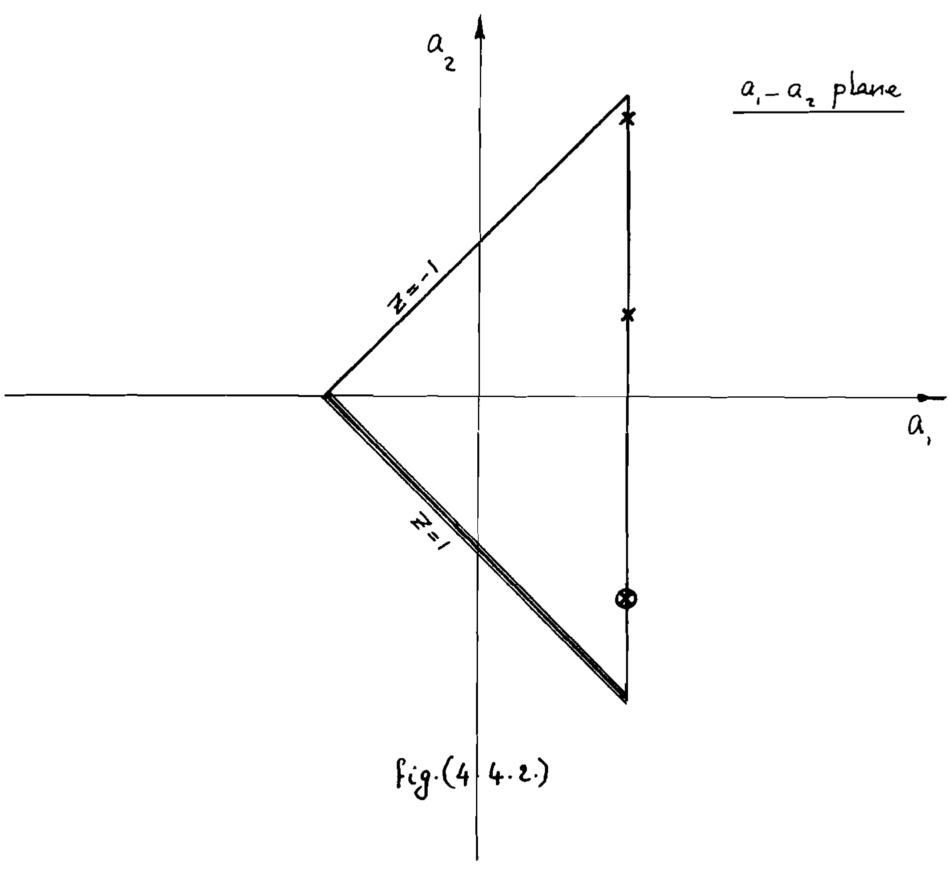


fig.(4.4.2.)

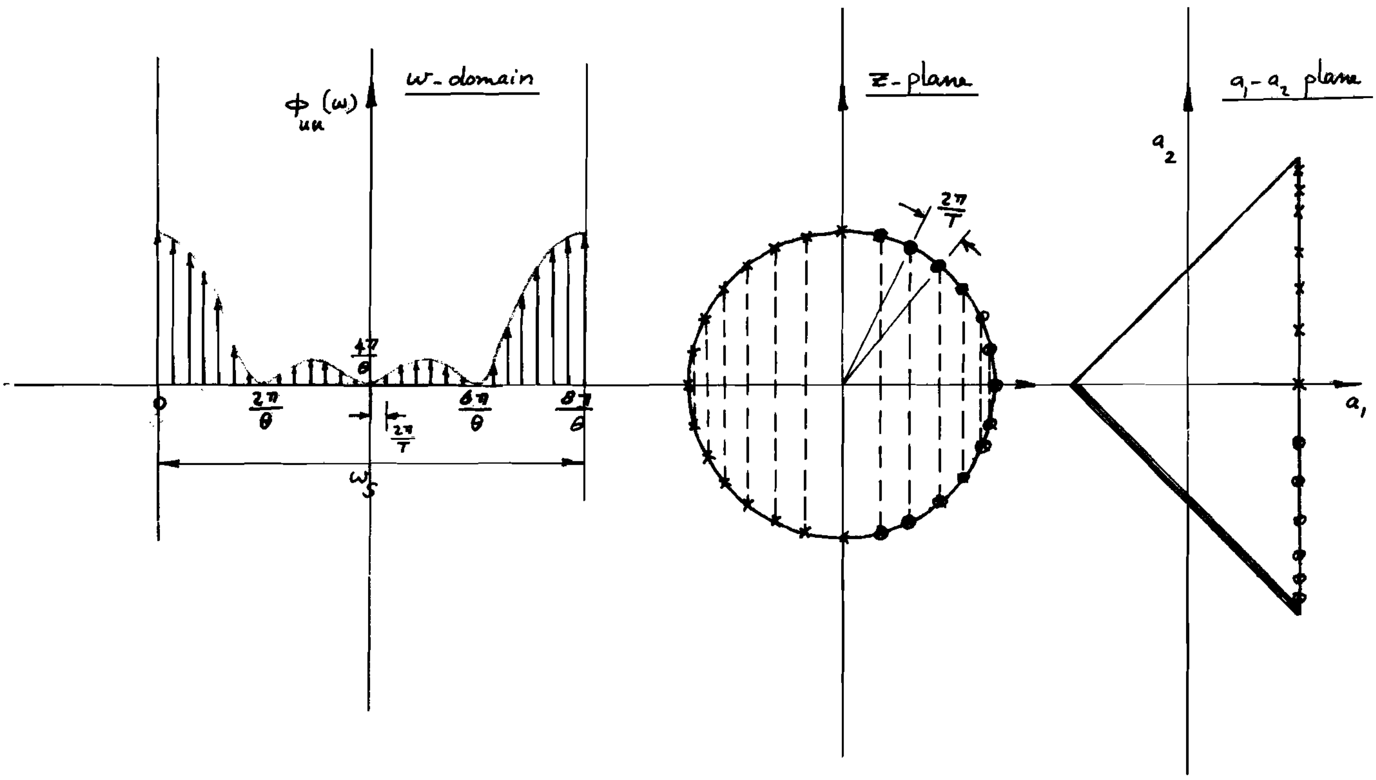


fig. (4.4.3.)

We conclude that for higher sampling rates the signal power is not uniformly distributed over the poles on the unit circle of the z-plane, this applies to the singularities on the vertical side of the triangle in the parameter plane as well. It is of much interest to note that we approach uniformity for lower sampling rates. This can be clearly seen from fig. (4.3.2.) where the signal spectrum is drawn from different sampling rates.

4.5. Effect of the input signal spectrum on the error criterion function:

We use "hill climbing" techniques as a tool for minimization;this leads to an estimate for the unknown model parameters. We expect that the behaviour of our estimation algorithm as well as the properties of the resultant estimate are highly dependent on the type and shape of the error criterion function to be minimized. We have readily discussed in chapter 3 the effect of the type of nonlinearity-in-the-parameters on the properties of the resultant estimate.

For the same type of error function, its shape in the parameter space has much effect on the behaviour and the convergence properties of the implemented minimization technique. Thus it seems very important to study whether the input signal has some effect on the shape of the error function and, if so, whether it is possible to find the best practical input signal which results in a desirable shape for the error function. We devote the remaining part of this chapter to the answers of these questions.

It is quite reasonable to expect that the effect, if it exists, is directly related to the spectrum of the input signal. In other words, this effect is due to the corresponding locations of the poles of the input signal on the boundary of the stability region in the parameter space. It is helpful to illustrate this by considering an example of a simple first order model given by

$$H_M(z) = \frac{w(z)}{u(z)} = \frac{1}{(z-\alpha)}$$

where a single parameter α is to be estimated in a least squares sense i.e. through the minimization of $\sum_{k=k_1}^{k_2} e^2(k)$

The process is given by

$$H_p(z) = \frac{y(z)}{u(z)} = \frac{1}{(z-a)}$$

where a is the unknown process parameter. The error e(z) is given by

$$e(z) = y(z) - w(z) = \left(\frac{1}{(z-a)} - \frac{1}{(z-\alpha)}\right) u(z)$$

$$= \frac{a-\alpha}{(z-a)(z-\alpha)} \quad u(z)$$

We will consider the cases of step input sequence, alternating input sequence and a combination of them.

For a step input sequence (pole at z= 1), u(z) is given by $u(z) = \frac{z}{(z-1)}$

$$e_s(z) = \frac{(a-\alpha)}{(z-a)(z-\alpha)} \cdot \frac{z}{(z-1)}$$

Making use of the final value theorem, an expression for $e_s(k)$ as $k \to \infty$ folows

$$\lim_{k\to\infty} e_s(k) = \lim_{z\to 1} \frac{(z-1)}{z} e_s(z)$$

$$= \lim_{z\to 1} \frac{(a-\alpha)}{(z-a)(z-\alpha)}$$

$$= \frac{(a-\alpha)}{(1-a)(1-\alpha)} = e_{s\infty}$$

Consequently we can write for the final squared error ($e^2_{s\infty}$)

$$e^2_{s\infty} = \frac{(a-\alpha)^2}{(1-a)^2(1-\alpha)^2}$$

Let the process parameter a = 0.5, then

$$e_{s\infty} = 2 - \frac{1}{1-\alpha} = \frac{1-2\alpha}{1-\alpha}$$

$$e^2_{s\infty} = \left(\frac{1-2\alpha}{1-\alpha}\right)^2$$

The error and the final squared error can be plotted as function of the model parameter α as shown in fig. (4.5.1.). Here the parameter axis coincides with the real axis of the z-plane. The shaded area indicates the borders of stability ($|z| \geqslant 1$). The location of the input signal pole (z = 1, for step input) is surrounded by a small circle.

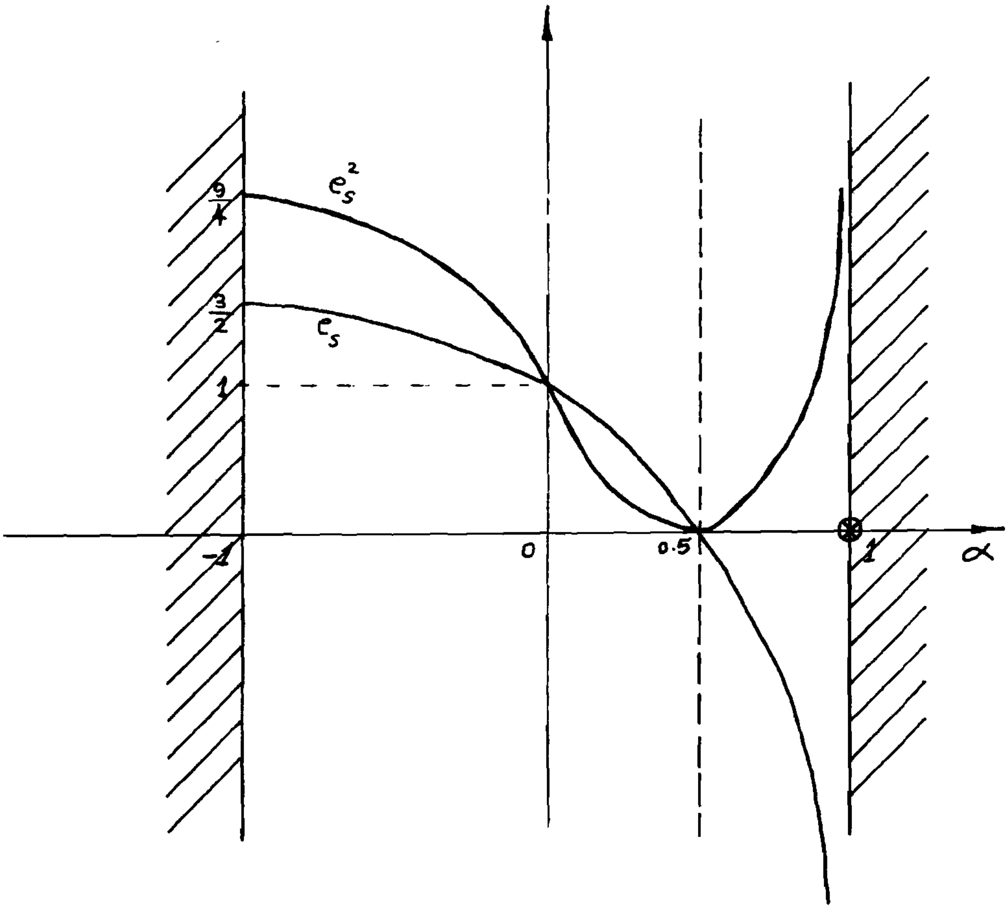


fig.(4.5.1.)

For alternating input sequence (pole at z = -1), $u(z)$ is given by

$$u(z) = \frac{z}{z+1} \quad , \quad \text{and}$$

$$e_a(z) = \frac{1}{(z-a)(z-\alpha)} \cdot \frac{z}{(z+1)}$$

Since the function $e_a(z)$ has a pole on the unit circle we can not use the final value theorem; alternatively we use the conventional inverse z-transform techniques

$$\frac{e_a(z)}{z} = \frac{1}{(z-a)(z+1)} - \frac{1}{(z-\alpha)(z+1)}$$

$$= \frac{1}{(1+a)}\left(\frac{1}{z-a} - \frac{1}{z+1}\right) - \frac{1}{1+\alpha}\left(\frac{1}{z-\alpha} - \frac{1}{z+1}\right)$$

$$= \left(\frac{1}{1+a} - \frac{1}{1+\alpha}\right)\cdot\frac{z}{z+1} + \frac{1}{1+a}\cdot\frac{z}{(z-a)} - \frac{1}{1+\alpha}\cdot\frac{z}{(z-\alpha)}$$

From tables of z-transform we can write the following

$$e_a(kT_s) = \frac{(\alpha-a)}{(1+a)(1+\alpha)} \cos\frac{\pi}{T_s}(kT_s) + \frac{1}{1+a}\, a^{kT_s/T_s} - \frac{1}{1+\alpha}\, \alpha^{kT_s/T_s}$$

$$e_a(k) = \frac{(\alpha-a)}{(1+a)(1+\alpha)} \cos \pi k + \frac{1}{1+a}\, a^{k} - \frac{1}{1+\alpha}\, \alpha^{k}$$

but, since $|\cos k\pi| = 1$, we can write for the final error ($e_{a\infty}$)

$$\lim_{k\to\infty} |e_a(k)| = \left| \frac{\alpha-a}{(1+\alpha)(1+a)} \right| = |e_{a\infty}|$$

where the conditions of stability $|a| < 1$ and $|\alpha| > 1$ were used. The final squared error($e^2_{a\infty}$)is given by

$$e^2_{a\infty} = \frac{(\alpha-a)^2}{(1+a)^2(1+\alpha)^2}$$

For the process parameter a = 0.5, then

$$e_{a\infty} = \frac{\alpha-0.5}{(1+0.5)(1+\alpha)} = \frac{2}{3}\cdot\left(\frac{\alpha-0.5}{1+\alpha}\right)$$

$$e^2_{a\infty} = \frac{4}{9}\cdot\left(\frac{\alpha-0.5}{1+\alpha}\right)^2$$

fig. (4.5.2.) shows a plot for $e_{a\infty}$ and $e^2_{a\infty}$ as a function of the model parameter α, also shown is the location of the input signal pole (z = -1, for alternating input) on the parameter axis.

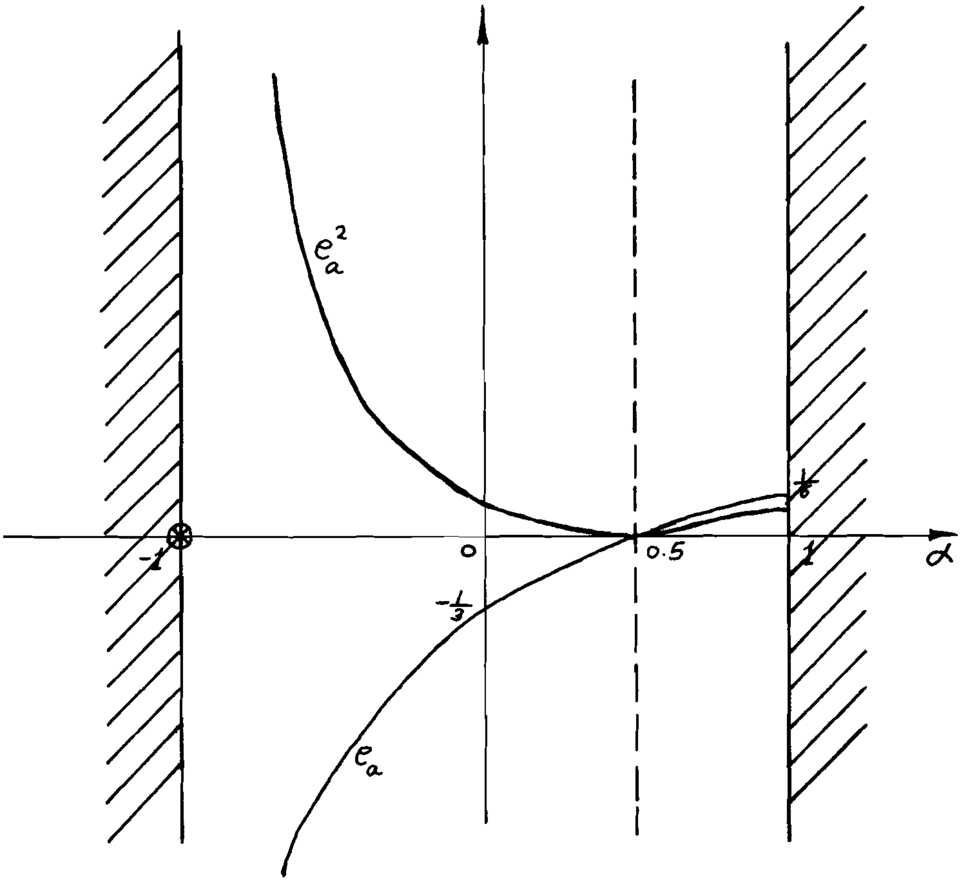


fig. (4.5.2.)

For a combined step and alternating inputs we can apply the superposition principle by adding their respective responses. In fig. (4.5.3.) is shown the corresponding final squared error ($e_{c\infty}^2$)

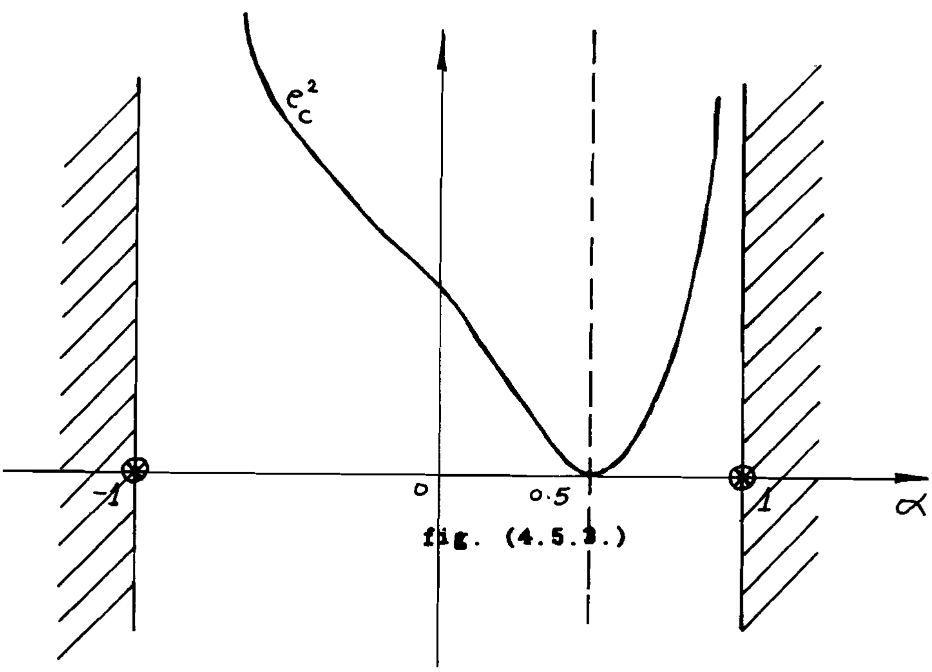


fig. (4.5.3.)

It is of much interest to note that e_∞ and, consequently, e_∞^2 approach infinity as the model pole approaches the input signal pole on the stability boundary. This is due to the fact that the model response w_∞ approaches infinity as the model pole approaches the input signal pole on the stability boundary; this can be easily checked in the last example.
It would be of value to present a proof for the validity of the above conclusion in the general case of higher order model when subjected to general periodic input sequence.
Assume that the model transfer function is given by

$$H(z) = C \cdot \frac{R(z)}{Q(z)}$$

where C is a constant, R(z) and Q(z) are rational polynomials in z.
Q(z) may be written as follows

$$Q(z) = (z-q_1)\cdots\cdots(z-q_m)\cdots\cdots(z-q_n) = \prod_{m=1}^{n}(z-q_m)$$

where the q_m's represent the model poles which may be either real or complex. It is assumed that the order of Q(z) is equal to or higher than that of R(z) and that H(z) contains no multiple pole. The magnitude of q_m is less than 1, i.e. $|q_m| < 1$, since the model is assumed to be stable.
Thus,

$$H(z) = \frac{w(z)}{u(z)} = C \cdot \frac{R(z)}{\prod_{m=1}^{n}(z-q_m)}$$

since the input signal is a stationary periodic sequence, its z-transform may be written as

$$u(z) = A \cdot \frac{N(z)}{D(z)}$$

where A is a constant, N(z) and D(z) are rational polynomials in z.
D(z) may be written as follows

$$D(z) = (z-d_1)\cdots\cdots(z-d_s)\cdots\cdots(z-d_\ell) = \prod_{s=1}^{\ell}(z-d_s)$$

where d_s's represent the input signal poles, they all ly on the unit circle, i.e. $|\alpha_s| = 1$. The order of D(z) is equal to or higher than that of N(z) and D(z) contains no multiple pole. u(z) can be written as

$$u(z) = A \cdot \frac{N(z)}{\prod_{s=1}^{\ell}(z-d_s)}$$

The z-transform of the output sequence is given by

$$w(z) = H(z) \,.\, u(z)$$

$$= C.A.\frac{R(z).\,N(z)}{\prod_{m=1}^{n}(z-q_m)\prod_{s=1}^{\ell}(z-d_s)}$$

$$= \frac{G(z)}{\prod_{m=1}^{n}(z-q_m)\prod_{s=1}^{\ell}(z-d_s)} = \frac{G(z)}{\prod_{i=1}^{n+\ell}(z-p_i)} = \frac{G(z)}{F(z)}$$

The corresponding output sequence can be expressed using the inverse z-transform

$$w(k) = \frac{1}{2\pi j}\oint_{\Gamma} w(z)\, z^{k-1}\, dz$$

where Γ is the path of integration in the z-plane which encloses all singularities of the integrand $w(z)z^{k-1}$. By applying Cauchy's theorem the value of the integral is given by the sum of all residues of $w(z)z^{k-1}$ inside the contour Γ. Thus

$$w(k) = \sum_{i=1}^{n+\ell} \text{residues of } w(z)z^{k-1} \text{ at } p_i$$

where p_i's are the poles of $w(z)z^{k-1}$

$$w(k) = \sum_{i=1}^{n+\ell}\frac{G(p_i)}{F'(p_i)}\cdot(p_i)^k$$

$$F'(p_i) = \left.\frac{dF(z)}{dz}\right|_{z=p_i} = \prod_{\substack{j=1\\ j\neq i}}^{n+\ell}(p_i-p_j)$$

And $w(k)$ can be written in the following form

$$w(k) = \sum_{m=1}^{n}\frac{G(q_m)\cdot(q_m)^k}{\prod_{\substack{t=1\\ t\neq m}}^{n}(q_m-q_t)\prod_{s=1}^{\ell}(q_m-d_s)} + \sum_{s=1}^{\ell}\frac{G(d_s)\cdot(d_s)^k}{\prod_{\substack{r=1\\ r\neq s}}^{\ell}(d_s-d_r)\prod_{m=1}^{n}(d_s-q_m)}$$

The terms included in the first summation give rise to a decaying responses, since $|q_m| < 1$. The terms included in the second summation give rise to periodic stationary responses, since $|d_s| = 1$. The presence of the expression $\prod_{m=1}^{n}(d_s-q_m)$ in the terms of the second summation explains the fact that if one or more of the model poles approach one or more of the periodic input signal poles, then the model response grows infinitely. This must be clear since the amplitude of the stationary oscillations due to the terms in the second summation will grow

infinitely as the value of (d_s-q_m) approaches zero (i.e. the model pole (q_m) approaches the signal pole (d_s)).
The above analysis explains the behaviour of the error (e) and its square (e_∞^2) in the parameter space in the general case when the model is subjected to a stationary periodic input sequence.

Now, we are in a position where we can say more about the effect of the signal poles distribution on the error criterion function in the parameter space. Consider the second order discrete model

$$H(z) = \frac{a_6 z^2 + a_5 z + a_4}{z^2 + a_2 z + a_1}$$

where two unknown parameters a_1 and a_2 are to be estimated in a least squares sense i.e. by minimizing the error criterion

$$E = \sum_{k=1}^{N} e^2(k) \quad ,$$

where

$$e(k) = y(k) - w(k) \quad ,$$

w(k) is the model output and y(k) is the process output, whose true parameters are $a_1 = 0$ and $a_2 = 0$. The process and the model are subjected to a periodic binary noise sequence whose poles distribution in the a_1-a_2 plane is given in fig.(4.4.2.) and will be repeated here for convenience.
The pole at z=1 is projected on the line $1+a_2+a_1=o$ and three conjugate pairs of complex poles on the unit circle are projected on the vertical side $a_1=1$, their locations are indicated by cross marks as shown by fig. (4.5.4.) . Since the model response w(k) and, consequently, e(k) approach infinity as the model poles approach one or more of the input signal poles, we expect that the error criterion function E possesses singular points on the locations of the input signal poles in the parameter plane. The contour lines of the error criterion function E are sketched in fig. (4.5.5.). It is obvious that none of the "hill climbing" techniques may successfully be used to locate the minimum if the starting point is chosen arbitrarily inside the stability triangle, convergence may be achieved if the starting point is sufficiently far from the irregularities of the function E along the stability borders. Constrained minimization techniques may be used more successfully if the opposite gradient direction along the stability borders points towards the interior of the stability region, which is not always the case.

4.6. The choice of a suitable input signal:

After having answered the first question about the effect of the input spectrum on the shape of the error criterion function, we come to the second question namely; is it possible to find a suitable input signal? By suitable input signal we mean one which affects the shape of the error criterion function in a desirable way (for the implemented minimization technique) and preferably avoiding us a constrained minimization problem.

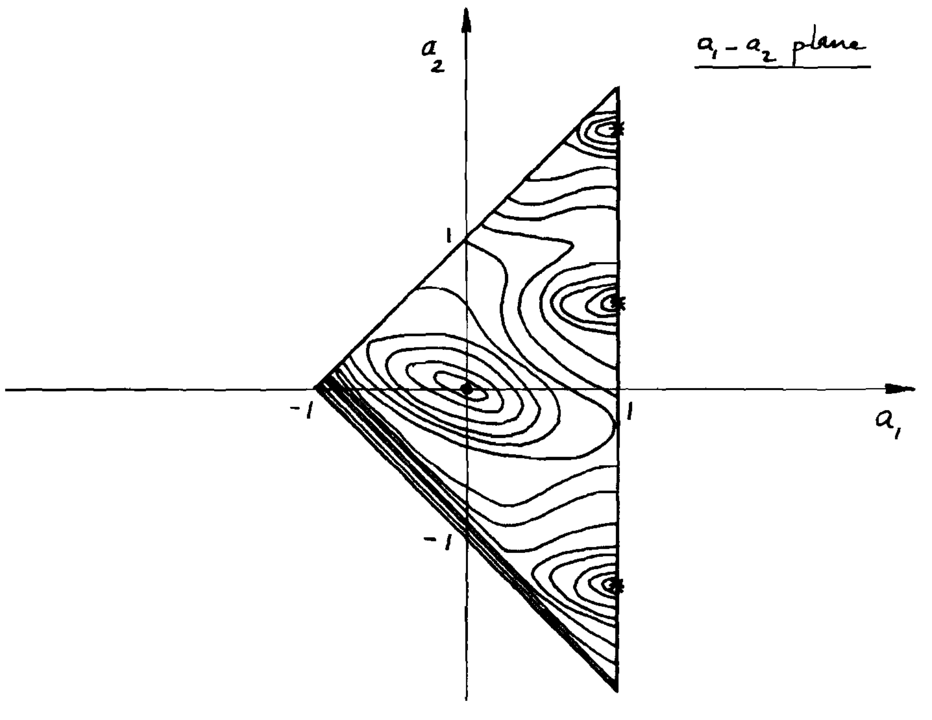


fig. (4.5.4.)

From fig. (4.5.4.) considered in the last section, we feel intuitively that if all stability borders are singular boundaries of the error criterion function then this will result in a suitable shape of the error criterion within the stability region where, only, one global minimum exists at the true values of the parameters. It is obvious that singularity along the triangle sides $1+a_2+a_1=o$ and $1-a_2+a_1=o$ are attained if the input signal possesses poles at the points z=1 (step input sequence) and z=-1 (alternating input sequence). Points of singularity along the vertical side $a_1=1$ corresponds to the poles located on the unit circle (excluding the points z=1 and z=-1). Consequently, it corresponds to the spectral components in any frequency strip of size ω_s. This implies that singularity along the vertical side is approached, only, by increasing the density of the equidistant spectral components in each of the frequency strips of size ω_s and distributing the signal power uniformly over these spectral components. An intuitive and acceptable solution to approach uniformity in case of MLS input is to decrease the sampling frequency such that the repetition of the spectrum $\frac{2\pi}{T_s}$ would occur within one period of the spectrum envelope $\frac{2\pi}{\theta}$; a decrease in the density of the spectral components in the frequency strips will result. This can be compensated by lowering the frequency of the input signal $\frac{2\pi}{T}$, which means increasing the number of states (each of duration θ) in one period of the signal (i.e. increasing the number of register elements of the MLS generator).

A theoretical optimal sample rate, which results in a uniform spectrum may be derived as follows for any signal of finite band width. If the power spectrum envelope is given by $\phi_{uu}(\omega)$ where

$$\phi_{uu}(\omega) = 0 \qquad \text{for all } \omega \geqslant \omega_c ,$$

if sampling is performed on the signal with frequency ω_s, the resultant spectrum of the sequence is given by

$$\phi^*_{uu}(\omega) = \sum_{n=W-N}^{W+N} \phi_{uu}(\omega - n\omega_s)$$

where

$$W = \omega \div \omega_s \qquad (\text{integer division})$$

N is an integer satisfying $N \geqslant \frac{\omega_c}{\omega_s}$.

Since $\phi^*_{uu}(\omega)$ is periodic with frequency period ω_s, it can be expanded in a Fourier series in the form

$$\phi^*_{uu}(\omega) = \sum_{k=-\infty}^{\infty} C_k \, e^{jkT_s\omega}$$

$$C_k \triangleq \frac{1}{\omega_s} \int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} \phi^*_{uu}(\omega) \, e^{-jkT_s\omega} \, d\omega$$

for a uniform spectrum $\phi^*_{uu}(\omega)$

$$C_o = \frac{1}{\omega_s} \int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} \phi^*_{uu}(\omega)\, d\omega \neq 0$$

and

$$C_k = \frac{1}{\omega_s} \int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} \phi^*_{uu}(\omega)\, e^{-jkT_s\omega}\, d\omega = 0 \qquad \text{for all } k \neq 0$$

If a lower bound for the optimal sampling rate (ω_{sl}) is to be known in priori, we can calculate a higher bound for N ($N \geqslant \frac{\omega_c}{\omega_{sl}}$) which can be used in the expression for $\phi^*_{uu}(\omega)$, then C_k is given by

$$C_k = \frac{1}{\omega_s} \int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} \sum_{n=-N}^{N} \phi(\omega - n\omega_s)\, e^{-jkT_s\omega}\, d\omega$$

the solution of the equation C_k=o for all k≠o, if it exists, will deliver the sampling fequency ω_s which results in a uniform spectrum. The solution ω_s is, in general, not unique. The highest $\hat{\omega}_s$ is to be the optimal; this is due to the fact that, in a fixed time interval more samples can be generated from the continuous signal for higher sampling rate; furthermore, for a periodic signal, the number of spectral components contained in each frequency strip of size ω_s, obviously, increases for higher sampling rate, which is desirable in our case.

In the following we show the effect of changing the sampling frequency ω_s around a chosen nominal value $\omega_s = \frac{2\pi}{T_s} = \frac{2\pi}{T} N$.
Note that N is an odd number given by (2^n-1) where n is the number of register elements used in the MLS generator.
From fig. (4.6.2) it is seen that if the sampling frequency ($\frac{2\pi}{T_s}$) is not an integer multiple of the signal frequency ($\frac{2\pi}{T}$), the density of the spectral components is approximately doubled; on the other hand the spectral components of the overlaped spectrum will not add together leading to a nonuniform distribution of the signal power over the spectral components. A rather uniform distribution is obtained for sampling frequencies of integer multiples of the signal frequency. This is because the spectral components of the overlaped spectrum add together.
An advantage can be gained in the case where

$$\omega_s = \frac{2\pi}{T}(N \pm 1)$$

is the presence of a spectral component at frequencies $\frac{\omega_s}{2} \pm k\omega_s$, $k = 0,1,2,\ldots$

This is transfered into a pole on the unit circle of the z-plane at the point z=-1, this means the presence of a singularity along the upper side of the stability traingle ($1+a_2-a_1=o$) in the parameter plane; a strong requirement of a suitable input signal.

So far, we have been concerned with the choice of the sampling rate for a suitable input sequence generated from continuous a periodic signal; the choice of a suitable period for the continuous signal can, simply, be found by realizing that the effect of periodicity is just, the discretization of the continuous spectrum corresponding to a certain period of the signal, the resulting spectral components are equidistant with a period of $\frac{2\pi}{T}$ (T is the signal period). It follows directly that an increase in the signal period will result in a higher spectral density in the frequency domain. This is a desirable effect on the spectrum of the input signal. A finite optimal for the signal frequency does not exist; practical constraints may put a lower limit on the signal frequency. For instance, in the case of binary noise signal with a clock frequency ($\frac{1}{\theta}$), the signal period is given by

$$T = (2^n - 1)\ \theta$$

An increase in the signal period is obtained by increasing the number of register elements used in the MLS generator while keeping the clock frequency unchanged. Another limitation is that the number of states in one period should not be large as a requirement for faster generation of input sequence, faster simulation on a digital computer and

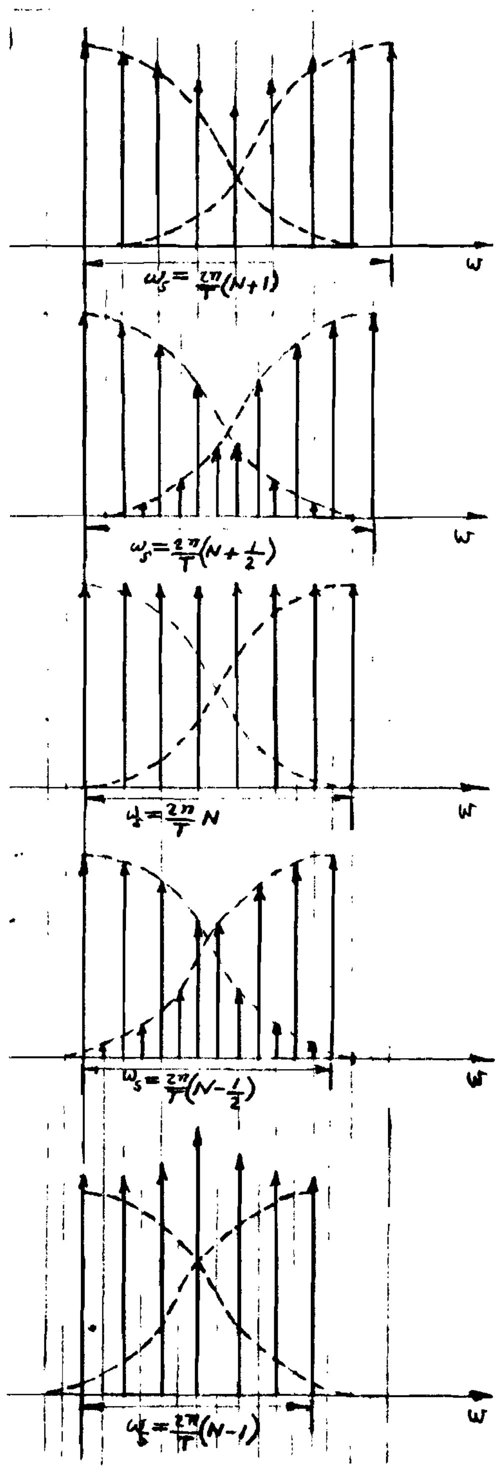


fig.(4.6.1.)

less memory storage requirements. The effect of the signal period on the error contour lines is demonstrated by increasing the number of register elements in the MLS generator and fixing the number of samples per state of the signal. The effect of the sampling rate on the error contour lines is demonstrated by fixing the number of register elements in the MLS generator and increasing the number of samples per state of the signal. This is shown in figures (4.6.2.) where a digital computer was used for the error contour lines plottings.

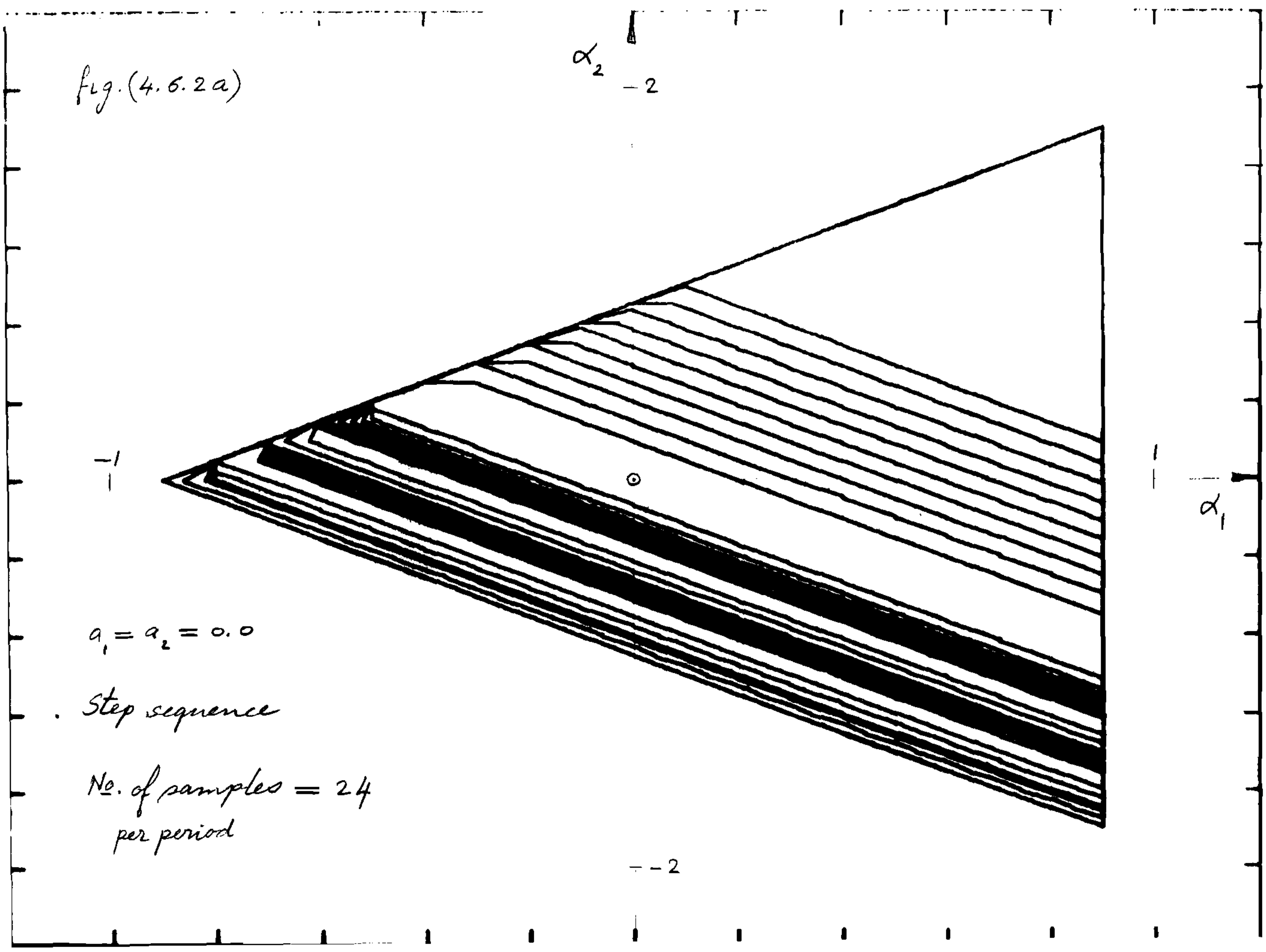


fig.(4.6.2a)
α_2
2
−1
1
α_1
−2
$a_1 = a_2 = 0.0$
Step sequence
No. of samples = 24 per period

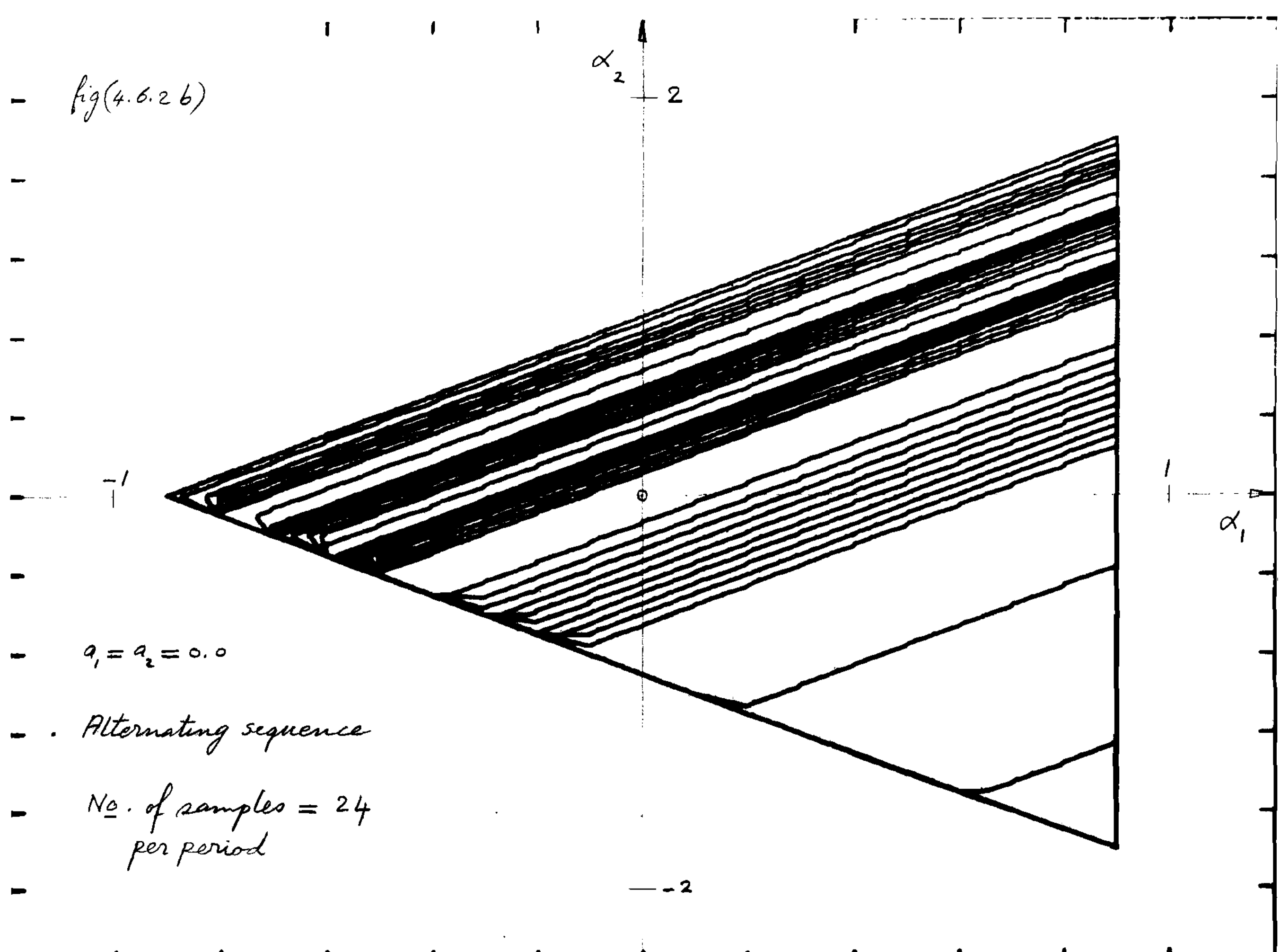
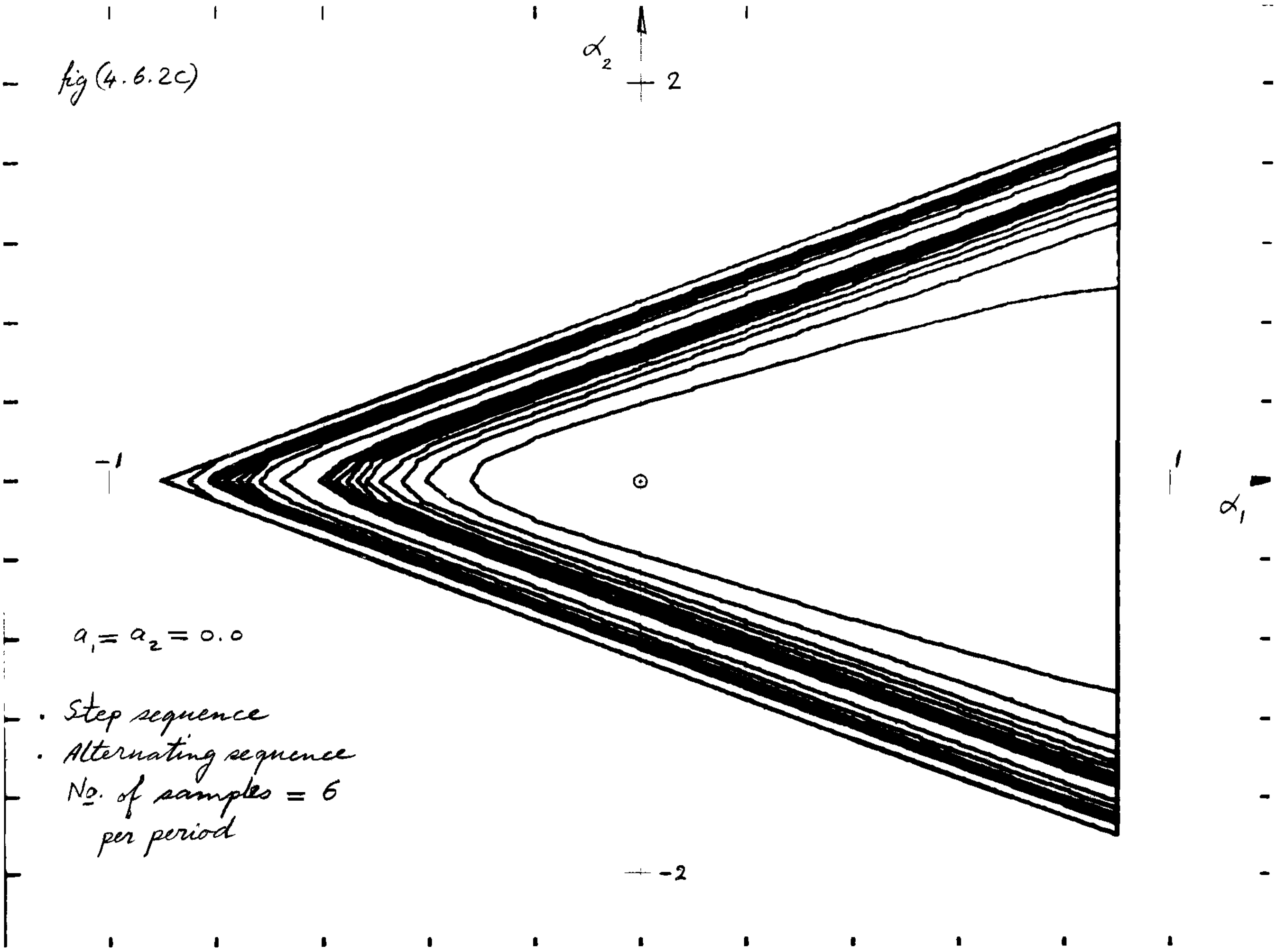

fig(4.6.2b)
α_2
2
−1
1
α_1
−2
$a_1 = a_2 = 0.0$
Alternating sequence
No. of samples = 24 per period

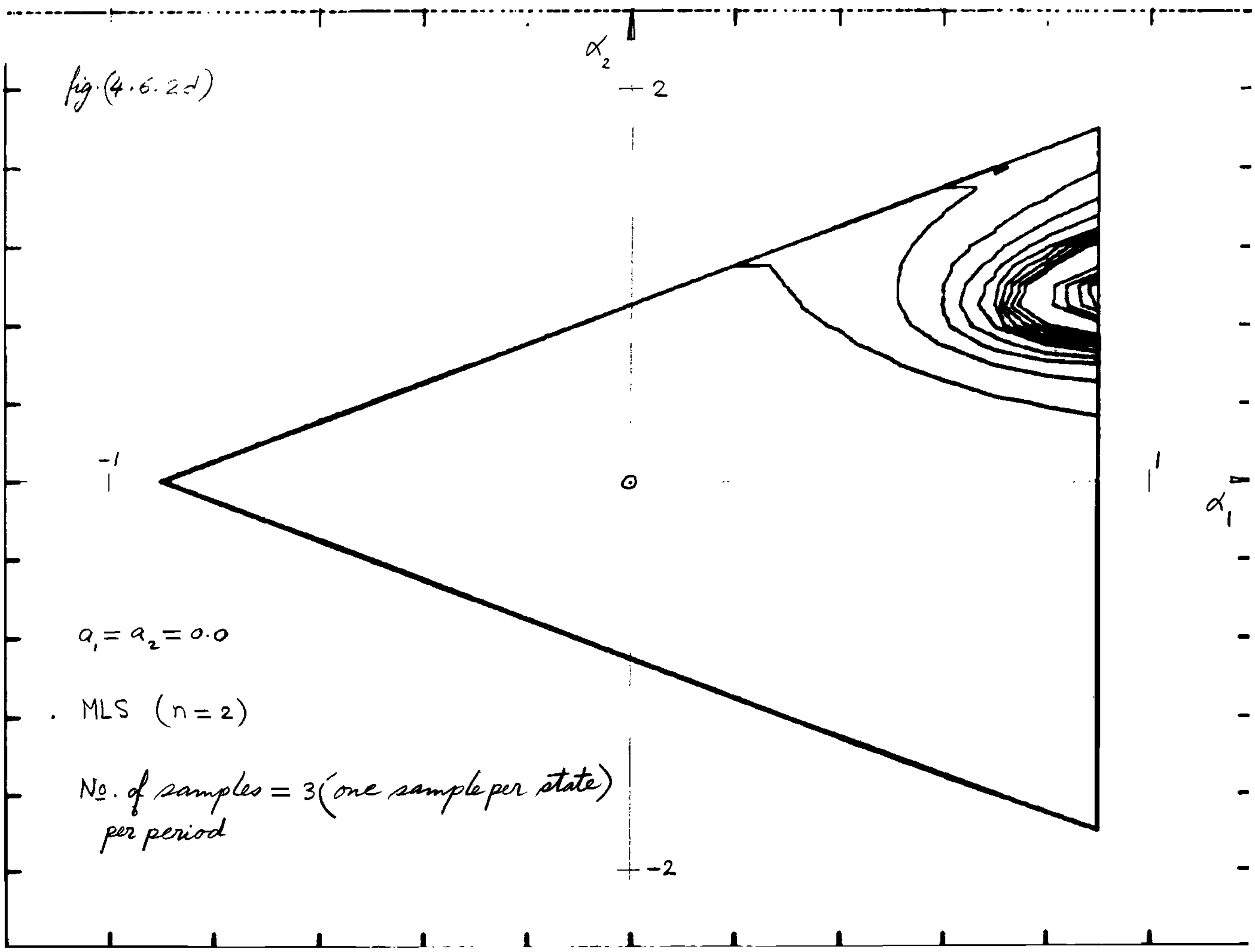
fig (4.6.2c)



$a_1 = a_2 = 0.0$

. Step sequence
. Alternating sequence
No. of samples = 6 per period

fig (4.6.2d)

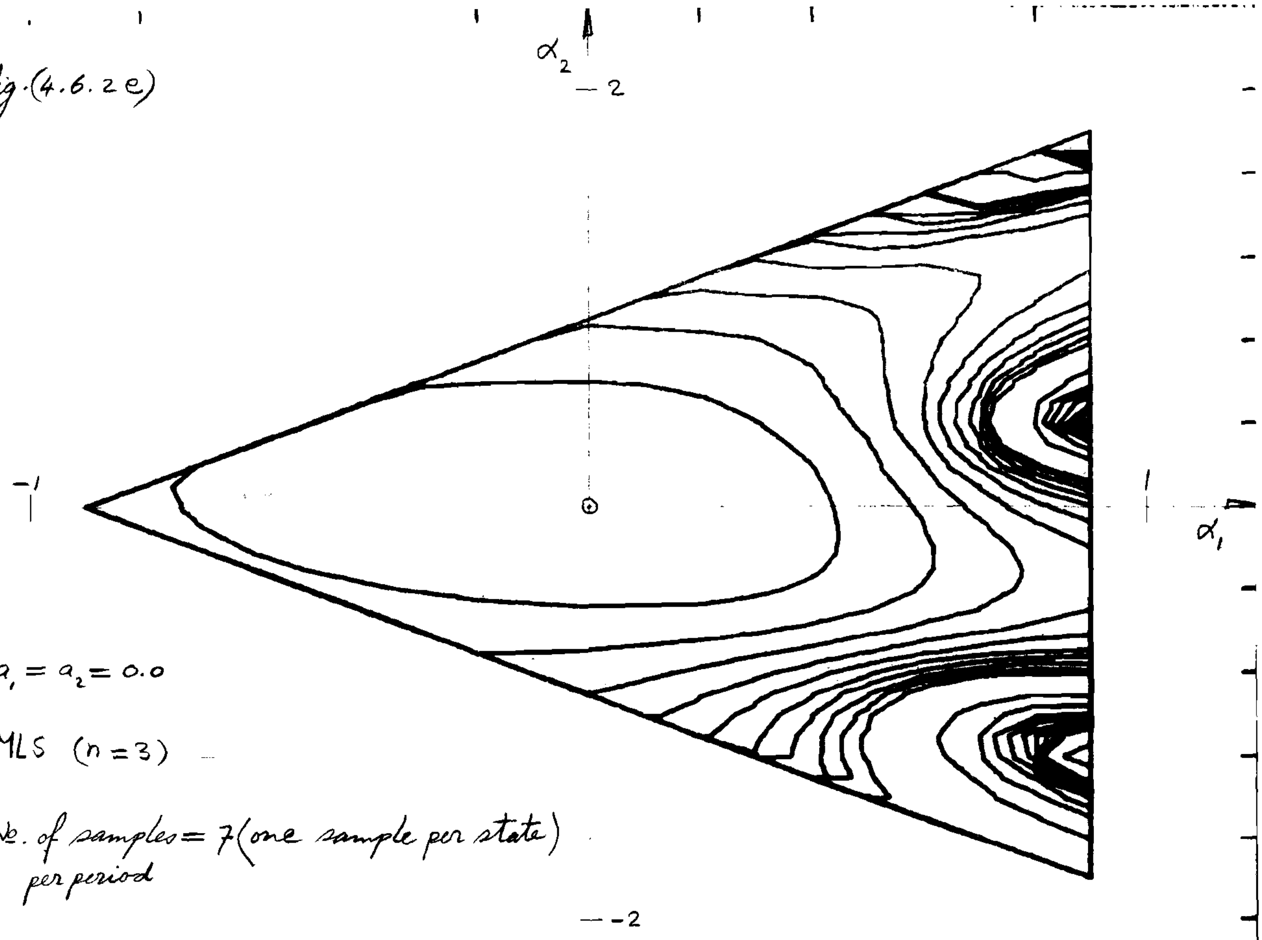


$a_1 = a_2 = 0.0$

. MLS (n = 2)

No. of samples = 3 (one sample per state) per period

fig.(4.6.2e)

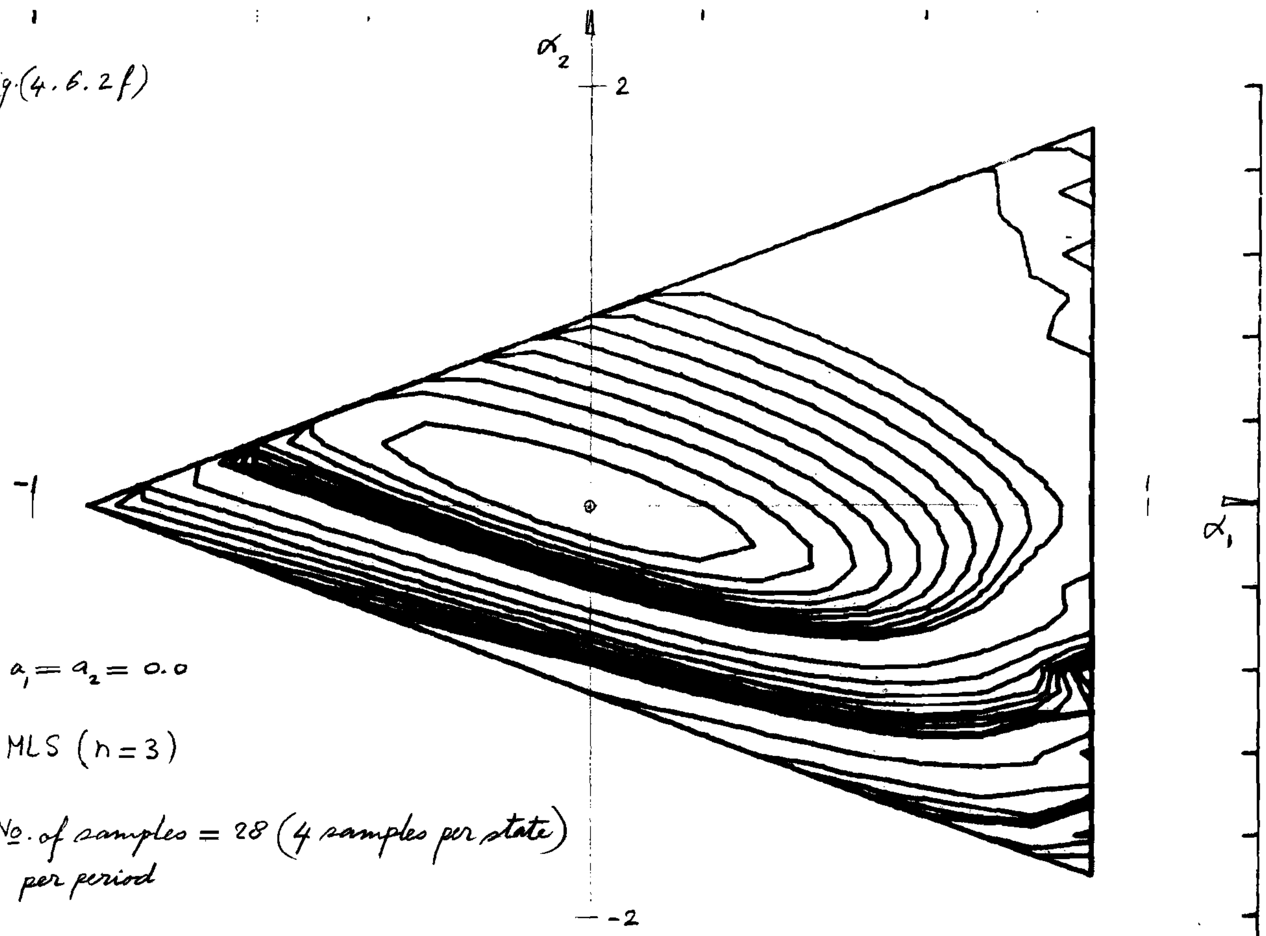


$a_1 = a_2 = 0.0$

. MLS (n=3)

No. of samples = 7 (one sample per state) per period

fig.(4.6.2f)



$a_1 = a_2 = 0.0$

. MLS (n=3)

No. of samples = 28 (4 samples per state) per period

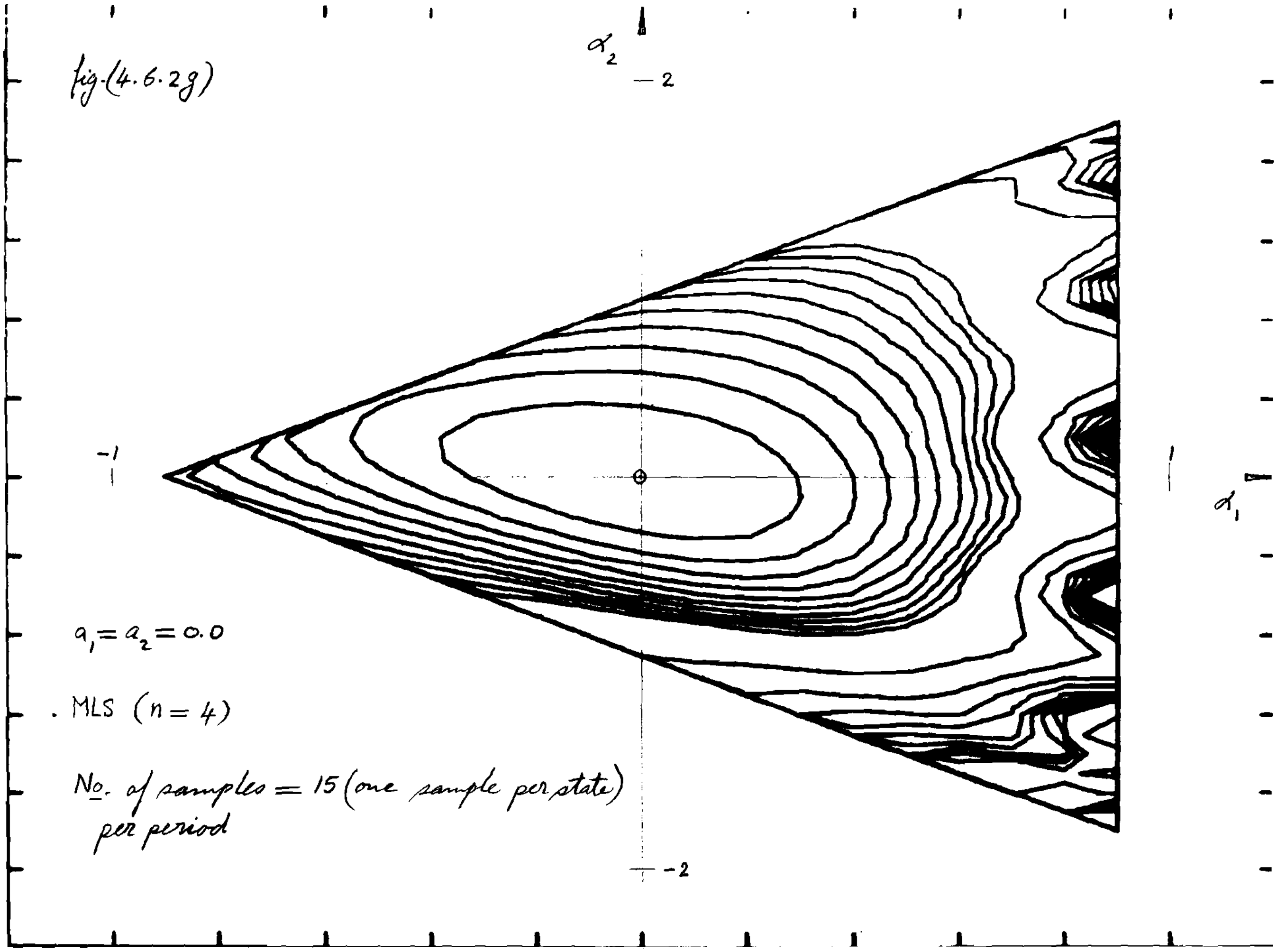
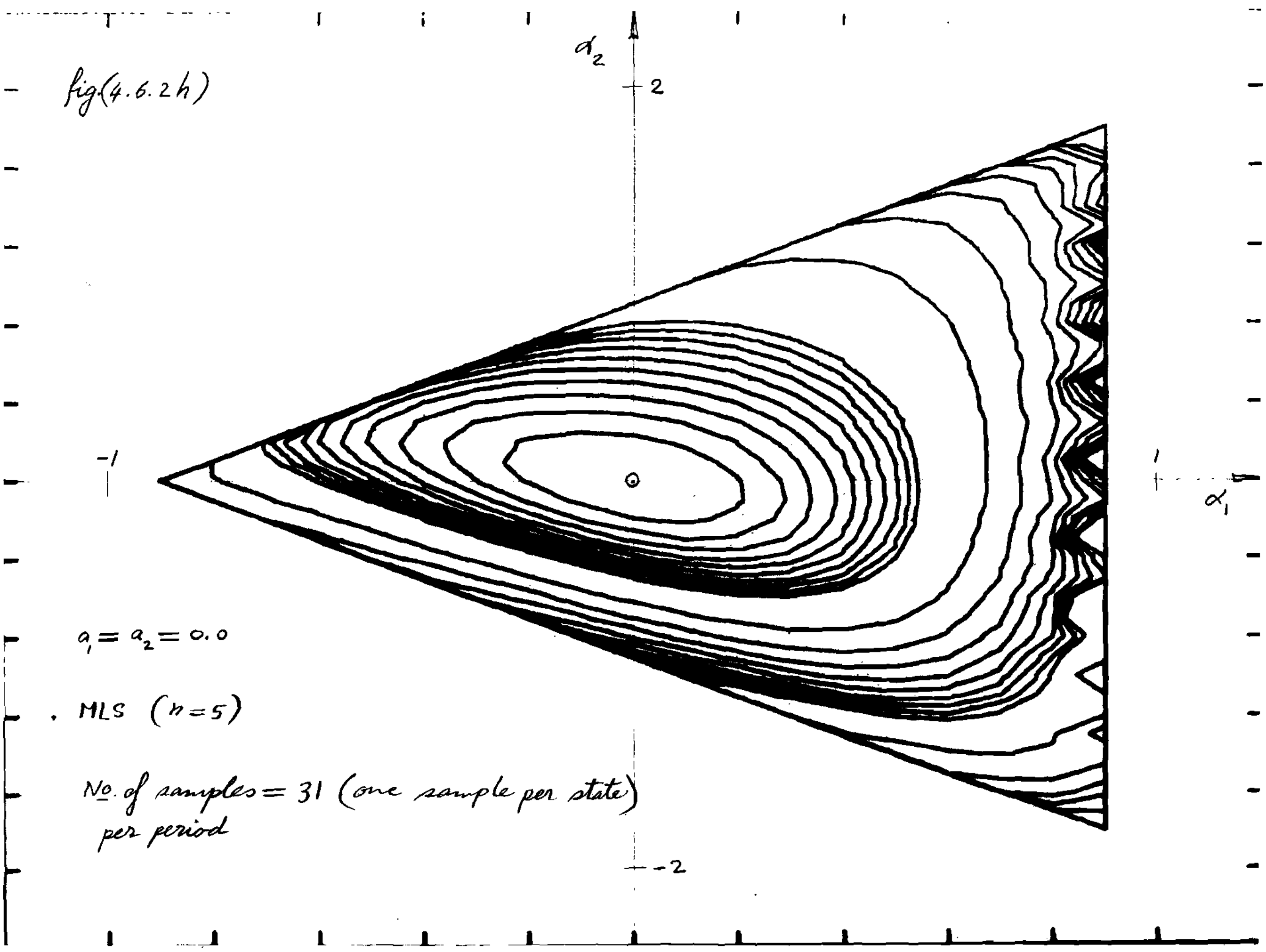


fig.(4.6.2g)

$a_1 = a_2 = 0.0$

. MLS ($n = 4$)

No. of samples = 15 (one sample per state) per period



fig(4.6.2h)

$a_1 = a_2 = 0.0$

. MLS ($n=5$)

No. of samples = 31 (one sample per state) per period

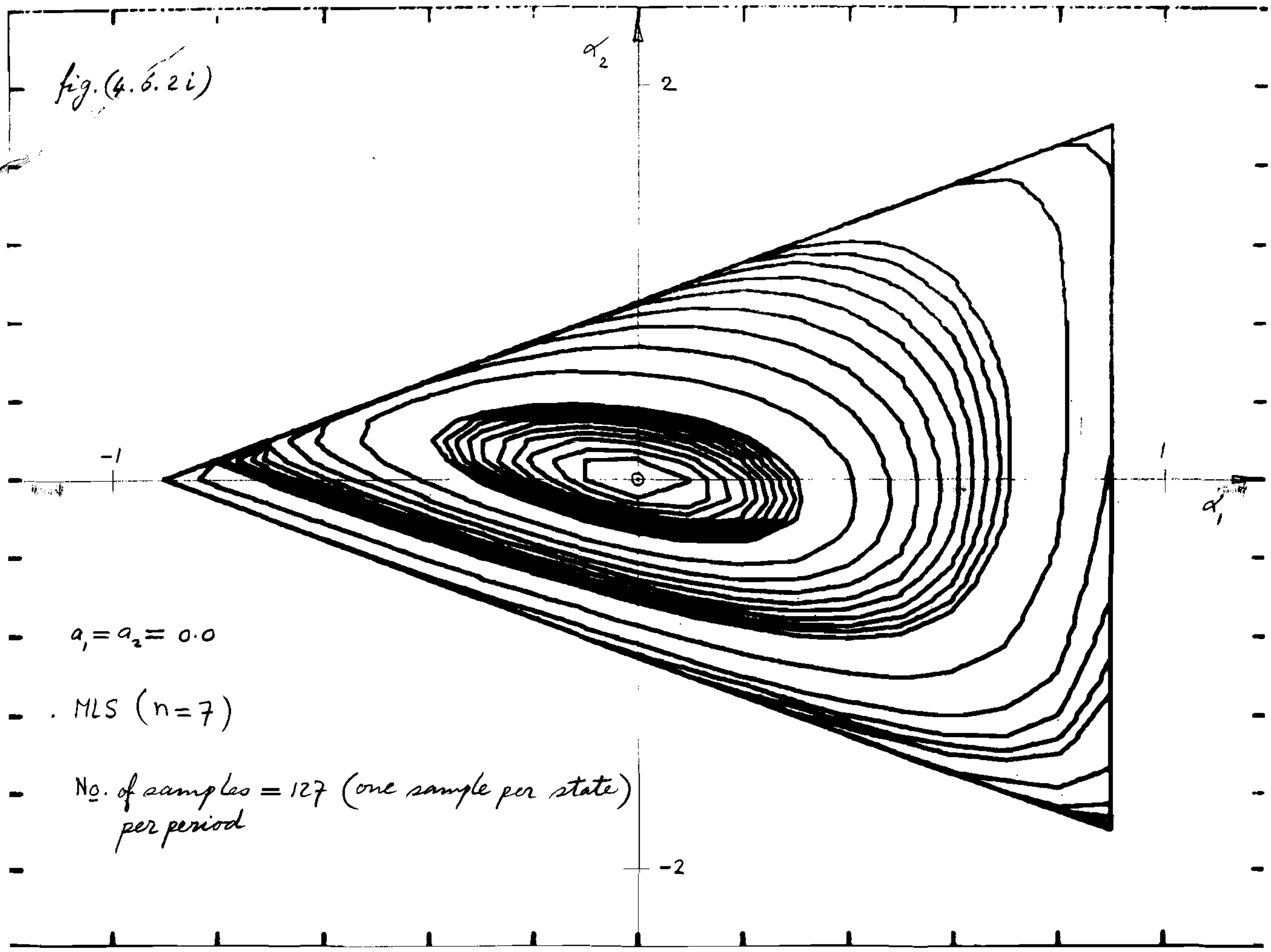


fig.(4.6.2i)
α_2
2
-1
1
α_1
-2
$a_1 = a_2 = 0.0$
. MLS (n=7)
No. of samples = 127 (one sample per state) per period

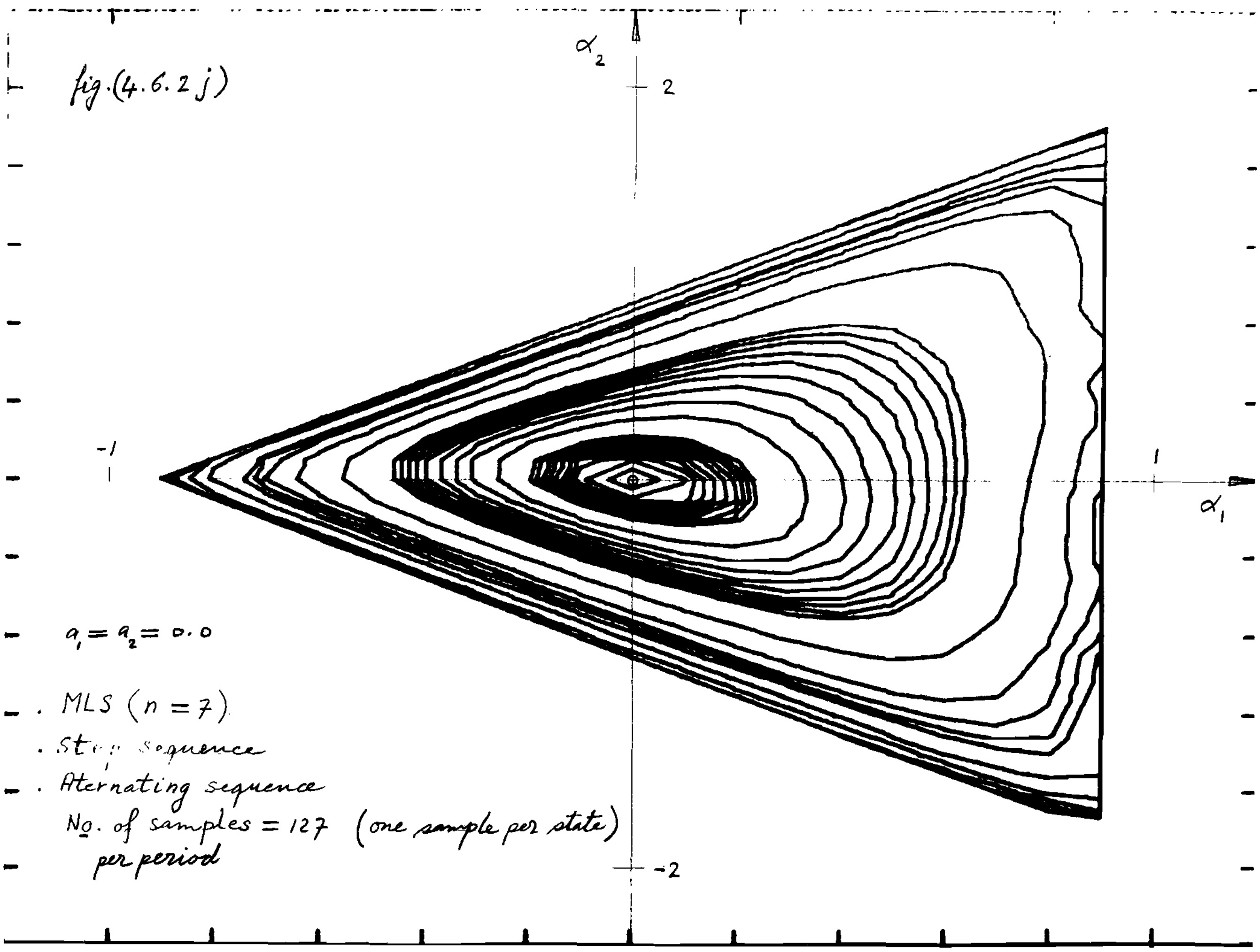


fig.(4.6.2j)
α_2
2
-1
1
α_1
-2
$a_1 = a_2 = 0.0$
. MLS (n = 7)
. Step sequence
. Alternating sequence
No. of samples = 127 (one sample per state) per period

5. Model adjustment techniques

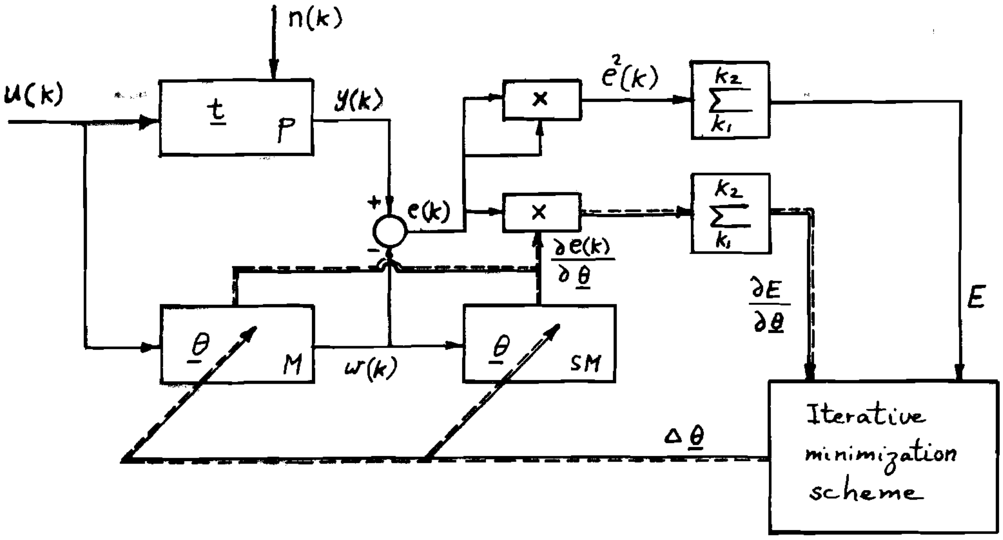


fig. 5.1.1.

Flow and processing of data in model adjustment techniques using least squares criterion.

5.1. Introduction:

Estimation in a least squares sense means the minimization of the error criterion function

$$E = \underline{e}^T \underline{e} = \sum_{k=k_1}^{k_2} e^2(k)$$

As we are concerned with models, nonlinear-in-the-parameters, equating the derivatives $\frac{\partial E}{\partial \underline{\theta}}$ to zero will lead to a set of nonlinear equations in the unknown parameters which, in general, can not be solved explicitly; solution of the set of nonlinear equations is approached iteratively. Alternatively, we can use hill climbing techniques on the error criterion function E where the location of the minimum is to be found, as the original problem suggests (see fig. 5.1.1.)

In the last chapter we have seen how the input signal can affect the optimization problem to be unconstrained one. In this chapter we concentrate our attention to the unconstrained minimization problem keeping in mind that constrained minimization problems can be converted into unconstrained ones by joining the constraints with a penalty function to the object function $E(\underline{\theta})$.

The methods we shall discuss are iterative "hill climbing" techniques. We start with an initial guess $\underline{\theta}_1$ which has to be in the same global region as the location of the required minimum. We generate a sequence $\underline{\theta}_2$, $\underline{\theta}_3$, ... which should converge to $\hat{\underline{\theta}}$ at which $E(\underline{\theta})$ is minimum. $\underline{\theta}_i$ is the i^{th} iterate and the computation of $\underline{\theta}_{i+1}$, is the i^{th} iteration. The model is adjusted to the new parameter set after each iteration. The sequence is terminated after a finite number (N) of iterations and we accept $\underline{\theta}_N$ as an approximation to $\hat{\underline{\theta}}$.

The sequence is generated by the relation

$$\underline{\theta}_{i+1} = \underline{\theta}_i + \Delta \underline{\theta}_i$$

where $\Delta\underline{\theta}_i$ is the i^{th} step given by

$$\Delta \underline{\theta}_i = \lambda_i \, \underline{p}_i$$

where $\underline{p}_i$ is a vector in the proposed i^{th} search direction, and λ_i is a scalar such that the i^{th} step is acceptable (i.e. $E(\underline{\theta}_{i+1}) < E(\underline{\theta}_i)$).

Here we discuss in some detail about the condition of acceptability. In the i^{th} iteration starting from $\underline{\theta}_i$, we search for $\underline{\theta}_{i+1}$ in a direction $\underline{p}$ along which $\underline{\theta}$ is given by

$$\underline{\theta}(\lambda) = \underline{\theta}_i + \lambda \, \underline{p}$$

In that direction the criterion function E varies as λ is changed, consequently, it becomes a function of λ alone.

$$E_{ip}(\lambda) = E\left(\underline{\theta}_i + \lambda \, \underline{p}\right)$$

its derivative is given by

$$v_{ip}(\lambda) = \frac{d\, E_{ip}(\lambda)}{d\lambda} = \left(\frac{\partial E}{\partial \underline{\theta}}\right)^T \frac{\partial \underline{\theta}}{\partial \lambda}$$

$$= \underline{g}^T(\underline{\theta}) \cdot \underline{p}$$

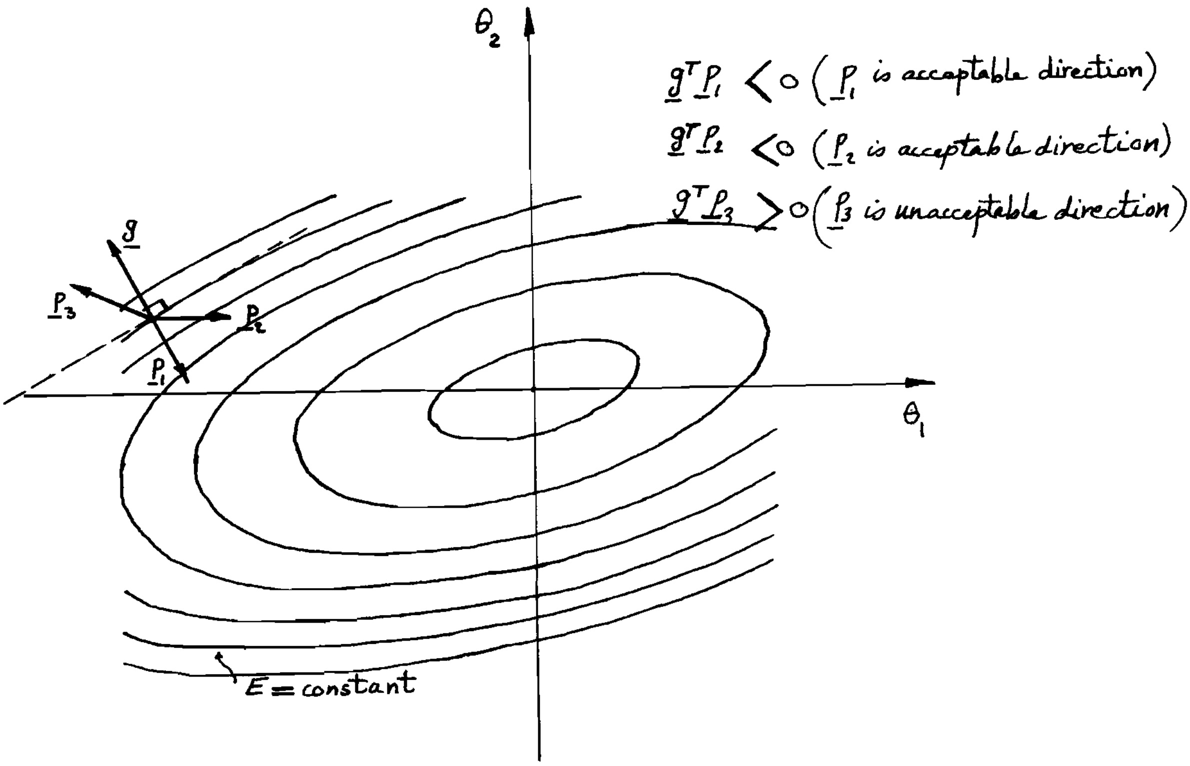


fig. 5.1.2.

where $\underline{g}(\underline{\theta})$ is the gradient vector at $\underline{\theta}$, the quantity

$$v_{ip}(o) = \underline{g}^T(\underline{\theta}_i) \cdot \underline{p}$$

is called the directional derivative of E relative to $\underline{p}$ at $\underline{\theta}_i$.
If $v_{ip}(o) < 0$ then $E(\underline{\theta})$ decreases when we start moving from $\underline{\theta}_i$ in the direction $\underline{p}$. If λ is a sufficiently small positive scalar, the step $\underline{p}$ is acceptable. On the other hand, if $v_{ip}(o) \geq 0$, there may not exist any positive value of λ for which $\underline{p}$ is an acceptable step. So, $\underline{p}$ is an acceptable direction if $v_{ip}(o) < 0$ (fig. 5.1.2.) may help as an illustration.
It can be shown that $\underline{p}$ is acceptable if and only if $\underline{p}$ can be written as

$$\underline{p} = -R\,\underline{g}_i$$

where R is a positive definite matrix and $\underline{g}_i = \underline{g}(\underline{\theta}_i)$. In that case

$$v_{ip}(o) = -\underline{g}_i^T R\, \underline{g}_i < 0$$

which proofs the above statement.
A minimization method, is acceptable if all its steps are acceptable. For any gradient method the i^{th} iteration is given by

$$\underline{\theta}_{i+1} = \underline{\theta}_i - \lambda_i R_i \underline{g}_i \qquad (5.1.1)$$

in which different strategies for the choice of λ_i and R_i leads to different minimization methods. The choice of a suitable minimization method is dependent on the problem at hand. In most of the cases the choice criteria is the speed of convergence to the minimum with the required accuracy. In most of the minimization methods which will be considered here the scalar λ_i is determined such that $v_{ip}(\lambda) \approx 0$, i.e. we search in the direction $\underline{p}_i$ for the value λ_i at which the criterion function $E_{ip}(\lambda)$ is minimum. This process is called linear search and is discussed in some detail.

5.2. Linear search (minimization along a line):

Linear search for the minimum is one of the most critical and time consuming tasks in any minimization method. This is because of the function and derivative evaluation needed during the search. It can be performed with different amount of complexity and sophistication depending on the required accuracy in locating the minimum. For instance, it will be clear that the variable metric methods may require more accuracy in locating the minimum than other methods such as the steepest descent. The method proposed here is a modified and extended version of the original method proposed by Davidon and used by Fletcher - Powell and Fletcher - Reeves. This extension is made to allow flexibility in the required accuracy for locating the minimum and to count for some critical situation where the minimum is in the close neighbourhood of the boundary of the feasible region.

As extrapolation step is estimated based on an available estimate of the unconstrained minimum ($\overset{*}{E}$) and the suppositions that the unconstrained minimum lies on the search direction, and the $E(\lambda)$ is quadratic. The step is then, given by

$$d = \frac{2\cdot(\overset{*}{E} - E(o))}{v(o)} \qquad (5.2.1)$$

where v(o) is necessarily negative as pointed previously.

An estimate of the unconstrained minimum may not be available and due to the fact that the unconstrained minimum will generally not lie on the search direction, the calculated step in equation(5.2.1) will overestimate λ_i; furthermore, $E(\lambda)$ may not be quadratic. For these reasons a modified formula for the calculation of the initial step length is given by

$$d = \frac{2\cdot(EST * E(o) - E(o))}{v(o)} \qquad (5.2.2.)$$

where EST may take any reasonable value between o and 1 ($o \leq EST < 1$) for positive values of E(o) and may take any value more than 1 ($EST > 1$) for negative values of E(o). The values of EST can be changed automatically through the minimization algorithm.

A practical restriction on the step length (h) is

$$h = d \quad \text{if} \quad 0 < d < \frac{1}{\sqrt{\underline{p}^T \underline{p}}}$$

$$= \frac{1}{\sqrt{\underline{p}^T \underline{p}}} \quad \text{otherwise}$$

this limits the step size (h.p) to be of, at most, a unit length along $\underline{p}$; this may garantee the step size to be within the scale of the problem. After that, doubled distance extrapolation is performed as shown in fig. (5.2.1.) and the sign of v(h), v(2h), v(4h), ... are examined until negative v(a) is followed by non-negative v(b), the λ is bounded in the interval (a,b) and can be calculated using a cubic interpolation formula, define

$$z = 3 \cdot \frac{E(a) - E(b)}{(b-a)} + v(a) + v(b)$$

$$w = \sqrt{z^2 - v(a) \cdot v(b)}$$

The estimate of λ is

$$\hat{\lambda} = b - \left(\frac{v(b) + w - z}{v(b) - v(a) + 2w}\right) \cdot (b-a)$$

if either E(a) or E(b) is less than $E(\hat{\lambda})$, the interpolation is repeated over the subinterval $(a,\hat{\lambda})$ or $(\hat{\lambda},b)$, respectively.

If during the extrapolation we exceeded the boundary of the feasible region (or, if fine extrapolation is required for better accuracy in locating the minimum), then one backward extrapolation is performed (with the last step length, $h_b = h$) followed by the usual double distance extrapolation (with reduced step length, $h_f = 0.1\ h$); this is shown in fig. (5.2.1.)

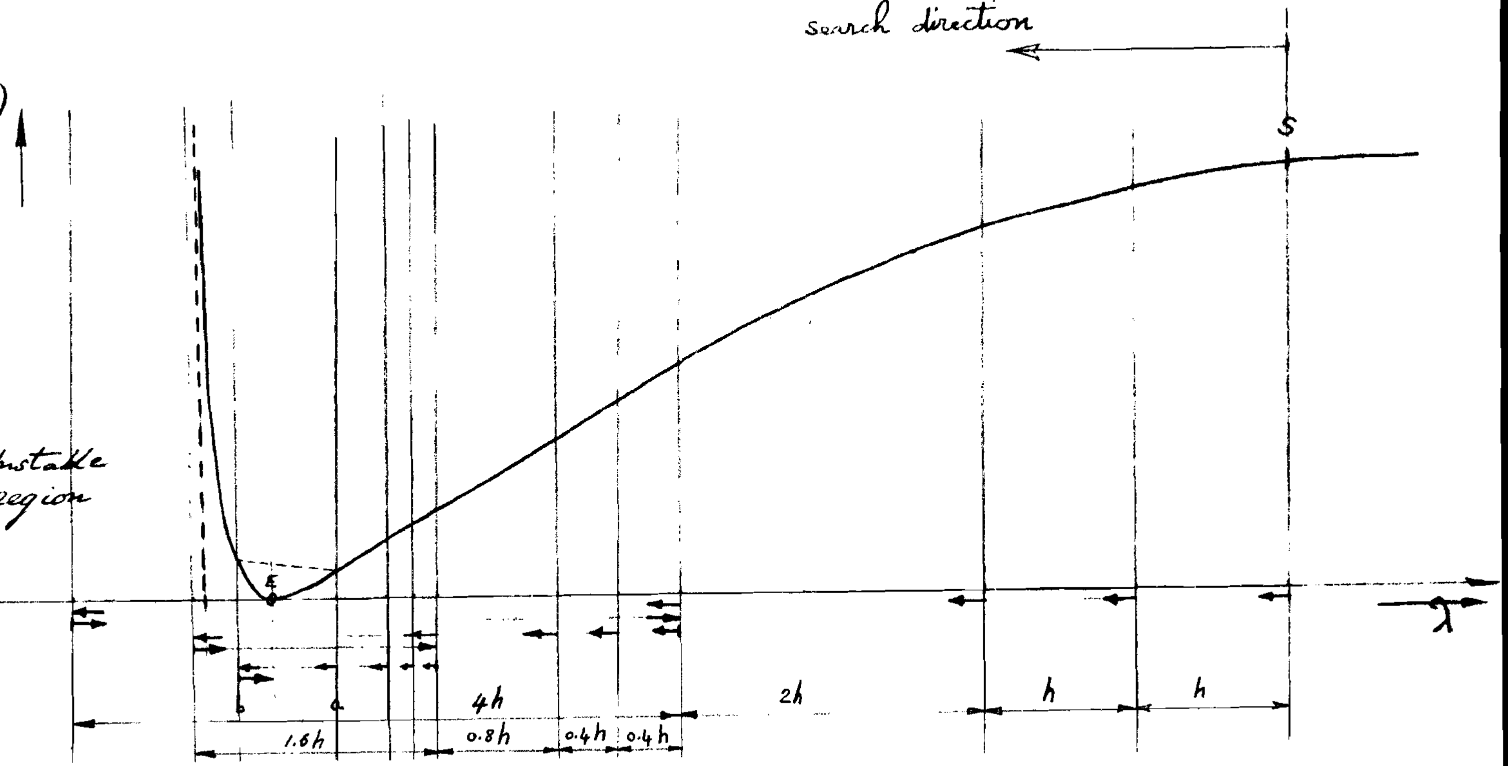


fig. 5.2.1. Linear search using doubled distance forward extrapolation , back extrapolation is used when we exceed the feasible region and followed by doubled distance forward extrapolation with a reduced step length.

fig. 5.2.2. Shows a flow chart for a linear search algorithm in which we follow the steps we discussed above.

In the remaining part of this chapter we shall discuss some minimization methods; their requirements and their convergence properties.

5.3. Steepest descent with minimization along a line:

This is the simplest of the gradient methods governed by the relation (5.1.1.) where the matrix $R_i = I$ (the identity matrix). The i^{th} iteration is given by

$$\underline{\theta}_{i+1} = \underline{\theta}_i - \lambda_i \underline{g}_i \qquad (5.3.1.)$$

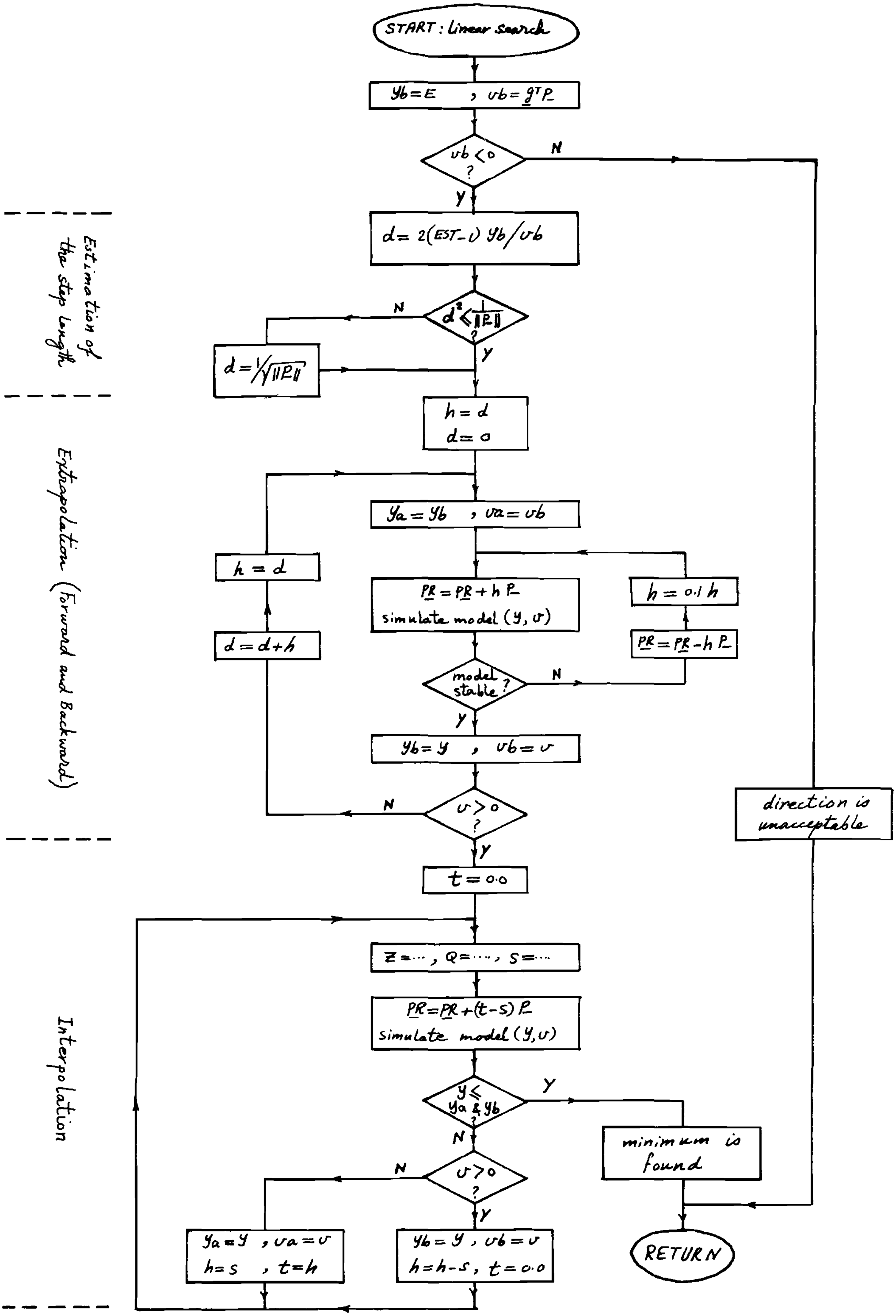

START : Linear search
Yb = E , vb = g^T P
vb < 0 ?
N
Y
d = 2(EST_V) Yb/vb
d² ≤ 1/‖P‖ ?
N
Y
d = 1/√‖P‖
h = d
d = 0
Ya = Yb , va = vb
h = d
PR = PR + h P
simulate model (y, v)
h = 0.1 h
d = d + h
PR = PR − h P
model stable ?
N
Y
Yb = y , vb = v
v > 0 ?
N
Y
direction is unacceptable
t = 0.0
Z = ··· , Q = ··· , S = ···
PR = PR + (t − S) P
simulate model (y, v)
y ≤ Ya & Yb ?
Y
N
minimum is found
v > 0 ?
N
Y
Ya = y , va = v
h = s , t = h
Yb = y , vb = v
h = h − s , t = 0.0
RETURN
Estimation of the step length
Extrapolation (Forward and Backward)
Interpolation

Fig. (5.2.2)

where λ_i is determined such that a minimum is reached in the search direction ($\underline{p}_i = -\underline{g}_i$). This direction is called the steepest descent direction since the object function initially decreases most rapidly in that direction. If the minimum in the current search direction is located accurately, the new search direction will be, approximately, orthogonal to the old one, which means

$$\underline{p}_{i+1}^T \underline{p}_i \approx 0$$

this is shown in fig. 5.3.1.

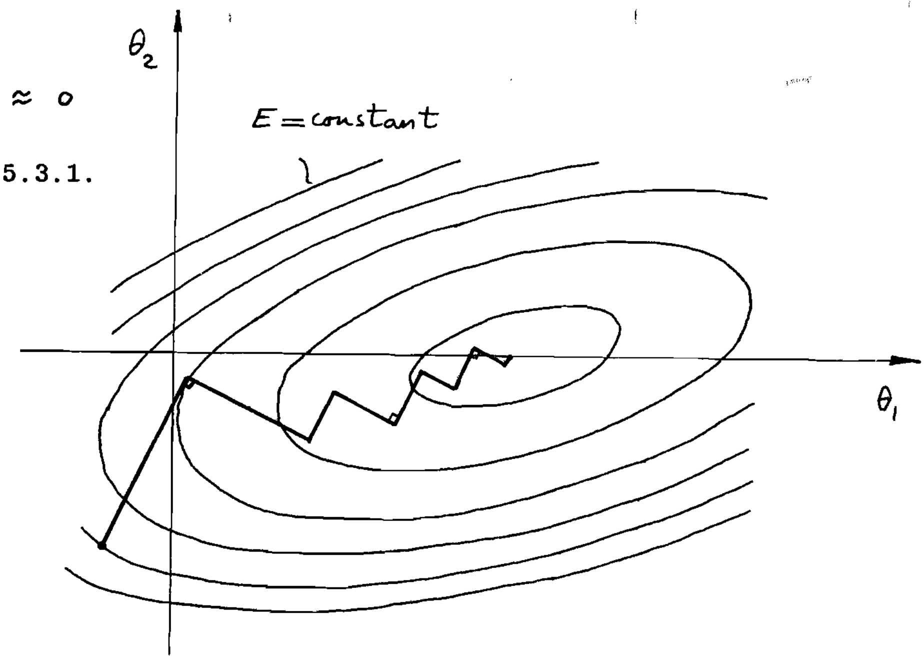


fig. 5.3.1. Steepest descent with minimization along a line; an optimization path in the two dimensional case.

It is obvious that this property may avoid us gradient evaluations after each iteration in the two dimensional case; however, the gradients can be evaluated after number of iteration to correct for the accumulated errors in the search direction.

Unfortunately, the convergence can be very slow near the minimum due to the fact that the gradient vanishes in the neighbourhood of the minimum. In situations where the curves of constant object function are steep in one direction but not in the other, inaccuracy in line minimization will lead to oscillatory behaviour in the optimization path as shown in fig. 5.3.2., convergence may not be achieved practically and the method fails to locate the minimum with reasonable accuracy.

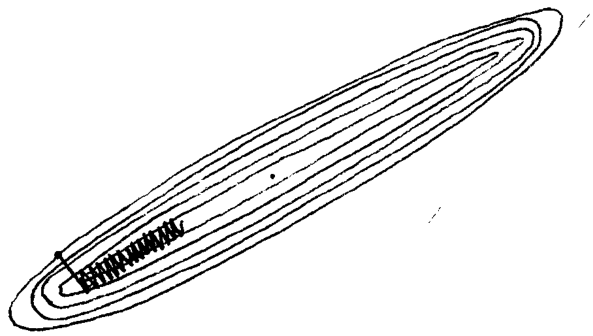


fig.5.3.2. Zigzag optimization path in case of very steep ellipsis.

5.4. Newton-Raphson method:

Quadratically convergent methods are those which converge in a specified number of iterations when applied for the minimization of quadratic functions in the unknown parameters. Any general function $E(\underline{\theta})$ can be approximated in a certain neighbourhood of a point $\underline{\theta}_i$ by the quadratic

$$Q(\underline{\theta}) = E(\underline{\theta}_i) + \underline{g}_i^T(\underline{\theta} - \underline{\theta}_i) + \frac{1}{2}(\underline{\theta} - \underline{\theta}_i)^T H_i (\underline{\theta} - \underline{\theta}_i) \qquad (5.4.1.)$$

where Taylor's expansion is utilized around the point $\underline{\theta}_i$ · $\underline{g}_i$ is the gradient $\frac{\partial E}{\partial \underline{\theta}}$ evaluated at the point $\underline{\theta}_i$, and H_i is the Hessian matrix containing the second partial derivatives evaluated at the point $\underline{\theta}_i$. A minimum for the quadratic Q $(\underline{\theta})$ exists if the Hessian matrix H_i is positive definite and the minimum is found by equating $\frac{\partial Q(\underline{\theta})}{\partial \underline{\theta}}$ to zero.

$$\underline{g}_i + H_i(\underline{\theta}_{i+1} - \underline{\theta}_i) = \underline{o}$$

which results in the i^{th} Newton-Raphson's iteration

$$\underline{\theta}_{i+1} = \underline{\theta}_i - H_i^{-1}\underline{g}_i \qquad (5.4.2.)$$

It conforms to the general formula for the gradient methods eq . (5.1.1.) with $\lambda_i = 1$ and $R_i = H_i^{-1}$. If $E(\underline{\theta})$ is quadratic, as it is the case in certain neighbourhood of the minimum, then H_i is constant and equals to the Hessian at the minimum i.e. $Q(\underline{\theta}) = E(\underline{\theta})$ at the minimum; hence $\underline{\theta}_{i+1}$ is a minimum for $E(\theta)$ and the method converges in a single iteration. If $E(\underline{\theta})$ is not quadratic the minimum is reached in more iterations provided that H_i is always positive definite; this guarantees acceptable iteration steps, see fig. 5.4.1. Disadvantages of this method are the requirements of second derivatives evaluation and the determination of the inverse matrix H_i^{-1}.

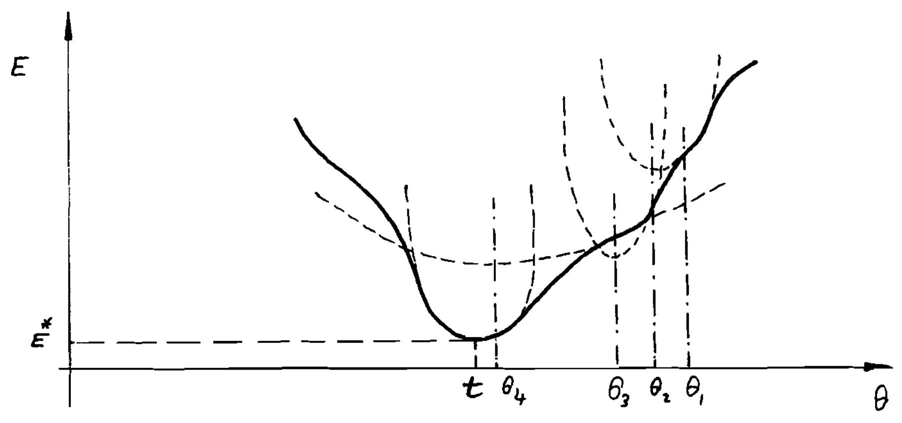


fig. 5.4.1. Illustration of Newton-Raphson convergence in one dimensional space.

Practical limitation is that the method does not converge in the case where the function to be minimized is not concave at some starting (or intermediate) point.

5.5. Gauss-Newton method:

This method is an approximation of the Newton-Raphson method when applied to least squares minimization. It avoids us the second derivatives evaluation. Consider the error equation $\underline{e}(\underline{\theta})$, it can be approximated in a certain neighbourhood of a point $\underline{\theta}_i$ by the following linear relation

$$\underline{L}(\underline{\theta}) = \underline{e}(\underline{\theta}_i) + G_i(\underline{\theta} - \underline{\theta}_i) \qquad (5.5.1.)$$

where

$$G_i = \left. \frac{\partial \underline{e}(\underline{\theta})}{\partial \underline{\theta}^T} \right|_{\underline{\theta} = \underline{\theta}_i}$$

where Taylor's expansion is utilized around $\underline{\theta}_i$ up to the 1st order term. Now, if we employ weighted least squares minimization then

$$E(\underline{\theta}) = \underline{L}^T(\underline{\theta})\, W\, \underline{L}(\underline{\theta})$$

$$= (\underline{e}_i + G_i\, \Delta\underline{\theta}_i)^T\, W\, (\underline{e}_i + G_i\, \Delta\underline{\theta}_i)$$

The minimum is found by equating $\frac{\partial E(\underline{\theta})}{\partial \underline{\theta}}$ to zero

$$2\, G_i^T W (\underline{e}_i + G_i \,\Delta \underline{\theta}_i) = \underline{0}$$

which results in the i^{th} iteration of the Gauss-Newton method

$$\underline{\theta}_{i+1} = \underline{\theta}_i - (G_i^T W G_i)^{-1} G_i^T W \underline{e}_i \qquad (5.5.2.)$$

but $\left.\frac{\partial E(\underline{\theta})}{\partial \underline{\theta}}\right|_{\underline{\theta}_i} = \underline{g}_i = 2\, G_i^T W \underline{e}_i$

and the i^{th} iteration can be written as

$$\underline{\theta}_{i+1} = \underline{\theta}_i - (2\, G_i^T W G_i)^{-1} \underline{g}_i \qquad (5.5.3.)$$

which conforms to the general formula of the gradient method eq. (5.1.1.) with $\lambda_i = 1$ and $R_i = (2G_i^T W G_i)^{-1}$. Note that $(2G_i^T W G_i)$ is an approximation of H_i used in the Newton Raphson method, since

$$H_i = \frac{\partial}{\partial \underline{\theta}\, \partial \underline{\theta}^T} \{ \underline{e}^T(\underline{\theta})\, W\, \underline{e}(\underline{\theta}) \} = 2\, G_i^T W G_i + 2 \sum_{k=1}^{n} \left(\sum_{r=1}^{n} w_{kr}\, e_r(\underline{\theta}_i) \right) M_k(\theta_i)$$

n is the length of the vector $\underline{e}(\theta)$, $M_k(ij) = \frac{\partial^2 e_k(\underline{\theta})}{\partial \theta_i \partial \theta_j}$ and w_{kr} is the element W(kr) of the matrix W.

where the second term is dropped in Gauss-Newton method. It is obvious that this approximation is valid in the neighbourhood of the minimum since $\underline{e}_i$ is small. It is interesting to note that if the error equation $\underline{e}$ ($\underline{\theta}$) is linear, then this method is equivalent to the Newton-Raphson mehtod; they both converge in a single iteration. Since the matrix $(G_i^T W G_i)$ is a positive definite (or at least semi definite) Gauss-Newton method is more applicable for the minimization of general functions.

5.6. Marquardt method:

In order to insure the positive definitness of the matrix R_i, the following modification is suggested by Marquardt (1963)

$$\underline{\theta}_{i+1} = \underline{\theta}_i - (A_i + \mu_i I)^{-1} \underline{g}_i \quad , \quad I \equiv \text{identity matrix} \qquad (5.6.1.)$$

which conforms to the general formula of the gradient methods eq. (5.1.1.) with $\lambda_i = 1$ and $R_i = (A_i + \mu_i I)^{-1}$, where μ_i is chosen such that R_i is positive definte this will guarantee acceptable iteration steps. We shall see that the Marquardt iteration is derived from the minimization of the quadratic approximation of $E(\underline{\theta})$ around $\underline{\theta}_i$, under the constraint

$$\Delta \underline{\theta}_i^T \Delta \underline{\theta}_i = C \quad , \text{ where } \Delta \underline{\theta}_i = \underline{\theta} - \underline{\theta}_i$$

consider the quadratic approximation of $E(\underline{\theta})$ around $\underline{\theta}_i$ is given by

$$Q(\underline{\theta}) = E(\underline{\theta}_i) + \underline{g}_i^T \cdot \Delta \underline{\theta}_i + \frac{1}{2} \Delta \underline{\theta}_i^T A_i \Delta \underline{\theta}_i$$

where A_i may be, the exact Hessian H_i used in Newton-Raphson method, or the approximate Hessain ($2 G_i^T W G_i$) used in Gauss-Newton method. If we join the constraint with a lagrange multiplier ($\frac{1}{2}\mu_i$) to the quadratic $Q(\underline{\theta})$

$$F(\underline{\theta}) = E(\underline{\theta}_i) + \underline{g}_i^T \Delta \underline{\theta}_i + \frac{1}{2} \Delta \underline{\theta}_i^T A_i \Delta \underline{\theta}_i + \frac{1}{2} \mu_i (\Delta \underline{\theta}_i^T \Delta \underline{\theta}_i - C)$$

Minimum is found by differentiating w.r.t. $\Delta \underline{\theta}_i$ and equating to zero

$$\underline{g}_i + A_i \Delta \underline{\theta}_i + \mu_i I \Delta \underline{\theta}_i = \underline{0}$$

which leads to the Marquardt iteration given above.
In other words the step $\Delta \underline{\theta}_i$ finds the minimum (of the quadratic approximation of $E(\underline{\theta})$ around $\underline{\theta}_i$) on the hypersphere given by $\Delta \underline{\theta}_i^T \Delta \underline{\theta}_i = C$.
The size of the hypersphere depends on μ_i where

if $\mu = 0 \longrightarrow R_i = A_i^{-1}$ (Newton method)
if $\mu \rightarrow \infty \longrightarrow R_i = \mu_i^{-1} I$ (steepest descent method)

Fig. 5.6.1. will help as an illustration for the Marquardt iteration.

μ_i can be chosen in each iteration in such a way to obtain an acceptable step (i.e. $E(\underline{\theta}_{i+1}) < E(\underline{\theta}_i)$. Many algorithms were proposed for the choice of μ_i within the minimization method; they are available in many sources and will not be discussed here. Marquardt method has shown to be very successful in practice. Disadvantages are the possible requirement for the second derivations in the matrix A_i and its inversion.

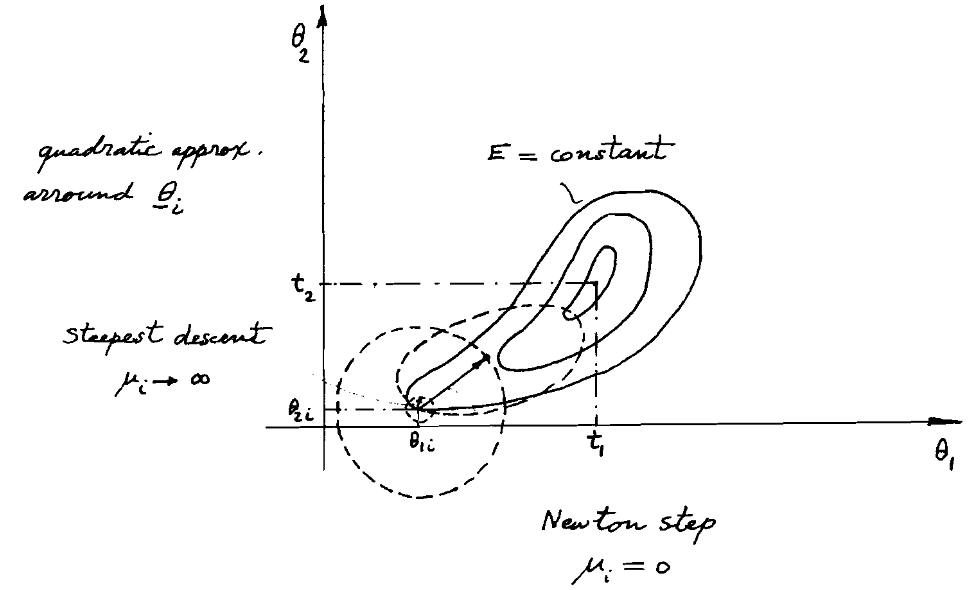


fig. 5.6.1. Illustration of Marquardt iteration in the two-dimensional case.

5.7. Variable Metric methods:

These are quadratically convergent methods where the evaluation of the second derivatives matrix and the determination of its inverse are not meeded; instead, more steps (usually n, the number of variable parameters) are required to reach the minimum of a quadratic function rather than a single step if the inverse of the Hessian of the quadratic function is known. These methods follow the general formula of the gradient methods repeated here for convenience

$$\underline{\theta}_{i+1} = \underline{\theta}_i - \lambda_i R_i \underline{g}_i$$

here λ_i is chosen such that $E(\underline{\theta})$ is minimum along the search direction $(- R_i \underline{g}_i)$. The matrix R_i is updated after each iteration in such a way that it is always positive definite and it becomes the inverse of the Hessian of the quadratic function after n iterations. The term "Variable metric" was introduced by Davidon (1959) as to indicate the automatic updating of the matrix R_i. This updating process is based on the construction of n mutually conjugate directions which can be used to form an expression for the inverse of the Hessian at the minimum; furthermore, a linear conbination of these directions form the step to the minimum of the quadratic function as will be seen in the following.

Consider the quadratic approximation of $E(\underline{\theta})$ around the point $\underline{\theta}_i$

$$Q(\underline{\theta}) = E(\underline{\theta}_i) + \underline{g}_i^T \Delta\underline{\theta}_i + \frac{1}{2} \Delta\underline{\theta}_i^T H_i \Delta\underline{\theta}_i \tag{5.7.1.}$$

where $\underline{g}_i = \left.\frac{\partial E}{\partial \underline{\theta}}\right|_{\underline{\theta}_i}$, $H_i = \left.\frac{\partial^2 E}{\partial \underline{\theta}^T \partial \underline{\theta}}\right|_{\underline{\theta}_i}$

is a symmetric positive definite matrix and $\Delta\underline{\theta}_i = \underline{\theta} - \underline{\theta}_i$

If $E(\underline{\theta})$ is quadratic then $E(\underline{\theta}) = Q(\underline{\theta})$.

The directions $\underline{p}_i$ and $\underline{p}_j$ are said to be conjugate w.r.t. H_i if

$$\underline{p}_i^T H_i \underline{p}_j = 0 \qquad \text{for } i \neq j \tag{5.7.2.}$$

If the directions $\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_n$ are mutually conjugate, then they are linearly independent and they span an n-dimensional space. This means that any vector in that space can be written as a linear combination of the $\underline{p}_i$'s, consider the step $\Delta\underline{\theta}_i = \underline{\theta} - \underline{\theta}_i$; it can be written as

$$\Delta\underline{\theta}_i = \sum_{i=1}^{n} \alpha_i \underline{p}_i \tag{5.7.3.}$$

Substitution in $Q(\underline{\theta})$, equation (5.7.1.) we can write

$$Q(\alpha_1, \ldots, \alpha_n) = E(\underline{\theta}_i) + \underline{g}_i^T \left(\sum_{i=1}^{n} \alpha_i \underline{p}_i\right) + \frac{1}{2}\left(\sum_{i=1}^{n} \alpha_i \underline{p}_i\right)^T H_i \left(\sum_{i=1}^{n} \alpha_i \underline{p}_i\right)$$

$$= E(\underline{\theta}_i) + \sum_{i=1}^{n} \left(\alpha_i \underline{g}_i^T \underline{p}_i + \frac{1}{2} \alpha_i^2 \underline{p}_i^T H_i \underline{p}_i\right)$$

using the mutual conjugancy of the $\underline{p}_i$'s. Now Q is a function of the scalars α_i's and the step to the minimum can be written as

$$\Delta\hat{\underline{\theta}}_i = \hat{\underline{\theta}} - \underline{\theta}_i = \sum_{i=1}^{n} \lambda_i \underline{p}_i \tag{5.7.4}$$

where the scalars λ_i's are determined in such a way as to minimize $Q(\alpha_1, \ldots, \alpha_n)$. It is obvious that the minimization w.r.t. each α_i can be done independently by minimizing the term

$$\left(\alpha_i \underline{g}_i^T \underline{p}_i + \frac{1}{2} \alpha_i^2 \underline{p}_i^T H_i \underline{p}_i\right)$$

since the analytical expression for λ_i

$$\lambda_i = -\frac{\underline{g}_i^T \underline{p}_i}{\underline{p}_i^T H_i \underline{p}_i} \tag{5.7.5.}$$

can not be used because of the absence of information about H_i ; alternatively, we use the fact that along the direction $\underline{p}_i$, Q is a function of, only, α_i. Hence linear search for the minimum of $Q(\alpha_i)$ along the direction $\underline{p}_i$ will provide us the value of λ_i. and the minimum of the quadratic $Q(\underline{\theta})$ is located by a sequence of linear search along the mutually conjugate direction $\underline{p}_1, \ldots, \underline{p}_n$. The inverse of the Hessian H_i can be constructed from the $\underline{p}_i$'s. This will be clear if we rewrite the step $\Delta\hat{\underline{\theta}}_i$ as follows by substituting for λ_i's their analytic expressions in equation (5.7.5)

$$\begin{aligned} \Delta \hat{\underline{\theta}}_i &= - \sum_{i=1}^{n} \left(\frac{\underline{g}_i^T \cdot \underline{p}_i}{\underline{p}_i^T H_i \underline{p}_i} \right) \underline{p}_i \\ &= - \left(\sum_{i=1}^{n} \frac{\underline{p}_i \, \underline{p}_i^T}{\underline{p}_i^T H_i \underline{p}_i} \right) \underline{g}_i \end{aligned} \tag{5.7.6.}$$

which if compared with the Newton step to the minimum of a quadratic

$$\Delta \hat{\underline{\theta}}_i = - H_i^{-1} \underline{g}_i$$

we can write immediately the relation

$$H_i^{-1} = \sum_{i=1}^{n} \frac{\underline{p}_i \, \underline{p}_i^T}{\underline{p}_i^T H_i \underline{p}_i} \tag{5.7.7.}$$

In the preceeding analysis we have seen that the construction of n mutually conjugate directions provides the necessary steps to the minimum of a quadratic function of n variables, as well as the determination of the inverse of its Hessian. All procedures which constructs such a set of mutually conjugate search directions are quadratically convergent (minimum is located after n linear searches) when applied to quadratic functions. They are iterative when applied to nonquadratic functions with the advantage of quadratic convergence properties when the minimum is approached; this because any general function can be well approximated with a quadratic in the vicinity of its minimum. This is a very important feature since it insures that the minimum will be located exactly avoiding the ossillatory behaviour of other methods such as the steepest descent.

Many ways have been proposed in the literature for the construction of the mutually conjugate directions. In the remaining part of this section we will mention the most popular and efficient ones; the first is due to Davidon (1959) and reformulated by Fletcher and Powell (1963), the second is the conjugate gradient method due to Hestenes and Stiefel (1952) and was applied to function minimization by Fletcher and Reeves (1964), A modification to the conjugate gradient method will lead to a procedure which posseses the properties of the Davidon-Fletcher-Powell procedure when applied to nonquadratic functions.

5.7.1. The conjugate direction method: (The Davidon-Fletcher-Powell method)

The search direction vector $\underline{p}_i$ is generated form the relation

$$\underline{p}_i = - R_i \, \underline{g}_i \qquad (5.7.8.)$$

where R_i is updated as follows

$$R_{i+1} = R_i + C_i + D_i \qquad (5.7.9.)$$

The matrices C_i and D_i are given by

$$C_i = \frac{\Delta \underline{\theta}_i \; \Delta \underline{\theta}_i^T}{\Delta \underline{\theta}_i^T \cdot \underline{\gamma}_i} \qquad (5.7.10)$$

$$D_i = - \frac{R_i \, \underline{\gamma}_i \, \underline{\gamma}_i^T R_i}{\underline{\gamma}_i^T R_i \, \underline{\gamma}_i} \qquad (5.7.11.)$$

$$\Delta \underline{\theta}_i = - \lambda_i \underline{p}_i = - \lambda_i R_i \, \underline{g}_i \qquad (5.7.12.)$$

$$\underline{\gamma}_i = \underline{g}_{i+1} - \underline{g}_i \qquad (5.7.13.)$$

$$\underline{\gamma}_i = H \, \underline{\theta}_{i+1} - H \, \underline{\theta}_i = H \Delta \underline{\theta}_i = - \lambda_i H \, \underline{p}_i \qquad (5.7.14.)$$

where H is the Hessian of the quadratic function under consideration. The initial matrix $R_1 (R_1 = D_1$, since $C_1 = o)$ can be chosen arbitrarily, provided it is symmetric and positive definite. The unit matrix is usually chosen and the first iteration is taken in the steepest descent direction.

The matrix C_i can be written as follows

$$C_i = \frac{\underline{p}_i \underline{p}_i^T}{\underline{p}_i^T H_i \underline{p}_i} \qquad (5.7.15.)$$

Reference to equation (5.7.6.) will show that the sequence of matrices C_i are successive approximations to the inverse of the Hessian and converges to the correct value after n iterations; hence $C_{n+1} = H^{-1}$. It is interesting to mention that the C_i's has no influence on the generation of the mutual conjugate directions (Sorenson, 1969). The sequence of matrices D_i are orthogonal projection operators; thus, D_{i+1} projects any vector onto a subspace orthogonal to the gradient difference vectors $\underline{\gamma}_1, \ldots \underline{\gamma}_i$ (or equivalently, $H\underline{p}_1, \ldots, H\underline{p}_i$) and spanned by directions $\underline{p}_{i+1}, \ldots \underline{p}_n$. The matrix D_{n+1} projects any vector onto a subspace orthogonal to n linearly independent gradient difference vectors which is not possible; hence, $D_{n+1} = o$.
The last paragraph implies the convergence of R_i to the inverse of the Hessian; i.e. $R_{n+1} = H^{-1}$ when the iterations are applied to a quadratic function. In the meanwhile the conjugate directions add linearly together resulting in the required step to the minimum as we have seen before. When the method is applied to nonquadratic functions, the direction vectors are not mutually conjugate and the matrix R_i is no longer an approximation to the inverse of the Hessian of the minimum. In the vicinity of the minimum the function can be well approximated with a quadratic and the convergence properties of the method starts to manifest themselves. An important property for the method when applied to nonquadratic functions is that the directions $\underline{p}_i$ is always orthogonal to the gradient difference vector $\underline{\gamma}_{i-1}$; this avoids the search directions to be parallel (or almost parallel) thus allowing new directions to be explored.

5.7.2. The conjugate gradient method: (The Fletcher-Reeves method)

First, we present a brief analysis of the method followed by a summary of the computational procedure.

In the following analysis many relations can be proved, only, by induction. We will proof some relations assuming the mutual conjugacy of the generated search directions

$$\underline{p}_i^T H \underline{p}_j = 0 \qquad , i \neq j \qquad (5.7.16.)$$

where H is the Hessian of the quadratic function under consideration. These relations are used to derive the formulaes for generating the mutually conjugate search directions.
Starting from an initial point $\underline{\theta}_1$, the initial direction is taken in the steepest descent direction, so

$$\underline{p}_1 = - \underline{g}_1 \qquad (5.7.17.)$$

A new point is determined using the well known iteration formula

$$\underline{\theta}_{i+1} = \underline{\theta}_i - \lambda_i \underline{p}_i \qquad (5.7.18.)$$

where λ_i is determined by linear search for the minimum along the direction $\underline{p}_i$, so that the following relation holds

$$\underline{g}_{i+1}^T \underline{p}_i = 0 \qquad (5.7.19.)$$

We will proof the following orthogonality relation

$$\underline{g}_{i+1}^T \underline{p}_j = 0 \qquad , j \leqslant i \qquad (5.7.20.)$$

From equation (5.7.18.) we can write

$$\underline{\theta}_{i+1} = \underline{\theta}_{j+1} - \sum_{r=j+1}^{i} \lambda_r \underline{p}_r \qquad , \quad j < i$$

multiplying both sides by the Hessian H, it follows

$$\underline{g}_{i+1} = \underline{g}_{j+1} - \sum_{r=j+1}^{i} \lambda_r H \underline{p}_r \qquad , \quad j < i$$

and therefore, using equations (5.7.16.) and (5.7.19.)

$$\underline{p}_j^T \underline{g}_{i+1} = \underline{p}_j^T \underline{g}_j - \sum_{r=j+1}^{i} \lambda_r \underline{p}_j^T H \underline{p}_r = 0 \qquad , \quad j < i$$

which in combination with equation (5.7.19.) results in equation (5.7.20.) This means that $\underline{g}_{i+1}$ is orthogonal to the subspace spanned by $\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_i$.

At the new point $\underline{\theta}_{i+1}$, a new search direction is generated using the relation

$$\underline{p}_{i+1} = -\underline{g}_{i+1} + \beta_i \underline{p}_i \tag{5.7.21.}$$

It is clear that $\underline{p}_i$ is a linear combination of the gradient vectors $\underline{g}_i, \underline{g}_{i-1}, \ldots, \underline{g}_1$; this implies that $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_i$ are within the subspace spanned by $\underline{p}_1, \underline{p}_2, \ldots \underline{p}_i$, and hence $\underline{g}_{i+1}$ is orthogonal to all previous gradient vectors $\underline{g}_i, \underline{g}_{i-1}, \ldots, \underline{g}_1$, then the following relation holds

$$\underline{g}_i^T \cdot \underline{g}_j = 0 \qquad , \quad i \neq j \tag{5.7.22.}$$

Now we are in a point to determine β_i such that the mutual conjugacy of the generated direction, expressed by

$$\underline{p}_{i+1}^T H \underline{p}_i = 0 \quad , \left(\text{or equivalently} \quad \underline{\gamma}_i^T \underline{p}_{i+1} = 0\right)$$

is satisfied, this results in

$$(\underline{g}_{i+1}^T - \underline{g}_i^T) \cdot (-\underline{g}_{i+1} + \beta_i \underline{p}_i) = 0$$

$$-\underline{g}_{i+1}^T \underline{g}_{i+1} + \beta_i \underline{g}_{i+1}^T \underline{p}_i + \underline{g}_i^T \underline{g}_{i+1} - \beta_i \underline{g}_i^T \underline{p}_i = 0$$

$$-\underline{g}_{i+1}^T \underline{g}_{i+1} - \beta_i \underline{g}_i^T (-\underline{g}_i + \beta_{i-1} \underline{p}_{i-1}) = 0$$

$$-\underline{g}_{i+1}^T \underline{g}_{i+1} + \beta_i \underline{g}_i^T \underline{g}_i - \beta_i \beta_{i-1} \underline{g}_i^T \underline{p}_{i-1} = 0$$

hence, the scalar should satisfy the relation

$$\beta_i = \frac{\underline{g}_{i+1}^T \underline{g}_{i+1}}{\underline{g}_i^T \cdot \underline{g}_i} \tag{5.7.23.}$$

We can summarize the conjugate gradient algorithm in the following steps,

1. Starting from the point $\underline{\theta}_1$ as an initial guess for the location of the minimum, we take the steepest descent direction as a search direction.
2. Linear search for the minimum is performed in direction $\underline{p}_i$ and $\underline{\theta}_{i+1}$ is the location of the minimum on that direction, $(\underline{\theta}_{i+1} = \underline{\theta}_i - \lambda_i \underline{p}_i)$.

3. New search direction $\underline{p}_{i+1}$ from the point $\underline{\theta}_{i+1}$ is determined following

$$\underline{p}_{i+1} = -\underline{g}_{i+1} + \beta_i \underline{p}_i \quad \text{where} \quad \beta_i = \frac{\underline{g}_{i+1}^T \cdot \underline{g}_{i+1}}{\underline{g}_i^T \cdot \underline{g}_i}$$

4. Repeat from the second step until either i = n+1 (n is the number of unknown variables) or the stopping criterion is satisfied.

This algorithm is guaranteed, apart from rounding errors to locate the minimum of a quadratic function of n variables in at most n steps. When applied to nonquadratic function, the algorithm is iterative rather than n-step algorithm and a test for convergence is required for termination. For this reason the algorithm is modified in such a way that the iterations are performed in cycles each of (n+1) steps. From the last point in the last cycle starting with the steepest descent direction we search for the minimum in (n+1) conjugate directions, where an additonal iteration is performed to compensate for the accumulation of rounding errors in the first n iterations. Cycles are repeated until the termination criterion is satisfied or no reduction in the function is obtained in number of cycles.

Advantages of this method are the simplicity of implementation and the modest storage requirement; where space for only three vectors being required ($\underline{\theta}$, $\underline{g}$ and $\underline{p}$) which makes the method preferable for problems with large number of variables. A disadvantage when it is applied to nonquadratic functions is the possibility of having successive search directions being almost parallel resulting in a very slow convergence; this is this avoided by starting a new cycle of iterations from the steepest descent direction after every (n+1) steps. In the neighbourhood of the minimum the quadratic convergence properties are valid and the minimum is located with a high accuracy.

5.7.3. The modified conjugate gradient method:

This method combines the computational simplicity of the conjugate gradient method and the basic property of the Davidon-Fletcher-Powell method when applied to nonquadratic functions; namely, the orthogonality of the search direction $\underline{p}_{i+1}$ to the gradient difference vector $\underline{\gamma}_i$.

It is theoretically identical to the conjugate gradient method when applied to quadratic functions.
Starting from an initial point $\underline{\theta}_1$, the initial direction is taken in the steepest descent direction, so

$$\underline{p}_1 = -\underline{g}_1$$

A new point is determined from

$$\underline{\theta}_{i+1} = \underline{\theta}_i - \lambda_i \underline{p}_i$$

where λ_i is determined by linear search along the direction $\underline{p}_i$ such that

$$\underline{g}_{i+1}^T \underline{p}_i = 0 \tag{5.7.24.}$$

Note that for nonquadratic functions, the relation

$$\underline{g}_{i+1}^T \underline{p}_j = 0 \qquad , \; j < i$$

does not hold, since the mutual conjugacy property of the generated directions is not preserved. Consequently, the relation

$$\underline{g}_i^T \cdot \underline{g}_j = 0 \qquad , \; i \neq j$$

does not hold as well for nonquadratic functions.
At the new point $\underline{\theta}_{i+1}$, a new direction is generated using the relation

$$\underline{p}_{i+1} = -\underline{g}_{i+1} + \beta_i \underline{p}_i \tag{5.7.25.}$$

where β_i is to be determined such that the following orthogonality property always holds

$$\underline{\gamma}_i^T \underline{p}_{i+1} = 0 \tag{5.7.26.}$$

$$i.e. \quad -\underline{\gamma}_i^T \underline{g}_{i+1} + \beta_i \underline{\gamma}_i^T \underline{p}_i = 0$$

$$\beta_i = \frac{\underline{\gamma}_i^T \cdot \underline{g}_{i+1}}{\underline{\gamma}_i^T \cdot \underline{p}_i} \tag{5.7.27.}$$

we can also write, using equation (5.7.24.) and (5.7.25.)

$$\beta_i = \frac{\underline{\gamma}_i^T \cdot \underline{g}_{i+1}}{\underline{g}_i^T \cdot \underline{g}_i}$$

The computational algorithm is the same as that of the conjugate gradient method except that equation (5.7.27.) is used for the determination of β_i.

5.8. Termination:

If one of the quadratically convergent methods is applied to quadratic function, then a possible criterion for stopping is to allow certain number of iterations (possibly depending on the number of variables) within which the method is known to be convergent. But if the function under consideration is nonquadratic and/or if the method is not quadratically convergent then we need to design a stopping criterion for termination. A natural choice is the vanishing of the gradient, but this is unlikely to be realized in practice because of rounding errors. A practical choice for the stopping criterion which is usually adopted in literature is the following test

$$E(\underline{\theta}_i) - E(\underline{\theta}_{i+\ell}) < \epsilon_E$$

and/or $$|\underline{\theta}_{i+\ell} - \underline{\theta}_i| < \epsilon_\theta$$

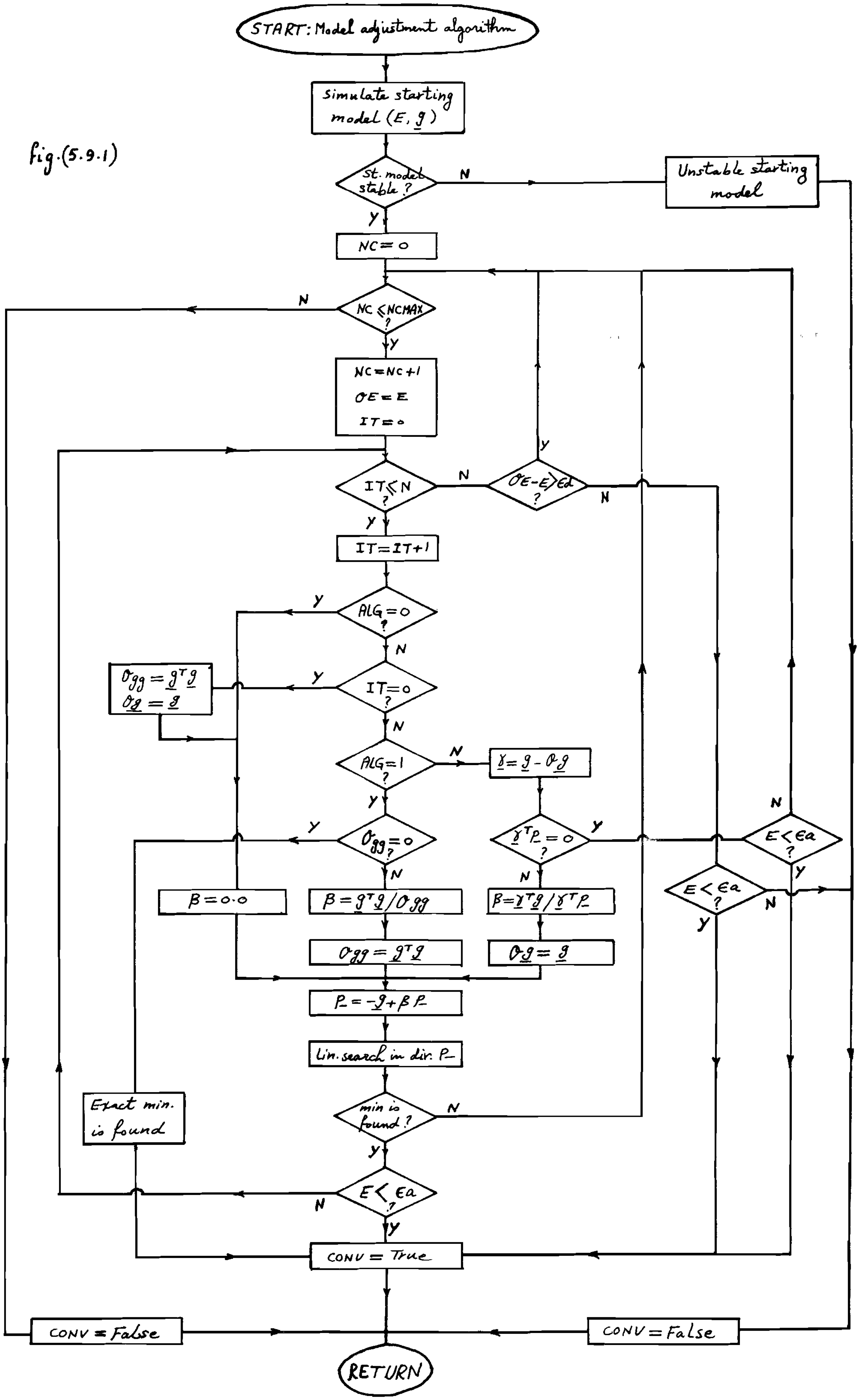
for prescribed small values of ϵ_E and ϵ_θ

i.e. the reduction in the function and/or the change in all the variables over a number of iterations (ℓ) is smaller than prescribed small values ϵ_E and ϵ_θ respectively. This is a suitable criterion since. The conjugate direction methods are particularly liable to converge slowly over a number of iterations, then to make sudden progress.

5.9. Conclusions:

For minimization of nonquadratic functions, it seems very reasonable to start with the steepest descent method until the minimum is approached or the progress becomes slow, then we switch to one of the quadratically convergent methods. The conjugate gradient is attractive method due to its simplicity and storage requirements. Fig. 5.9.1. shows a flow chart for a minimization algorithm in which one of three methods (steepest descent, conjugate gradient and modified conjugate gradient) can be chosen according to program parameter, this parameter can be changed in the program to switch from one method to another.

fig.(5.9.1)


START: Model adjustment algorithm
Simulate starting model (E, g)
St. model stable ?
Unstable starting model
NC = 0
NC ≤ NCMAX ?
NC = NC+1
OE = E
IT = 0
IT ≤ N ?
OE-E > Ed ?
IT = IT+1
ALG = 0 ?
IT = 0 ?
Ogg = gTg
Og = g
ALG = 1 ?
γ = g - Og
Ogg = 0 ?
γTP = 0 ?
E < Ea ?
β = 0.0
β = gTg/Ogg
β = γTg/γTP
Ogg = gTg
Og = g
P = -g + βP
Lin. search in dir. P
min is found ?
Exact min. is found
E < Ea ?
CONV = True
CONV = False
CONV = False
RETURN
Y
N

6. Results

This chapter is devoted to the presentation and discussion of some results of parameter estimation routines in which we apply some of the ideas and techniques discussed in previous chapters. All routines are implemented on a digital computer* which is used to simulate the process as well as the model and to generate the input sequence. Plotting routines† were implemented and used to draw the error function contour lines in the parameter-plane.

6.1. Discription of simulations:

6.1.1. Input signal:

Since the process and the model are simulated digitaly, the input signal is a sequence of numbers representing a sequence of samples taken from a continuous signal. In chapter 4, we discussed in detail the choice of a suitable input signal with a special emphasis on the case where the continuous signal is a periodic Pseuds Randam Binary Noise (PRBN) generated by a Maximum Length Sequence (MLS) generator. The choice of the number of register elements used in the generator was discussed and found to be constrained practically by the length of the signal period. The choice of the sampling rate was studied in detail and we found a practical sampling frequency to be an integer multiple of the continuous signal frequency (around the clock frequency of the generator).
Most of the results were obtained using as an input a PRBN sequence taken from a MLS generator with 7 register elements (i.e. $N = 2^7 - 1 = 127$ signal states) and sampling frequency equals to the clock frequency of the MLS generator (i.e. one sample for each state). This sequence of samples ($u(k)$, $k = 1, \ldots N$) in addition to a combination of step and alternating sequences are applied periodically till stationarity of the model is achieved, the corresponding stationary output sequence is taken for further processing.

*) PDP-11 minicomputer was used for simulations and routines implementation

†) Drawings were realized using a Tektronix display device connected to the PDP-11 minicomputer.

6.1.2. Process:

The adopted process is a second order linear process given by its discrete-time transfer function

$$H_p(z) = \frac{a_6 z^2 + a_5 z + a_4}{a_3 z^2 + a_2 z + a_1}$$

The process is simulated on a digital computer and is subjected to a suitable chosen periodic input sequence (u(k)) which, together with the stationary process output sequence (y(k)) form the data needed for the estimation procedure.

6.1.3. Model and sensitivity model:

The model is assumed to be of the same order as the process and is given by its discrete time transfer function

$$H_M(z) = \frac{\alpha_6 z^2 + \alpha_5 z + \alpha_4}{\alpha_3 z^2 + \alpha_2 z + \alpha_1}$$

The model is simulated on a digital computer and is subjected to the same periodic input sequence as the process. The stationary model output sequence (w(k)) together with the stationary process output sequence (y(k)) form the error sequence (e(k)).
The error sensitivity w.r.t. the parameters α_6, α_5 and α_4 $\left(\frac{\partial e(k)}{\partial \alpha_6}, \frac{\partial e(k)}{\partial \alpha_5}, \frac{\partial e(k)}{\partial \alpha_4}\right)$ can be explicitly derived from the model. The sensitivity w.r.t. the parameters α_3, α_2 and α_1 $\left(\frac{\partial e(k)}{\partial \alpha_3}, \frac{\partial e(k)}{\partial \alpha_2}, \frac{\partial e(k)}{\partial \alpha_1}\right)$ can be derived from a sensitivity model (as discussed in section 3.2.1.).
The sensitivity model is simulated simultaneously with the model and is subjected to the model output sequence (w(k)). The error criterion values as well as its sensitivity w.r.t. the parameters are used in the proposed model adjustment technique.

6.1.4. Error criterion function:

We adopt a least squares error criterion given by the sum of squared error sequence in one period of N samples

$$E = \sum_{k=1}^{N} e^2(k)$$

where the error sequence (e(k), k=1, ..., N) is taken after model stationarity is attained. The error sample e(k) is given by

$$e(k) = y(k) - w(k)$$

Utilizing z-transformation we can write

$$e(z) = y(z) - w(z)$$

$$= y(z) - \left(\frac{\alpha_6 z^2 + \alpha_5 z + \alpha_4}{\alpha_3 z^2 + \alpha_2 z + \alpha_1}\right) \cdot u(z)$$

It is obvious that e(z) and consequently e(k) are linear in the parameters α_6, α_5 and α_4 and they are nonlinear in the parameters α_3, α_2 and α_1. We conclude that the error criterion E is a quadratic function of the parameters α_6, α_5 and α_4 and it is a nonquadratic function of the parameters α_3, α_2 and α_1. This will make it convenient to test different minimization algorithms on various types of error criterion. This must be clear, since if we fix the parameters α_3, α_2 and α_1 to their true values and let the parameters α_6, α_5 and α_4 be adjustable; this will lead to a purely quadratic error criterion. On the contrary if we fix the parameters α_6, α_5 and α_4 and let the parameters α_3, α_2 and α_1 be adjustable; this will lead to a nonquadratic error criterion. A combined error criterion which is quadratic in some of the parameters and nonquadratic in other parameters results when adjusting one or more of the parameters α_6, α_5 and α_4 together with one or more of the parameters α_3, α_2 and α_1.

6.1.5. Minimization algorithm: (Model adjustment technique)

In general, the error criterion E may be nonquadratic function. Consequently we use implicit model adjustment techniques to approach the minimum of E. Starting in a steepest descent direction from an initial point $\underline{\alpha}_1$ in the parameter space, a new point is given by

$$\underline{\alpha}_{i+1} = \underline{\alpha}_i + \lambda_i \underline{p}_i$$

where $\underline{p}_i$ (the search direction is generated according to the adopted minimization method and λ_i is a scalar determined such that $E(\underline{\alpha}_{i+1})$ is minimum along the search direction $\underline{p}_i$.

We adopt the conjugate gradient method for its attractive properties such as simplicity, quadratic convergence and modest storage requirements. The modified conjugate gradient and the steepes descent methods were also used for reasons of comparison. They all use the following relations to generate the search directions

$$\underline{p}_1 = -\underline{g}_1$$

$$\underline{p}_{i+1} = -\underline{g}_{i+1} + \beta_i \underline{p}_i$$

where $$\underline{g}_{i+1} = \frac{\partial E}{\partial \underline{\alpha}}\Big|_{\underline{\alpha}_{i+1}}$$

and β_i is different for each algorithm

Steepest descent: $\beta_i = 0$

Conjugate gradient: $\beta_i = \underline{g}_{i+1}^T \cdot \underline{g}_{i+1} / \underline{g}_i^T \cdot \underline{g}_i$

Modified conjugate gradient: $\beta_i = \underline{g}_{i+1}^T \cdot \underline{\gamma}_i / \underline{p}_i^T \cdot \underline{\gamma}_i$

where

$$\underline{\gamma}_i = \underline{g}_{i+1} - \underline{g}_i$$

In the following we will demonstrate the behaviour of the mentioned algorithm when applied to various types of error criterion function for the estimation of the parameters of different processes.

6.2. Result classification

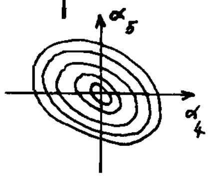
Quadratic error criterion

process: $H_{p_1}(z) = \frac{0.07601z + 0.05447}{z^2}$

input: PRBN sequence (5 reg. elements) + alternating sequence (31 samples per period)

adjustable parameters:

- α_4 and α_5
  - algorithm*: 0, 1 and 2
  - plott:



optimization path in the $\alpha_4 - \alpha_5$ plane

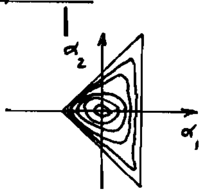
Nonqudratic error criterion

process: $H_{p_1}(z) = \frac{0.07601z + 0.05447}{z^2}$

input: PRBN sequence (7 reg. elements) + alternating sequence (127 samples per period)

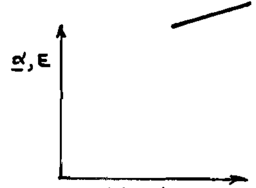
adjustable parameters:

- α_1 and α_2
  - algorithm: 0 and 2
  - plott:

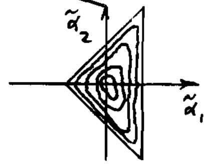


optimization path in the $\alpha_1 - \alpha_2$ plane

- all parameters
  - algorithm: 0 and 1
  - plott:



iteration no.



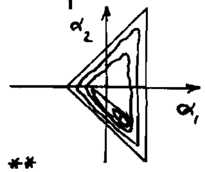
normalized projection** of the optimiz. path in the $\tilde{\alpha}_1 - \tilde{\alpha}_2$ plane

process: $H_{p_2}(z) = \frac{0.07601z + 0.05447}{z^2 - 1.237z + 0.3679}$

input: PRBN sequence (7 reg. elements) + alternating sequence (127 samples per period)

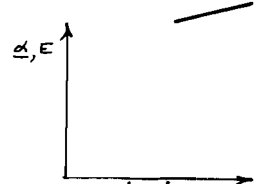
adjustable parameters:

- α_1 and α_2
  - algorithm: 0 and 2
  - plott:

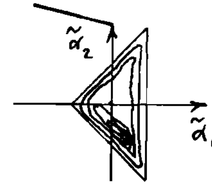


optimiz. path in the $\alpha_1 - \alpha_2$ plane

- all parameters
  - algorithm: 0 and 1
  - plott:



iteration no.



normalized projection of the optimization path in the $\tilde{\alpha}_1 - \tilde{\alpha}_2$ plane

---

*) 0 ≡ steepest descent
1 ≡ Conjugate gradient
2 ≡ modified conjugate gradient

**) Optimization path is drawn after normalizing the parameters w.r.t. α_3 (i.e. $\tilde{\alpha}_i = \frac{\alpha_i}{\alpha_3}$, $i = 1, 2, \ldots, 6$)

EPEST DESCENT METHOD (○)

	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	E
USTABLE PARAMETERS	F	F	F	T	T	F	
E PARAMETERS	0.0000E+00	0.0000E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.0000E+00
RT PARAMETERS	0.0000E+00	0.0000E+00	0.1000E+01	0.2000E+01	0.5000E+01	0.0000E+00	0.373[illegible]E+06
RATION NO.							
1	0.0000E+00	0.0000E+00	0.1000E+01	0.3173E+01	0.3023E+01	0.0000E+00	0.1085E+06
2	0.0000E+00	0.0000E+00	0.1000E+01	0.6205E+00	0.1508E+01	0.0000E+00	0.3157E+05
3	0.0000E+00	0.0000E+00	0.1000E+01	0.9617E+00	0.9335E+00	0.0000E+00	0.9186E+04
4	0.0000E+00	0.0000E+00	0.1000E+01	0.2191E+00	0.4927E+00	0.0000E+00	0.2672E+04
5	0.0000E+00	0.0000E+00	0.1000E+01	0.3184E+00	0.3255E+00	0.0000E+00	0.7774E+03
6	0.0000E+00	0.0000E+00	0.1000E+01	0.1024E+00	0.1972E+00	0.0000E+00	0.2262E+03
7	0.0000E+00	0.0000E+00	0.1000E+01	0.1313E+00	0.1486E+00	0.0000E+00	0.6579E+02
8	0.0000E+00	0.0000E+00	0.1000E+01	0.6840E-01	0.1113E+00	0.0000E+00	0.1914E+02
9	0.0000E+00	0.0000E+00	0.1000E+01	0.7681E-01	0.9712E-01	0.0000E+00	0.5568E+01
10	0.0000E+00	0.0000E+00	0.1000E+01	0.5852E-01	0.8627E-01	0.0000E+00	0.1620E+01
11	0.0000E+00	0.0000E+00	0.1000E+01	0.6097E-01	0.8215E-01	0.0000E+00	0.4712E+00
12	0.0000E+00	0.0000E+00	0.1000E+01	0.5565E-01	0.7899E-01	0.0000E+00	0.1371E+00
13	0.0000E+00	0.0000E+00	0.1000E+01	0.5636E-01	0.7780E-01	0.0000E+00	0.3988E-01
14	0.0000E+00	0.0000E+00	0.1000E+01	0.5481E-01	0.7688E-01	0.0000E+00	0.1160E-01
15	0.0000E+00	0.0000E+00	0.1000E+01	0.5502E-01	0.7653E-01	0.0000E+00	0.3376E-02
16	0.0000E+00	0.0000E+00	0.1000E+01	0.5457E-01	0.7626E-01	0.0000E+00	0.9818E-03
17	0.0000E+00	0.0000E+00	0.1000E+01	0.5463E-01	0.7616E-01	0.0000E+00	0.2856E-03
18	0.0000E+00	0.0000E+00	0.1000E+01	0.5450E-01	0.7608E-01	0.0000E+00	0.8308E-04
19	0.0000E+00	0.0000E+00	0.1000E+01	0.5452E-01	0.7606E-01	0.0000E+00	0.2417E-04
20	0.0000E+00	0.0000E+00	0.1000E+01	0.5448E-01	0.7603E-01	0.0000E+00	0.7029E-05
21	0.0000E+00	0.0000E+00	0.1000E+01	0.5448E-01	0.7602E-01	0.0000E+00	0.2046E-05

IMATED PARAMETERS ARE:
1)= 0.0000000000E+00
2)= 0.0000000000E+00
3)= 0.1000000000E+01
4)= 0.5448353863E-01
5)= 0.7602280378E-01
6)= 0.0000000000E+00
P -- PARAMETER ESTIMATION

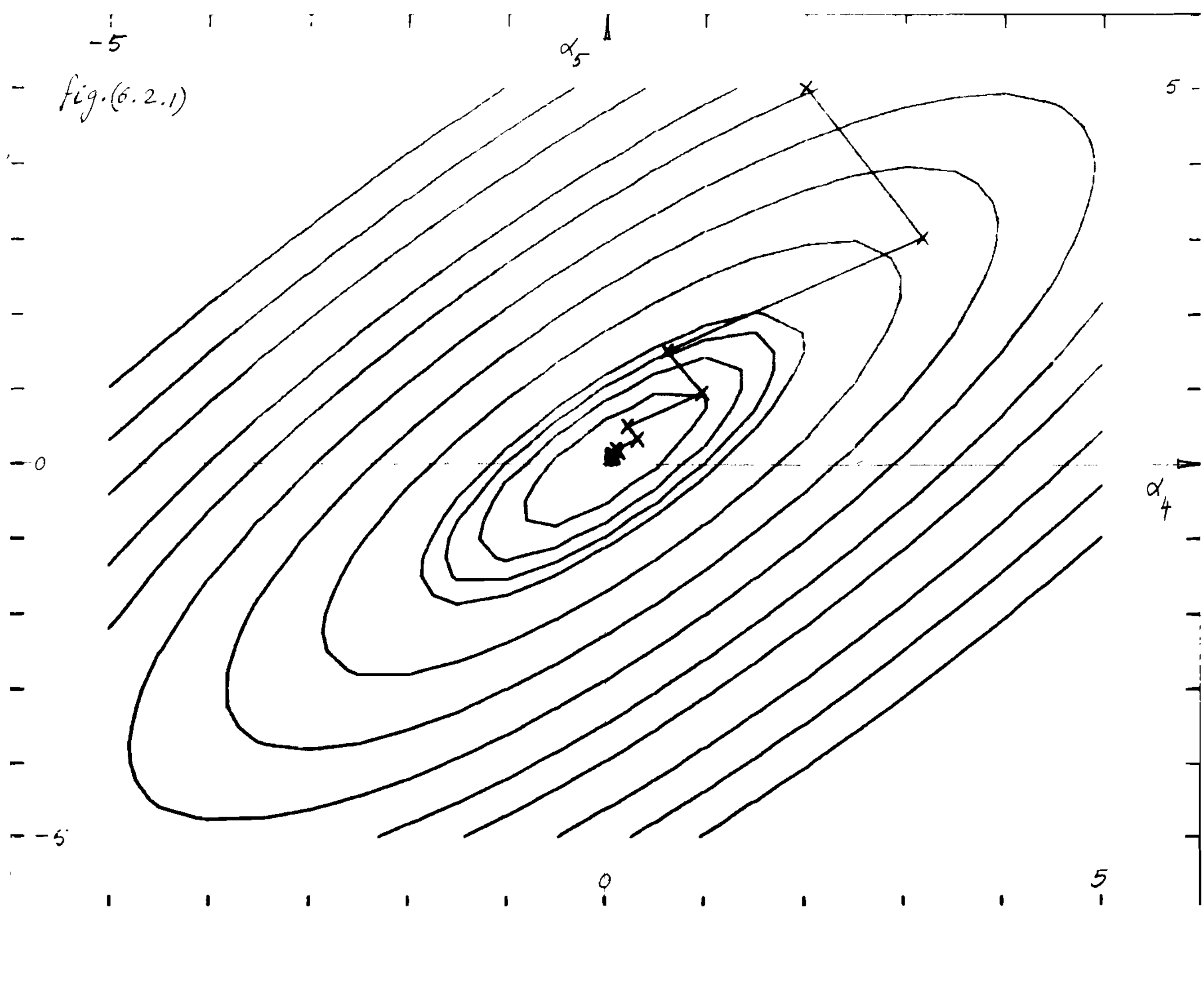
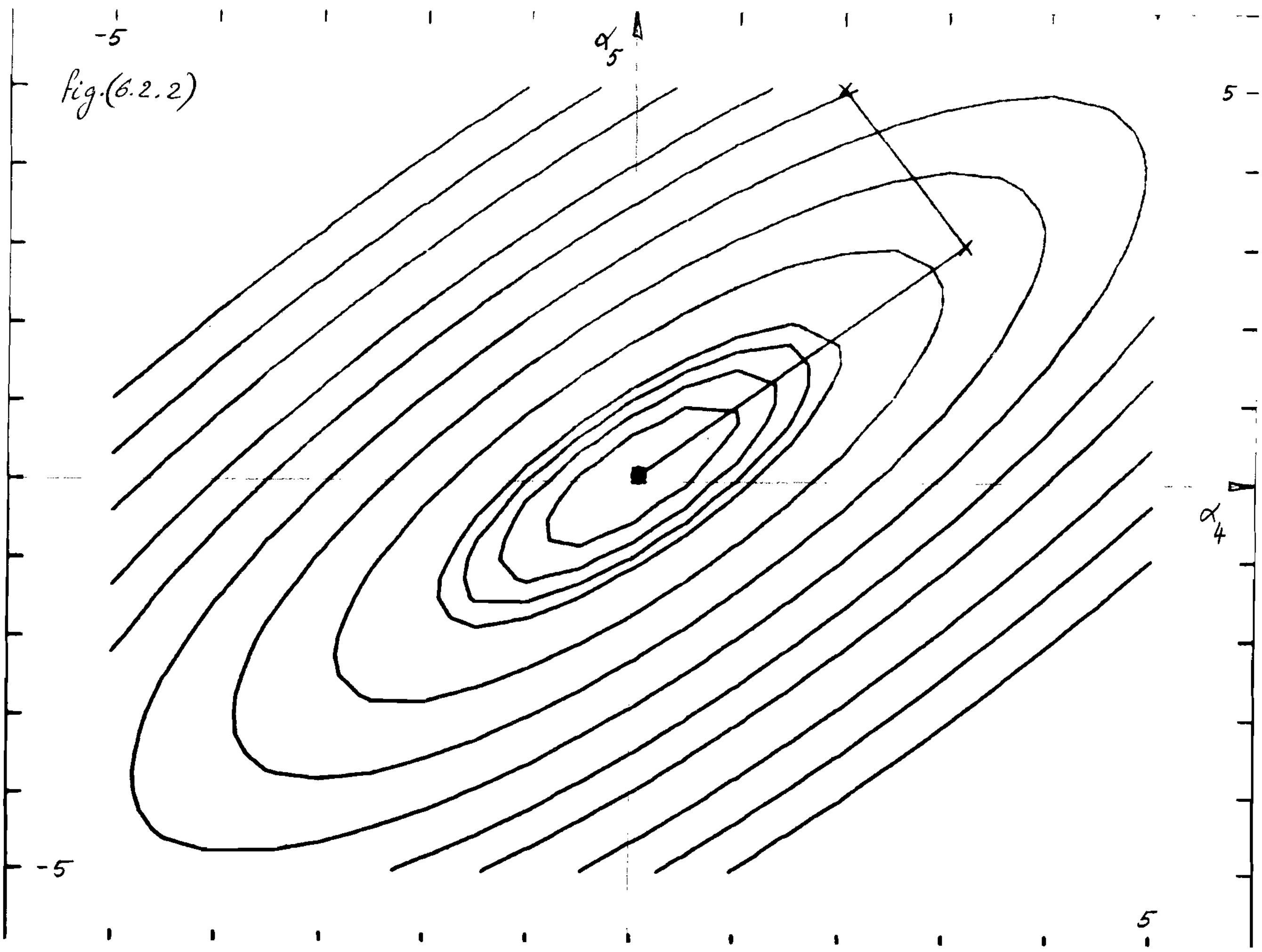


fig.(6.2.1)

JUGATE GRADIENT METHOD (1)

	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	E
JUSTABLE PARAMETERS	F	F	F	T	T	F	
JE PARAMETERS	0.0000E+00	0.0000E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.0000E+00
ART PARAMETERS	0.0000E+00	0.0000E+00	0.1000E+01	0.2000E+01	0.5000E+01	0.0000E+00	0.3731E+06
ERATION NO.							
1	0.0000E+00	0.0000E+00	0.1000E+01	0.3173E+01	0.3023E+01	0.0000E+00	0.1085E+06
2	0.0000E+00	0.0000E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.3114E-06

TIMATED PARAMETERS ARE:
1)= 0.0000000000E+00
2)= 0.0000000000E+00
3)= 0.1000000000E+01
4)= 0.5447125435E-01
5)= 0.7600784302E-01
6)= 0.0000000000E+00
OP -- PARAMETER ESTIMATION



MODIFIED CONJUGATE GRADIENT METHOD (2)

	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	E
ADJUSTABLE PARAMETERS	F	F	F	T	T	F	
TRUE PARAMETERS	0.0000E+00	0.0000E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.0000E+00
START PARAMETERS	0.0000E+00	0.0000E+00	0.1000E+01	0.2000E+01	0.5000E+01	0.0000E+00	0.3731E+06
ITERATION NO.							
1	0.0000E+00	0.0000E+00	0.1000E+01	0.3173E+01	0.3023E+01	0.0000E+00	0.1085E+06
2	0.0000E+00	0.0000E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2805E-08

ESTIMATED PARAMETERS ARE:
P(1)= 0.0000000000E+00
P(2)= 0.0000000000E+00
P(3)= 0.1000000000E+01
P(4)= 0.5447006226E-01
P(5)= 0.7600975037E-01
P(6)= 0.0000000000E+00
STOP -- PARAMETER ESTIMATION

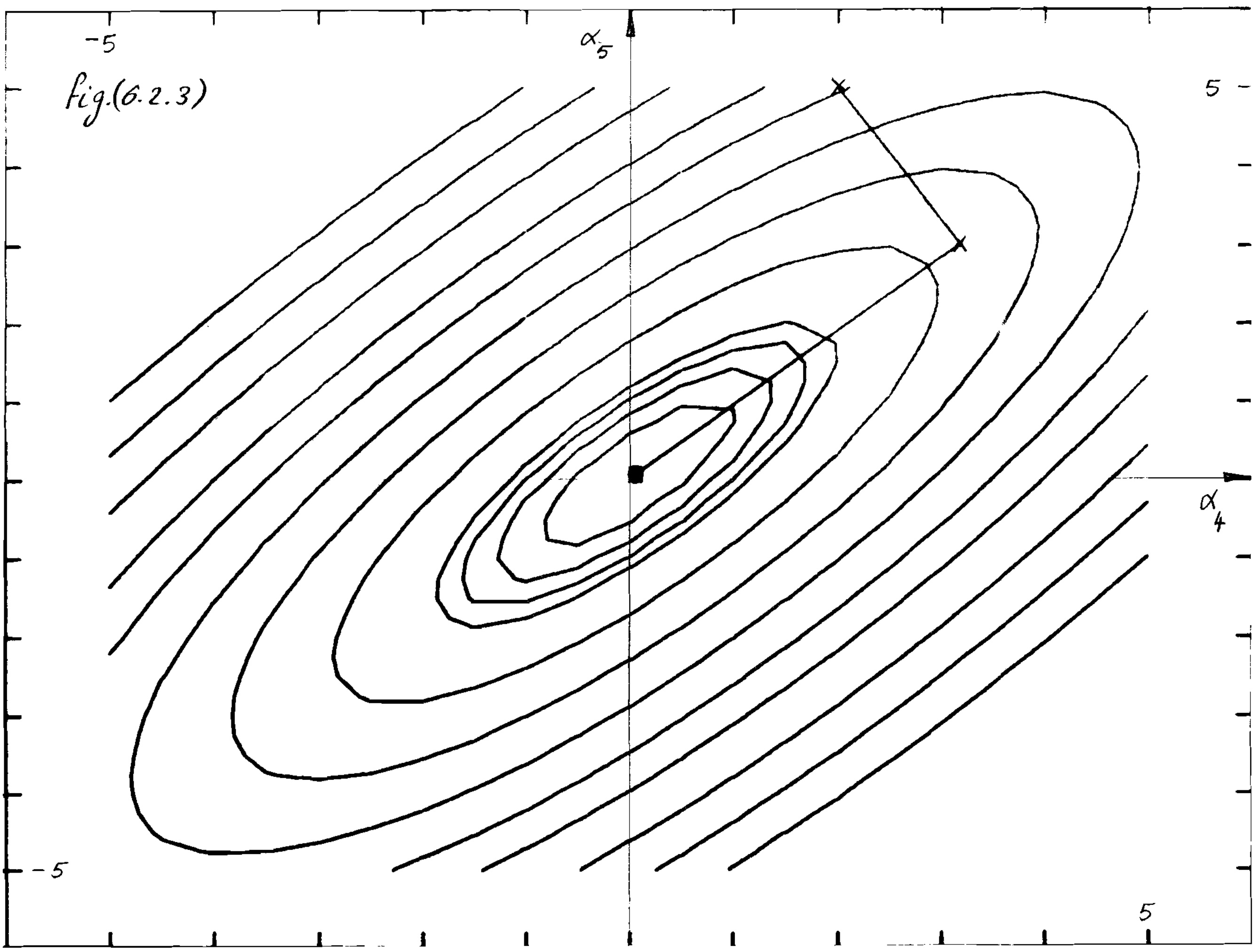


fig.(6.2.3)

EPEST DESCENT METHOD (0)

	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	E
JUSTABLE PARAMETERS	T	T	F	F	F	F	
JE PARAMETERS	0.0000E+00	0.0000E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.0000E+00
ART PARAMETERS	0.9000E+00	0.1800E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.4459E+04
ERATION NO.							
1	0.9473E+00	0.1748E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2207E+04
2	0.8678E+00	0.1696E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1642E+04
3	0.9202E+00	0.1612E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1049E+04
4	0.8793E+00	0.1617E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.8760E+03
5	0.8912E+00	0.1298E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.4805E+03
6	0.7042E+00	0.1254E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2756E+03
7	0.7187E+00	0.3905E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1393E+03
8	0.7192E-01	0.3726E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2563E+02
9	0.1419E+00	-0.1895E-01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.3604E+01
10	0.1885E-01	-0.3808E-01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2663E+00
11	0.1200E-01	-0.1759E-02	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2993E-01
12	0.1644E-02	-0.3628E-02	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2463E-02
13	0.1063E-02	-0.1463E-03	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2088E-03
14	0.1364E-03	-0.3019E-03	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1709E-04
15	0.8625E-04	-0.1200E-04	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1402E-05
16	0.1106E-04	-0.2473E-04	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1144E-06

TIMATED PARAMETERS ARE:
1)= 0.1104897274E-04
2)= -0.2472803409E-04
3)= 0.1000000000E+01
4)= 0.5446999893E-01
5)= 0.7601000369E-01
6)= 0.0000000000E+00
OP — PARAMETER ESTIMATION

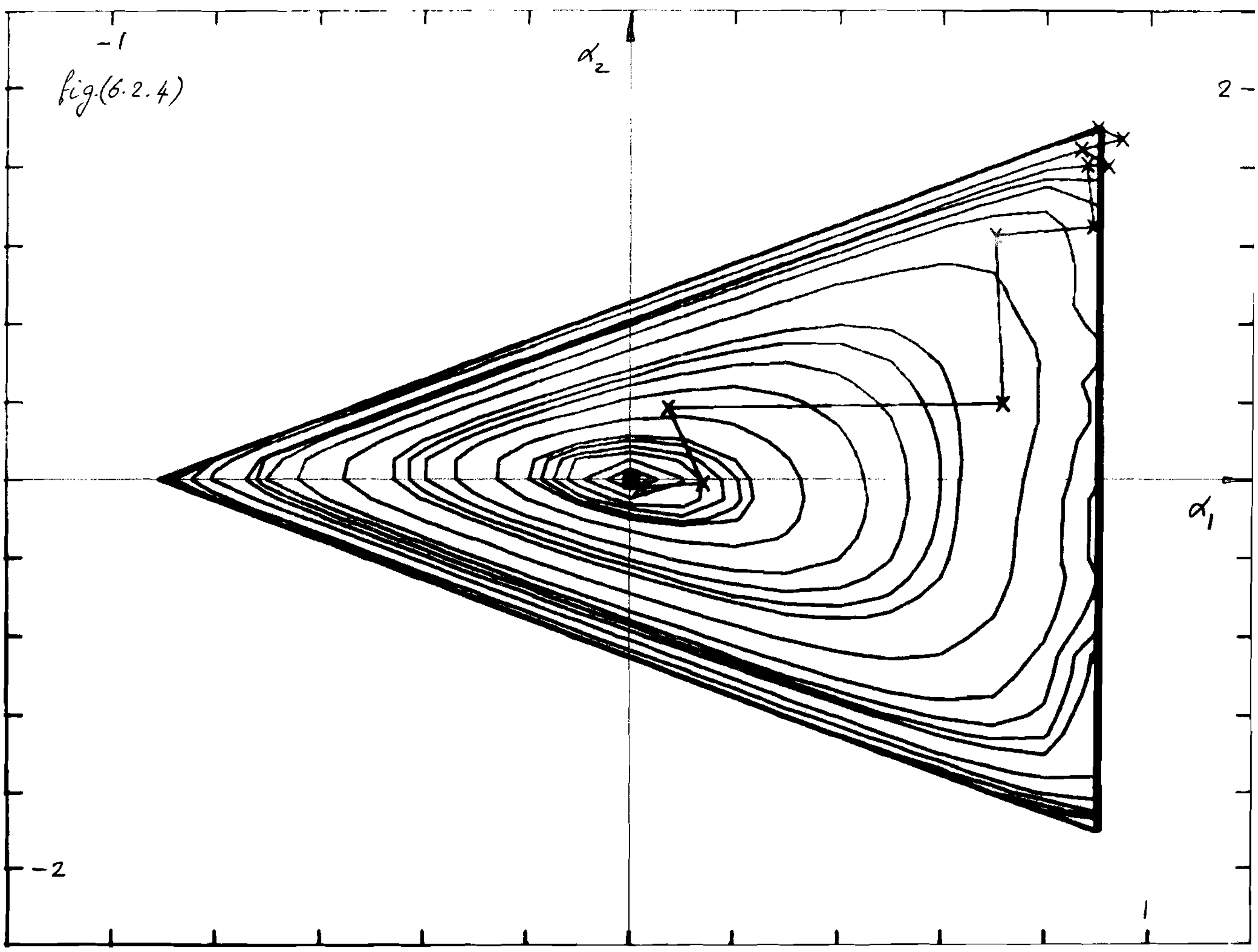
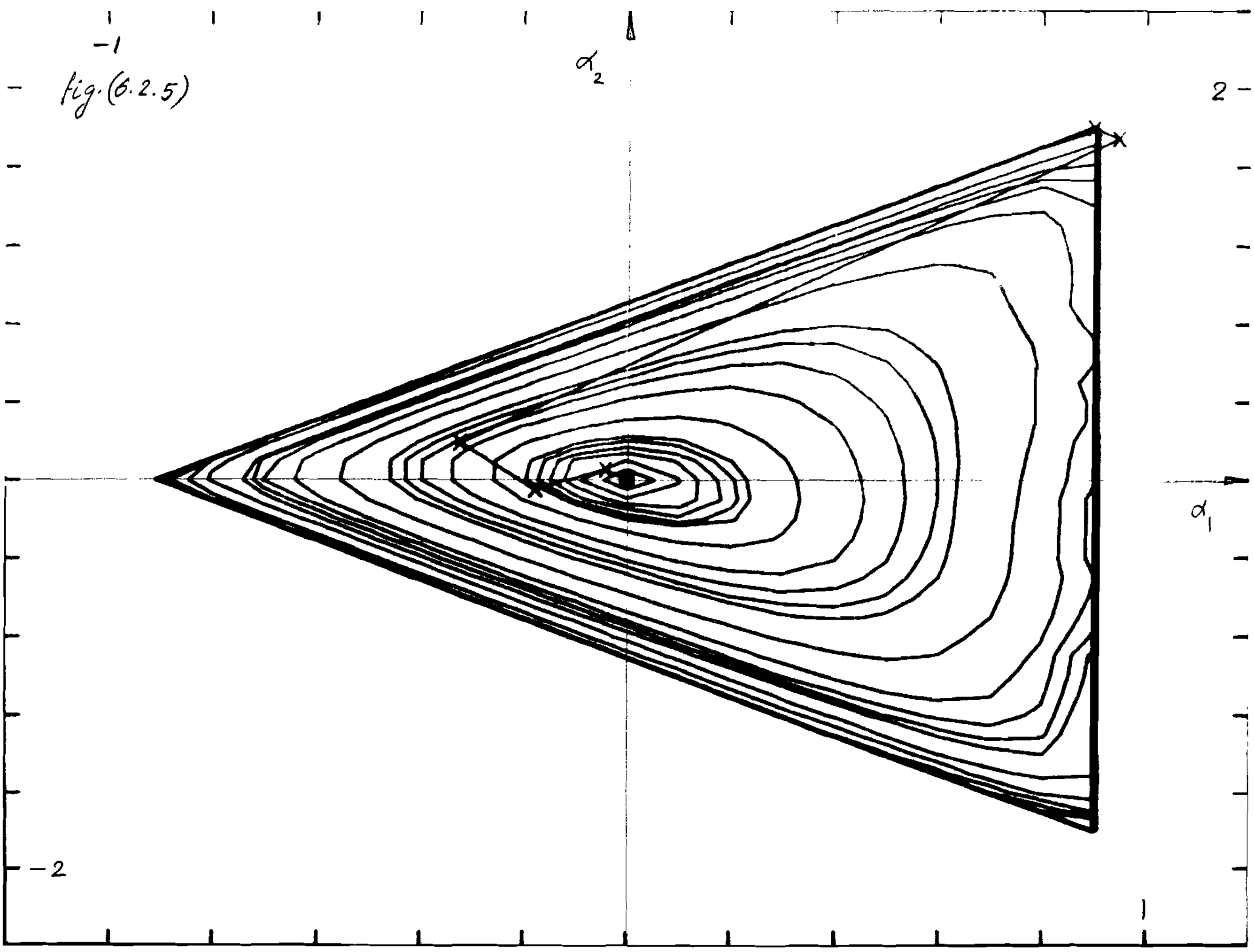


fig.(6.2.4)

DIFIED CONJUGATE GRADIENT METHOD (2)

	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	E
JUSTABLE PARAMETERS	T	T	F	F	F	F	
UE PARAMETERS	0.0000E+00	0.0000E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.0000E+00
ART PARAMETERS	0.9000E+00	0.1800E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.4459E+04
ERATION NO.							
1	0.9473E+00	0.1748E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2207E+04
2	-0.3209E+00	0.1914E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.5122E+02
3	-0.1761E+00	-0.4289E-01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.9921E+01
4	-0.4170E-01	0.5250E-01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.7075E+00
5	0.2951E-02	0.2345E-02	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.3680E-02
6	0.5029E-04	-0.5574E-04	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.7946E-06

TIMATED PARAMETERS ARE:
1)= 0.5029209206E-04
2)= -0.5573867747E-04
3)= 0.1000000000E+01
4)= 0.5446999893E-01
5)= 0.7601000369E-01
6)= 0.0000000000E+00
OP -- PARAMETER ESTIMATION



STEEPEST DESCENT METHOD (0)

	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	E
ADJUSTABLE PARAMETERS	T	T	T	T	T	T	
TRUE PARAMETERS	0.0000E+00	0.0000E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.0000E+00
START PARAMETERS	0.5000E+00	0.5000E+00	0.1000E+01	0.2000E+01	0.3000E+01	0.4000E+01	0.1024E+07
ITERATION NO.							
1	0.7208E+00	-0.6032E+00	0.1000E+01	-0.1445E+00	0.3513E+00	-0.1109E+00	0.6449E+04
2	0.6335E+00	-0.5930E+00	0.1000E+01	0.1057E-01	0.6235E-01	0.9604E-02	0.4596E+02
3	0.6283E+00	-0.5922E+00	0.1000E+01	0.1052E-01	0.6055E-01	0.2521E-02	0.4361E+02
4	0.6208E+00	-0.5913E+00	0.1000E+01	0.1363E-01	0.6037E-01	0.6568E-02	0.4233E+02
5	0.6169E+00	-0.5907E+00	0.1000E+01	0.1091E-01	0.6204E-01	0.3404E-02	0.4114E+02
6	0.6101E+00	-0.5899E+00	0.1000E+01	0.1388E-01	0.6039E-01	0.6241E-02	0.4004E+02
7	0.6065E+00	-0.5895E+00	0.1000E+01	0.1137E-01	0.6271E-01	0.3726E-02	0.3898E+02
8	0.6000E+00	-0.5888E+00	0.1000E+01	0.1414E-01	0.6106E-01	0.6443E-02	0.3796E+02
9	0.5966E+00	-0.5883E+00	0.1000E+01	0.1175E-01	0.6328E-01	0.4033E-02	0.3699E+02
10	0.5904E+00	-0.5877E+00	0.1000E+01	0.1440E-01	0.6170E-01	0.6624E-02	0.3606E+02
11	0.5872E+00	-0.5873E+00	0.1000E+01	0.1210E-01	0.6380E-01	0.4312E-02	0.3516E+02
12	0.5813E+00	-0.5868E+00	0.1000E+01	0.1463E-01	0.6228E-01	0.6790E-02	0.3430E+02
13	0.5782E+00	-0.5864E+00	0.1000E+01	0.1242E-01	0.6420E-01	0.4570E-02	0.3347E+02
14	0.5728E+00	-0.5859E+00	0.1000E+01	0.1483E-01	0.6281E-01	0.6948E-02	0.3267E+02
15	0.5690E+00	-0.5856E+00	0.1000E+01	0.1271E-01	0.6469E-01	0.4808E-02	0.3190E+02
16	0.5643E+00	-0.5851E+00	0.1000E+01	0.1502E-01	0.6329E-01	0.7082E-02	0.3117E+02
17	0.5614E+00	-0.5848E+00	0.1000E+01	0.1290E-01	0.6507E-01	0.5029E-02	0.3045E+02
18	0.5563E+00	-0.5844E+00	0.1000E+01	0.1519E-01	0.6372E-01	0.7210E-02	0.2977E+02
19	0.5536E+00	-0.5842E+00	0.1000E+01	0.1322E-01	0.6542E-01	0.5233E-02	0.2911E+02
20	0.5487E+00	-0.5838E+00	0.1000E+01	0.1534E-01	0.6418E-01	0.7328E-02	0.2847E+02
21	0.5461E+00	-0.5836E+00	0.1000E+01	0.1344E-01	0.6574E-01	0.5483E-02	0.2786E+02
22	0.5414E+00	-0.5832E+00	0.1000E+01	0.1547E-01	0.6449E-01	0.7437E-02	0.2728E+02
23	0.5389E+00	-0.5830E+00	0.1000E+01	0.1364E-01	0.6603E-01	0.5599E-02	0.2669E+02
24	0.5344E+00	-0.5827E+00	0.1000E+01	0.1560E-01	0.6482E-01	0.7537E-02	0.2614E+02
25	0.5308E+00	-0.5825E+00	0.1000E+01	0.1383E-01	0.6629E-01	0.5764E-02	0.2560E+02
26	0.5277E+00	-0.5823E+00	0.1000E+01	0.1571E-01	0.6518E-01	0.7629E-02	0.2508E+02
27	0.5254E+00	-0.5820E+00	0.1000E+01	0.1400E-01	0.6653E-01	0.5917E-02	0.2458E+02
28	0.5213E+00	-0.5818E+00	0.1000E+01	0.1581E-01	0.6540E-01	0.7715E-02	0.2410E+02
29	0.5190E+00	-0.5816E+00	0.1000E+01	0.1415E-01	0.6675E-01	0.6060E-02	0.2363E+02
30	0.5150E+00	-0.5814E+00	0.1000E+01	0.1589E-01	0.6560E-01	0.7794E-02	0.2318E+02
31	0.5129E+00	-0.5812E+00	0.1000E+01	0.1429E-01	0.6696E-01	0.6194E-02	0.2274E+02
32	0.5090E+00	-0.5811E+00	0.1000E+01	0.1597E-01	0.6589E-01	0.7868E-02	0.2231E+02
33	0.5070E+00	-0.5809E+00	0.1000E+01	0.1443E-01	0.6713E-01	0.6320E-02	0.2190E+02
34	0.5033E+00	-0.5807E+00	0.1000E+01	0.1605E-01	0.6610E-01	0.7930E-02	0.2150E+02
35	0.5013E+00	-0.5806E+00	0.1000E+01	0.1454E-01	0.6729E-01	0.6437E-02	0.2112E+02
36	0.4977E+00	-0.5804E+00	0.1000E+01	0.1611E-01	0.6630E-01	0.8000E-02	0.2074E+02
37	0.4958E+00	-0.5803E+00	0.1000E+01	0.1465E-01	0.6744E-01	0.6548E-02	0.2038E+02
38	0.4923E+00	-0.5801E+00	0.1000E+01	0.1617E-01	0.6648E-01	0.8058E-02	0.2003E+02
39	0.4906E+00	-0.5800E+00	0.1000E+01	0.1470E-01	0.6758E-01	0.6651E-02	0.1968E+02
40	0.4872E+00	-0.5799E+00	0.1000E+01	0.1622E-01	0.6664E-01	0.8113E-02	0.1936E+02
41	0.4854E+00	-0.5798E+00	0.1000E+01	0.1485E-01	0.6770E-01	0.6748E-02	0.1903E+02
42	0.4821E+00	-0.5796E+00	0.1000E+01	0.1626E-01	0.6680E-01	0.8164E-02	0.1872E+02

CONVERGENCE NOT ACHIEVED IN 43 ITERATIONS
STOP -- PARAMETER ESTIMATION

See fig. (6.2.6a) and fig. (6.2.6b)

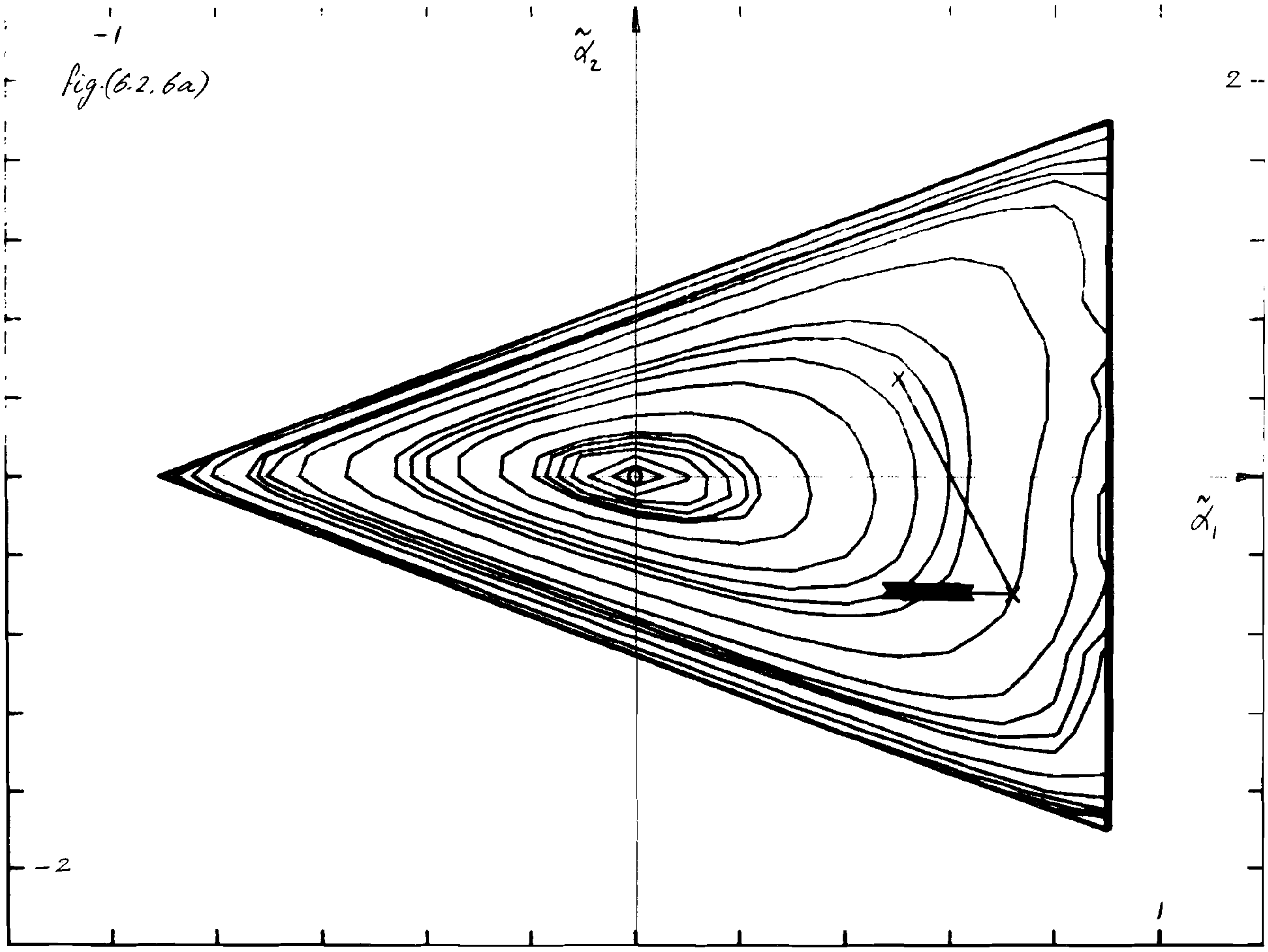


fig.(6.2.6a)

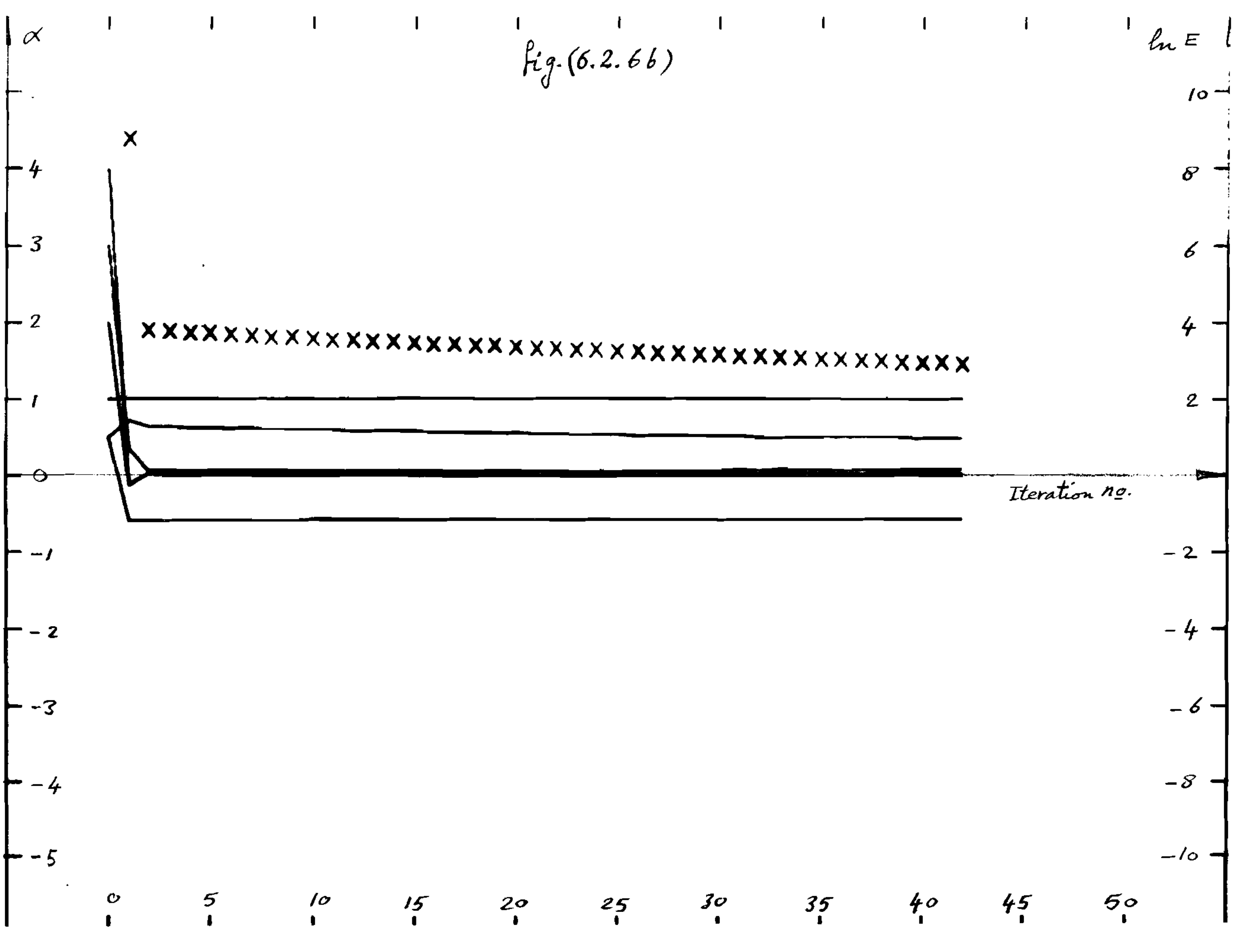


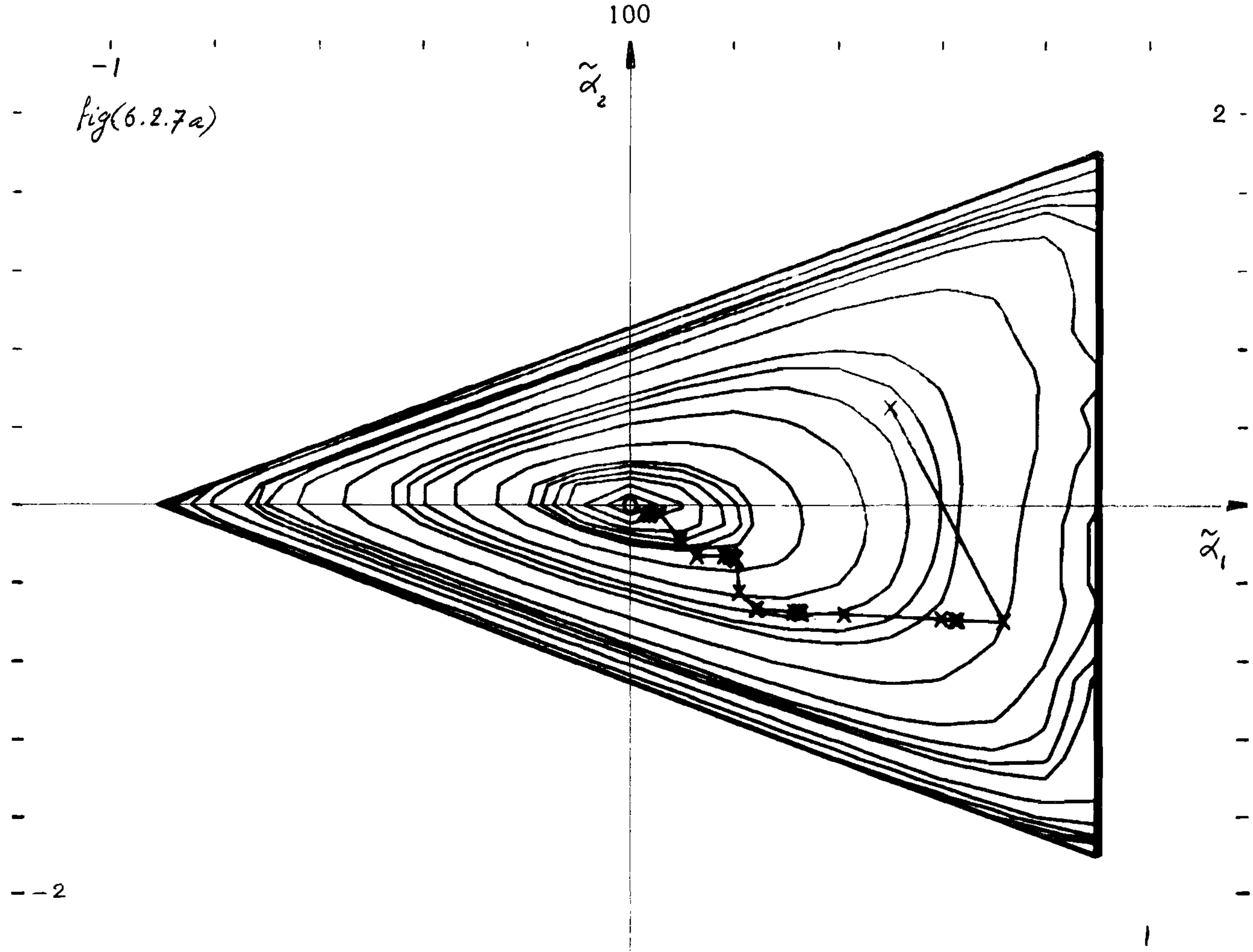
fig.(6.2.6b)

CONJUGATE GRADIENT METHOD (1)

	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	E
ADJUSTABLE PARAMETERS	T	T	T	T	T	T	
TRUE PARAMETERS	0.0000E+00	0.0000E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.0000E+00
START PARAMETERS	0.5000E+00	0.5000E+00	0.1000E+01	0.2000E+01	0.3000E+01	0.4000E+01	0.1024E+07
ITERATION NO.							
1	0.7208E+00	-0.6032E+00	0.1000E+01	-0.1445E+00	0.3513E+00	-0.1109E+00	0.6449E+04
2	0.6335E+00	-0.5930E+00	0.1000E+01	0.1055E-01	0.6232E-01	0.9580E-02	0.4595E+02
3	0.6282E+00	-0.5922E+00	0.1000E+01	0.1057E-01	0.6045E-01	0.2496E-02	0.4359E+02
4	0.6015E+00	-0.5888E+00	0.1000E+01	0.1836E-01	0.5804E-01	0.3303E-02	0.4033E+02
5	0.4140E+00	-0.5642E+00	0.1000E+01	0.2468E-01	0.7858E-01	0.4819E-02	0.1702E+02
6	0.3314E+00	-0.5511E+00	0.1000E+01	0.1741E-01	0.6569E-01	0.9362E-02	0.7236E+01
7	0.3270E+00	-0.5504E+00	0.1000E+01	0.1600E-01	0.6762E-01	0.8138E-02	0.6686E+01
8	0.3266E+00	-0.5505E+00	0.1000E+01	0.1715E-01	0.6804E-01	0.8792E-02	0.6578E+01
9	0.3259E+00	-0.5504E+00	0.1000E+01	0.1747E-01	0.6892E-01	0.8435E-02	0.6536E+01
10	0.3164E+00	-0.5486E+00	0.1000E+01	0.1931E-01	0.6803E-01	0.6284E-02	0.6346E+01
11	0.2454E+00	-0.5328E+00	0.1000E+01	0.1805E-01	0.6776E-01	0.8229E-02	0.5050E+01
12	0.2430E+00	-0.5320E+00	0.1000E+01	0.1775E-01	0.6685E-01	0.8033E-02	0.5015E+01
13	0.2425E+00	-0.5292E+00	0.1000E+01	0.1776E-01	0.6729E-01	0.7669E-02	0.4992E+01
14	0.2110E+00	-0.4471E+00	0.1000E+01	0.2720E-01	0.7202E-01	0.6559E-02	0.4379E+01
15	0.2116E+00	-0.4464E+00	0.1000E+01	0.2476E-01	0.6912E-01	0.5571E-02	0.3820E+01
16	0.2110E+00	-0.4462E+00	0.1000E+01	0.2479E-01	0.6866E-01	0.6028E-02	0.3789E+01
17	0.2116E+00	-0.4456E+00	0.1000E+01	0.2410E-01	0.6888E-01	0.6806E-02	0.3767E+01
18	0.2048E+00	-0.2814E+00	0.1000E+01	0.3709E-01	0.7357E-01	0.4163E-02	0.2582E+01
19	0.2046E+00	-0.2783E+00	0.1000E+01	0.3696E-01	0.7388E-01	0.3927E-02	0.2558E+01
20	0.2024E+00	-0.2784E+00	0.1000E+01	0.3821E-01	0.7371E-01	0.4915E-02	0.2513E+01
21	0.1967E+00	-0.2814E+00	0.1000E+01	0.3657E-01	0.7467E-01	0.4552E-02	0.2424E+01
22	0.1966E+00	-0.2614E+00	0.1000E+01	0.3718E-01	0.7419E-01	0.4889E-02	0.2346E+01
23	0.1944E+00	-0.2615E+00	0.1000E+01	0.3790E-01	0.7380E-01	0.3682E-02	0.2307E+01
24	0.1822E+00	-0.2616E+00	0.1000E+01	0.3930E-01	0.7527E-01	0.3749E-02	0.2147E+01
25	0.1287E+00	-0.2601E+00	0.1000E+01	0.3774E-01	0.7229E-01	0.3770E-02	0.1498E+01
26	0.1284E+00	-0.2600E+00	0.1000E+01	0.3761E-01	0.7234E-01	0.3654E-02	0.1493E+01
27	0.1279E+00	-0.2591E+00	0.1000E+01	0.3750E-01	0.7197E-01	0.3493E-02	0.1488E+01
28	0.9787E-01	-0.1828E+00	0.1000E+01	0.4178E-01	0.7389E-01	0.1631E-02	0.1077E+01
29	0.9788E-01	-0.1828E+00	0.1000E+01	0.4256E-01	0.7384E-01	0.2485E-02	0.8003E+00
30	0.9782E-01	-0.1827E+00	0.1000E+01	0.4289E-01	0.7374E-01	0.2547E-02	0.7915E+00
31	0.9714E-01	-0.1803E+00	0.1000E+01	0.4365E-01	0.7369E-01	0.1901E-02	0.7815E+00
32	0.9622E-01	-0.1730E+00	0.1000E+01	0.4435E-01	0.7362E-01	0.2006E-02	0.7563E+00
33	0.5691E-01	-0.4278E-01	0.1000E+01	0.5246E-01	0.7675E-01	0.5960E-03	0.2410E+00
34	0.5667E-01	-0.4216E-01	0.1000E+01	0.5233E-01	0.7670E-01	0.5814E-03	0.2387E+00
35	0.5565E-01	-0.4016E-01	0.1000E+01	0.5196E-01	0.7647E-01	0.8648E-03	0.2304E+00
36	0.5486E-01	-0.4043E-01	0.1000E+01	0.5237E-01	0.7684E-01	0.8601E-03	0.2236E+00
37	0.5454E-01	-0.4063E-01	0.1000E+01	0.5230E-01	0.7691E-01	0.7811E-03	0.2215E+00
38	0.3000E-01	-0.4548E-01	0.1000E+01	0.5867E-01	0.7553E-01	-0.5786E-03	0.1854E+00
39	0.2706E-01	-0.4850E-01	0.1000E+01	0.5159E-01	0.7553E-01	0.5545E-03	0.6301E-01
40	0.2700E-01	-0.4851E-01	0.1000E+01	0.5160E-01	0.7551E-01	0.5787E-03	0.6275E-01
41	0.2893E-01	-0.4842E-01	0.1000E+01	0.5153E-01	0.7548E-01	0.5687E-03	0.6250E-01
42	0.2527E-01	-0.4525E-01	0.1000E+01	0.5157E-01	0.7558E-01	0.8810E-03	0.5848E-01

CONVERGENCE NOT ACHIEVED IN 43 ITERATIONS
STOP -- PARAMETER ESTIMATION

see figures (6.2.7a) and (6.2.7b)


fig(6.2.7a)
100
-1
2
--2
$\tilde{\alpha}_2$
$\tilde{\alpha}_1$

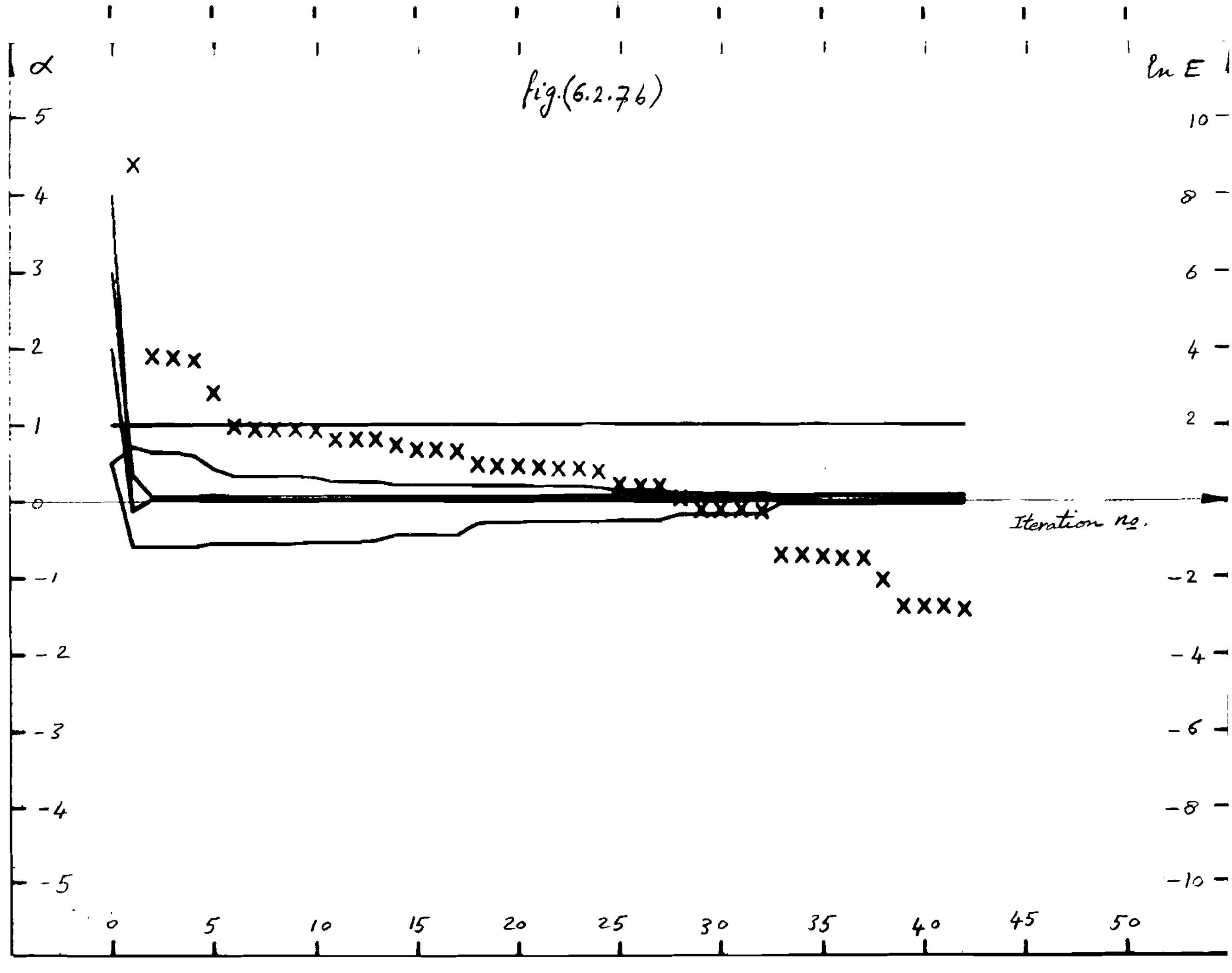


fig.(6.2.7b)
α
ln E
Iteration no.
5
4
3
2
1
0
-1
-2
-3
-4
-5
10
8
6
4
2
-2
-4
-6
-8
-10
0
5
10
15
20
25
30
35
40
45
50

PEST DESCENT METHOD (0)

	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	E
STABLE PARAMETERS	T	T	F	F	F	F	
PARAMETERS	0.3679E+00	-0.1237E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.0000E+00
T PARAMETERS	0.0000E+00	0.5000E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1937E+04
ATION NO.							
1	0.2627E+00	0.1205E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1624E+04
2	-0.3349E+00	-0.5254E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.3343E+03
3	0.2406E+00	-0.1064E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1117E+03
4	0.2193E+00	-0.1087E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.3236E+02
5	0.2234E+00	-0.1096E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.3099E+02
6	0.2277E+00	-0.1094E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2988E+02
7	0.2271E+00	-0.1097E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2924E+02
8	0.2332E+00	-0.1098E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2886E+02
9	0.2318E+00	-0.1101E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2764E+02
10	0.2417E+00	-0.1105E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2664E+02
11	0.2396E+00	-0.1109E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2489E+02
12	0.2453E+00	-0.1110E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2446E+02
13	0.2433E+00	-0.1115E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2383E+02
14	0.2468E+00	-0.1113E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2311E+02
15	0.2464E+00	-0.1116E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2261E+02
16	0.2512E+00	-0.1117E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2225E+02
17	0.2494E+00	-0.1121E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2192E+02
18	0.2527E+00	-0.1120E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2107E+02
19	0.2525E+00	-0.1123E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2066E+02
20	0.2559E+00	-0.1122E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.2033E+02
21	0.2553E+00	-0.1125E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1978E+02
22	0.2606E+00	-0.1128E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1934E+02
23	0.2593E+00	-0.1129E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1853E+02
24	0.2629E+00	-0.1129E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1824E+02
25	0.2617E+00	-0.1133E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1790E+02
26	0.2646E+00	-0.1130E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1734E+02
27	0.2645E+00	-0.1134E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1698E+02
28	0.2679E+00	-0.1134E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1670E+02
29	0.2669E+00	-0.1136E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1633E+02
30	0.2696E+00	-0.1137E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1598E+02
31	0.2693E+00	-0.1139E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1557E+02
32	0.2729E+00	-0.1139E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1528E+02
33	0.2721E+00	-0.1140E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1490E+02
34	0.2750E+00	-0.1142E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1453E+02
35	0.2743E+00	-0.1145E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1420E+02
36	0.2768E+00	-0.1144E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1389E+02
37	0.2763E+00	-0.1146E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1381E+02
38	0.2789E+00	-0.1147E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1381E+02
39	0.2790E+00	-0.1148E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1293E+02
40	0.2816E+00	-0.1149E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1268E+02
41	0.2811E+00	-0.1151E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1237E+02
42	0.2836E+00	-0.1151E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1213E+02

ERGENCE NOT ACHIEVED IN 43 ITERATIONS
— PARAMETER ESTIMATION

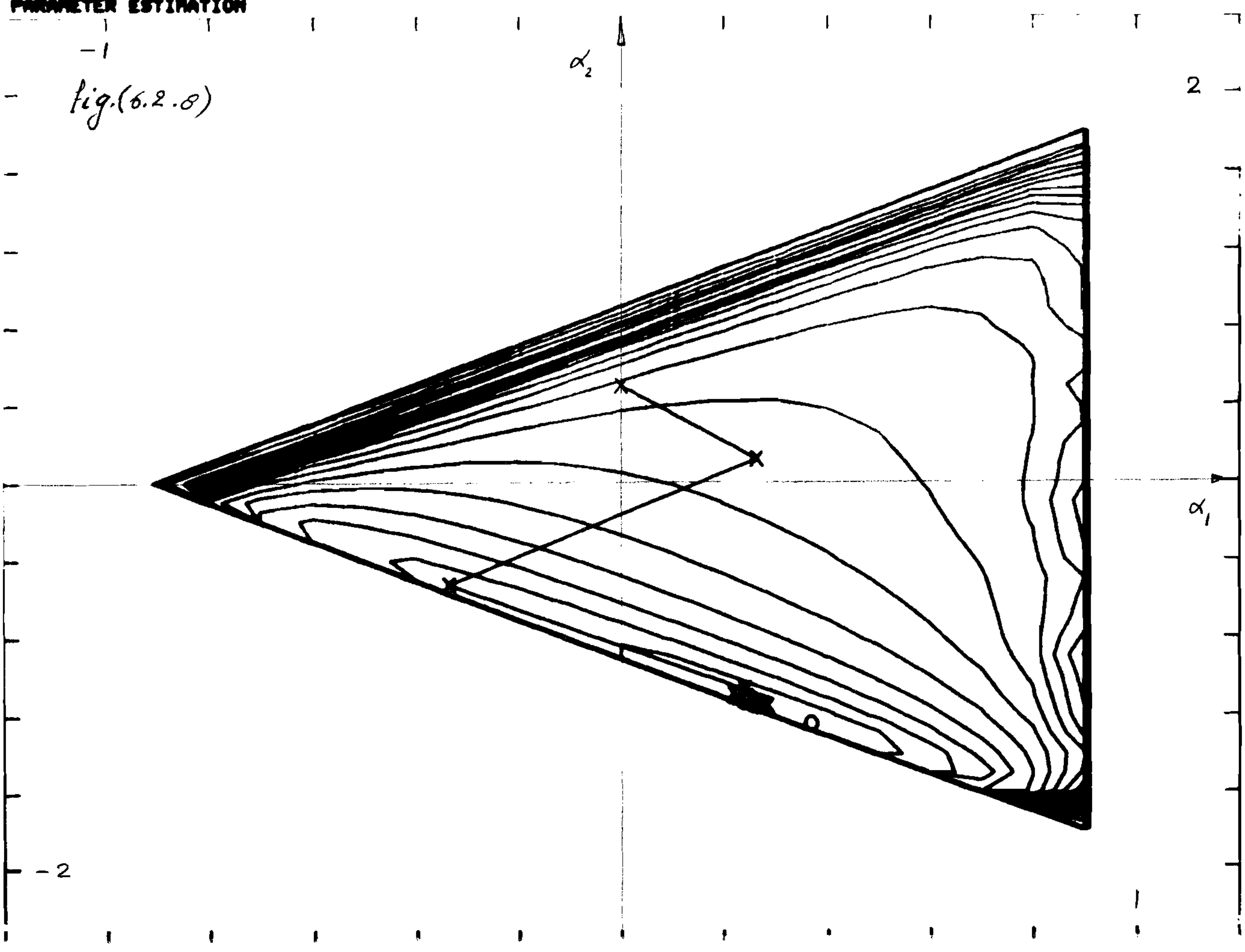


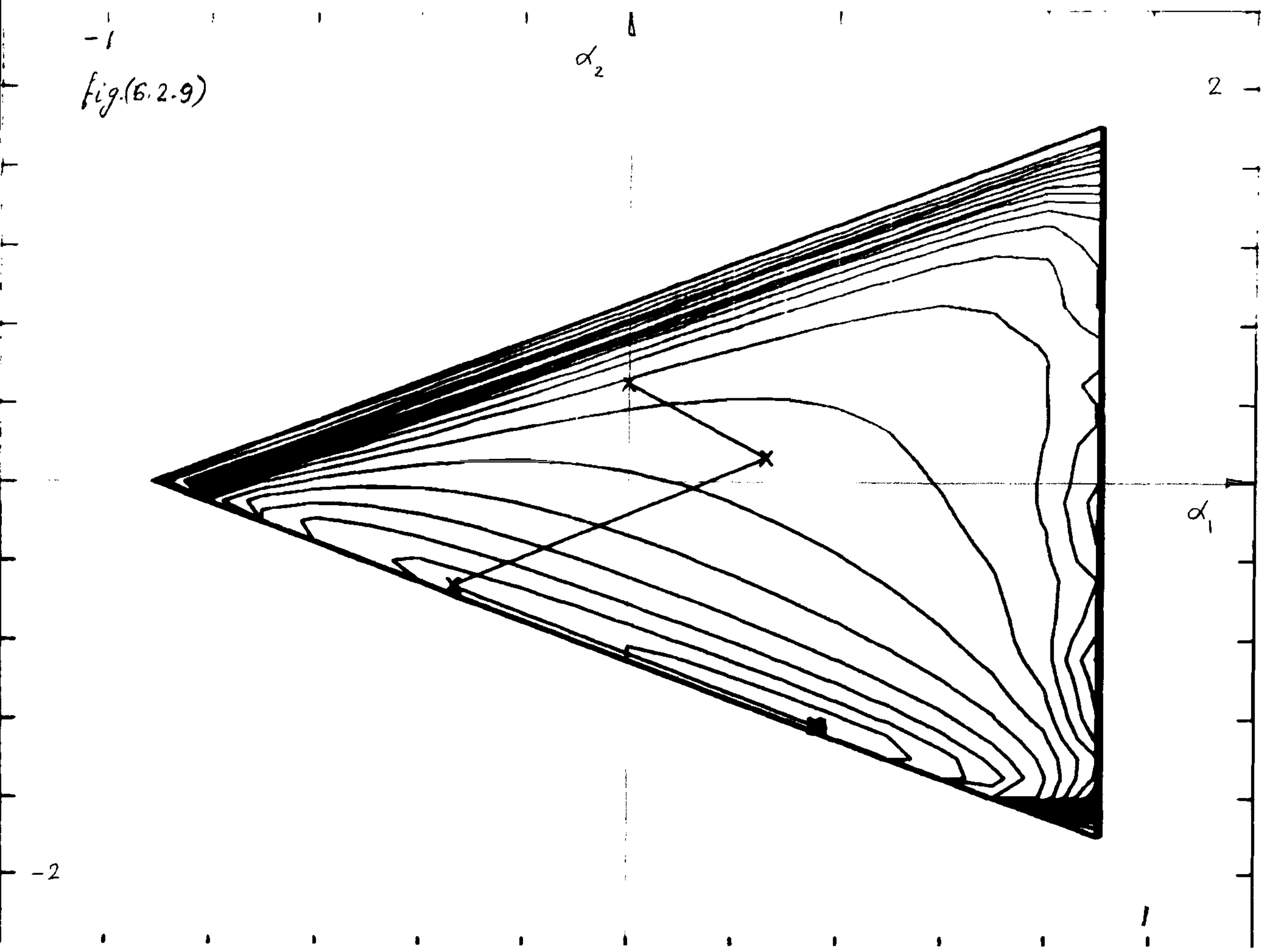
fig.(6.2.8)

ED CONJUGATE GRADIENT METHOD (2)

	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	E
ABLE PARAMETERS	T	T	F	F	F	F	
ARAMETERS	0.3679E+00	-0.1237E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.0000E+00
PARAMETERS	0.0000E+00	0.5000E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1937E+04
ION NO.							
1	0.2627E+00	0.1205E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1624E+04
2	-0.3335E+00	-0.5268E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.3333E+03
3	-0.3329E+00	-0.5332E+00	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.3323E+03
4	0.3612E+00	-0.1233E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.4870E+00
5	0.3620E+00	-0.1231E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.6819E-01
6	0.3679E+00	-0.1237E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.1155E-05
7	0.3679E+00	-0.1237E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.7006E-10

TED PARAMETERS ARE:
0.3679000437E+00
-0.1236999989E+01
0.1000000000E+01
0.5446999893E-01
0.7601000369E-01
0.0000000000E+00
- PARAMETER ESTIMATION

fig.(6.2.9)



STEEPEST DESCENT METHOD (O)

	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	E
ADJUSTABLE PARAMETERS	T	T	T	T	T	T	
TRUE PARAMETERS	0.3679E+00	-0.1237E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.0000E+00
START PARAMETERS	0.0000E+00	0.5000E+00	0.1000E+01	0.2000E+01	0.3000E+01	0.4000E+01	0.1010E+08
ITERATION NO.							
1	0.9708E+00	-0.9737E+00	0.1000E+01	-0.1573E+00	0.1767E+00	-0.1526E+00	0.9285E+04
2	0.9308E+00	-0.9431E+00	0.1000E+01	-0.3141E-01	-0.8805E-02	-0.8224E-01	0.3193E+04
3	0.8626E+00	-0.9017E+00	0.1000E+01	-0.9639E-02	0.3648E-01	-0.1682E-01	0.1781E+04
4	0.8596E+00	-0.9018E+00	0.1000E+01	0.1739E-01	0.2006E-01	-0.1496E-01	0.1664E+04
5	0.8140E+00	-0.8928E+00	0.1000E+01	0.1482E-01	0.2261E-01	-0.4827E-02	0.1647E+04
6	0.8099E+00	-0.8936E+00	0.1000E+01	0.2699E-01	0.2802E-01	-0.1701E-01	0.1621E+04
7	0.7620E+00	-0.8886E+00	0.1000E+01	0.2666E-01	0.2393E-01	-0.5943E-02	0.1598E+04
8	0.7573E+00	-0.8901E+00	0.1000E+01	0.3239E-01	0.3723E-01	-0.1789E-01	0.1569E+04
9	0.7444E+00	-0.8909E+00	0.1000E+01	0.4162E-01	0.3322E-01	-0.1318E-01	0.1556E+04
10	0.7349E+00	-0.8916E+00	0.1000E+01	0.3707E-01	0.4052E-01	-0.1929E-01	0.1543E+04
11	0.7218E+00	-0.8931E+00	0.1000E+01	0.4614E-01	0.3713E-01	-0.1300E-01	0.1528E+04
12	0.7120E+00	-0.8944E+00	0.1000E+01	0.4141E-01	0.4459E-01	-0.1968E-01	0.1513E+04
13	0.6987E+00	-0.8968E+00	0.1000E+01	0.5097E-01	0.4104E-01	-0.1302E-01	0.1496E+04
14	0.6886E+00	-0.8988E+00	0.1000E+01	0.4592E-01	0.4893E-01	-0.2006E-01	0.1479E+04
15	0.6750E+00	-0.9022E+00	0.1000E+01	0.5606E-01	0.4521E-01	-0.1292E-01	0.1460E+04
16	0.6640E+00	-0.9049E+00	0.1000E+01	0.5061E-01	0.5361E-01	-0.2041E-01	0.1440E+04
17	0.6505E+00	-0.9095E+00	0.1000E+01	0.6141E-01	0.4966E-01	-0.1268E-01	0.1418E+04
18	0.6397E+00	-0.9132E+00	0.1000E+01	0.5550E-01	0.5865E-01	-0.2074E-01	0.1395E+04
19	0.6251E+00	-0.9192E+00	0.1000E+01	0.6714E-01	0.5444E-01	-0.1232E-01	0.1369E+04
20	0.6138E+00	-0.9239E+00	0.1000E+01	0.6063E-01	0.6413E-01	-0.2109E-01	0.1341E+04
21	0.5985E+00	-0.9317E+00	0.1000E+01	0.7330E-01	0.5957E-01	-0.1181E-01	0.1310E+04
22	0.5865E+00	-0.9379E+00	0.1000E+01	0.6602E-01	0.7009E-01	-0.2150E-01	0.1276E+04
23	0.5704E+00	-0.9479E+00	0.1000E+01	0.7998E-01	0.6508E-01	-0.1116E-01	0.1237E+04
24	0.5578E+00	-0.9559E+00	0.1000E+01	0.7167E-01	0.7667E-01	-0.2206E-01	0.1194E+04
25	0.5404E+00	-0.9688E+00	0.1000E+01	0.8726E-01	0.7096E-01	-0.1037E-01	0.1145E+04
26	0.5264E+00	-0.9791E+00	0.1000E+01	0.7752E-01	0.8356E-01	-0.2294E-01	0.1089E+04
27	0.5080E+00	-0.9959E+00	0.1000E+01	0.9617E-01	0.7714E-01	-0.9586E-02	0.1024E+04
28	0.4984E+00	-0.1010E+01	0.1000E+01	0.8331E-01	0.9083E-01	-0.8442E-01	0.9475E+03
29	0.4726E+00	-0.1032E+01	0.1000E+01	0.1034E+00	0.8348E-01	-0.8733E-02	0.8578E+03
30	0.4546E+00	-0.1060E+01	0.1000E+01	0.8830E-01	0.9762E-01	-0.2699E-01	0.7500E+03
31	0.4340E+00	-0.1078E+01	0.1000E+01	0.1106E+00	0.8949E-01	-0.8817E-02	0.6248E+03
32	0.4136E+00	-0.1108E+01	0.1000E+01	0.9079E-01	0.1011E+00	-0.3079E-01	0.4778E+03
33	0.3964E+00	-0.1132E+01	0.1000E+01	0.1107E+00	0.9485E-01	-0.1263E-01	0.3307E+03
34	0.3778E+00	-0.1153E+01	0.1000E+01	0.8975E-01	0.9646E-01	-0.3847E-01	0.1983E+03
35	0.3688E+00	-0.1172E+01	0.1000E+01	0.1002E+00	0.9458E-01	-0.2068E-01	0.1137E+03
36	0.3589E+00	-0.1182E+01	0.1000E+01	0.8628E-01	0.8680E-01	-0.2806E-01	0.6493E+02
37	0.3550E+00	-0.1191E+01	0.1000E+01	0.9015E-01	0.8700E-01	-0.2274E-01	0.4080E+02
38	0.3508E+00	-0.1193E+01	0.1000E+01	0.8198E-01	0.8238E-01	-0.2389E-01	0.2790E+02
39	0.3488E+00	-0.1198E+01	0.1000E+01	0.8306E-01	0.8330E-01	-0.2028E-01	0.2054E+02
40	0.3468E+00	-0.1198E+01	0.1000E+01	0.7710E-01	0.8038E-01	-0.1931E-01	0.1559E+02
41	0.3457E+00	-0.1200E+01	0.1000E+01	0.7765E-01	0.8133E-01	-0.1870E-01	0.1207E+02
42	0.3447E+00	-0.1199E+01	0.1000E+01	0.7313E-01	0.7961E-01	-0.1546E-01	0.9453E+01

CONVERGENCE NOT ACHIEVED IN 43 ITERATIONS
STOP -- PARAMETER ESTIMATION

See figures (6 2.10 a) and (6.2.10 b)

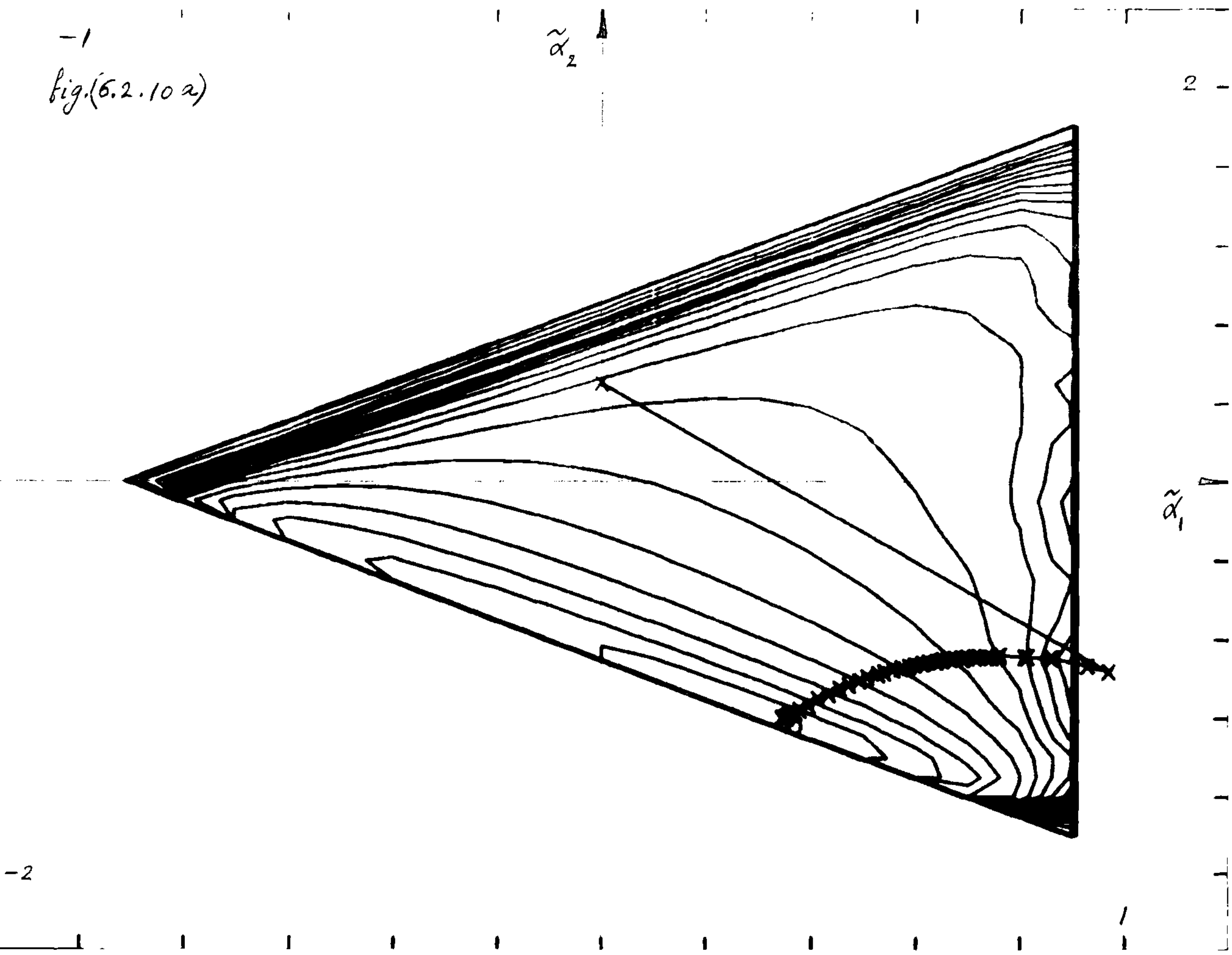


fig.(6.2.10a)

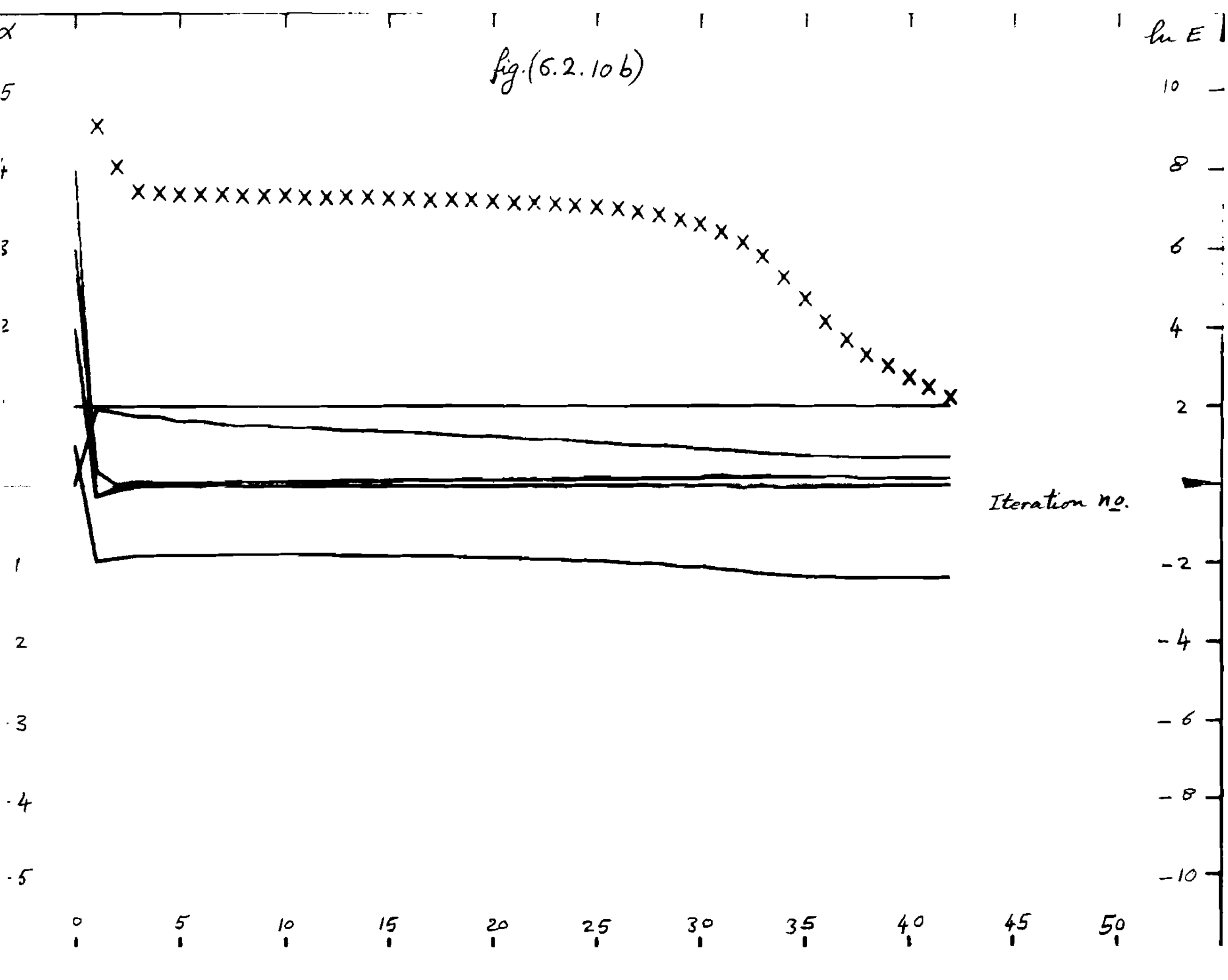


fig.(6.2.10b)

CONJUGATE GRADIENT METHOD (1)

	A(1)	A(2)	A(3)	A(4)	A(5)	A(6)	E
ADJUSTABLE PARAMETERS	T	T	T	T	T	T	
TRUE PARAMETERS	0.3679E+00	-0.1237E+01	0.1000E+01	0.5447E-01	0.7601E-01	0.0000E+00	0.0000E+00
START PARAMETERS	0.0000E+00	0.5000E+00	0.1000E+01	0.2000E+01	0.3000E+01	0.4000E+01	0.1010E+08
ITERATION NO.							
1	0.9708E+00	-0.9737E+00	0.1000E+01	-0.1573E+00	0.1767E+00	-0.1526E+00	0.9285E+04
2	0.9308E+00	-0.9431E+00	0.1000E+01	-0.3141E-01	-0.8805E-02	-0.8224E-01	0.3193E+04
3	0.8658E+00	-0.9022E+00	0.1000E+01	0.1566E-01	-0.1500E-01	-0.1247E-01	0.1819E+04
4	0.8574E+00	-0.8966E+00	0.1000E+01	0.2129E-01	0.1915E-01	-0.6381E-02	0.1672E+04
5	0.8488E+00	-0.8947E+00	0.1000E+01	0.1698E-01	0.2377E-01	-0.1542E-01	0.1658E+04
6	0.7606E+00	-0.8793E+00	0.1000E+01	0.2883E-01	0.1208E-01	-0.2612E-01	0.1624E+04
7	-0.2023E-01	-0.7493E+00	0.1000E+01	0.1756E+00	0.8479E-01	-0.5676E-01	0.3910E+03
8	-0.2185E-01	-0.7522E+00	0.1000E+01	0.1580E+00	0.1027E+00	-0.6791E-01	0.1821E+03
9	-0.2952E-01	-0.7644E+00	0.1000E+01	0.1530E+00	0.1158E+00	-0.5139E-01	0.1314E+03
10	-0.3595E-01	-0.7663E+00	0.1000E+01	0.9963E-01	0.1074E+00	-0.4705E-02	0.5088E+02
11	-0.3991E-01	-0.7740E+00	0.1000E+01	0.9732E-01	0.1055E+00	-0.5346E-02	0.4814E+02
12	-0.4004E-01	-0.7755E+00	0.1000E+01	0.9621E-01	0.1027E+00	-0.5981E-02	0.4743E+02
13	-0.1718E-01	-0.8050E+00	0.1000E+01	0.9324E-01	0.1025E+00	-0.8395E-02	0.4616E+02
14	0.8132E-01	-0.9275E+00	0.1000E+01	0.8767E-01	0.9431E-01	-0.1079E-01	0.3477E+02
15	0.8434E-01	-0.9222E+00	0.1000E+01	0.8621E-01	0.9193E-01	-0.1090E-01	0.2816E+02
16	0.8626E-01	-0.9205E+00	0.1000E+01	0.8764E-01	0.9219E-01	-0.8106E-02	0.2637E+02
17	0.8881E-01	-0.9191E+00	0.1000E+01	0.8702E-01	0.9632E-01	-0.6922E-02	0.2581E+02
18	0.9207E-01	-0.9224E+00	0.1000E+01	0.8706E-01	0.9477E-01	-0.5802E-02	0.2496E+02
19	0.1300E+00	-0.9606E+00	0.1000E+01	0.7935E-01	0.9149E-01	-0.1712E-02	0.2199E+02
20	0.2162E+00	-0.1059E+01	0.1000E+01	0.6483E-01	0.9007E-01	0.6830E-02	0.1525E+02
21	0.3346E+00	-0.1204E+01	0.1000E+01	0.4530E-01	0.7983E-01	0.7231E-02	0.6588E+01
22	0.3360E+00	-0.1206E+01	0.1000E+01	0.4890E-01	0.7807E-01	0.8068E-02	0.4276E+01
23	0.3358E+00	-0.1203E+01	0.1000E+01	0.5062E-01	0.7611E-01	0.6878E-02	0.3100E+01
24	0.3370E+00	-0.1206E+01	0.1000E+01	0.5751E-01	0.7701E-01	-0.1017E-02	0.6317E+00
25	0.3380E+00	-0.1204E+01	0.1000E+01	0.5862E-01	0.7803E-01	-0.1113E-02	0.4032E+00
26	0.3392E+00	-0.1204E+01	0.1000E+01	0.5868E-01	0.7815E-01	-0.1290E-02	0.3492E+00
27	0.3473E+00	-0.1214E+01	0.1000E+01	0.5716E-01	0.7792E-01	-0.1423E-02	0.2501E+00
28	0.3561E+00	-0.1224E+01	0.1000E+01	0.5614E-01	0.7748E-01	-0.8665E-03	0.1439E+00
29	0.3563E+00	-0.1224E+01	0.1000E+01	0.5606E-01	0.7711E-01	-0.8816E-03	0.8295E-01
30	0.3564E+00	-0.1224E+01	0.1000E+01	0.5628E-01	0.7688E-01	-0.6012E-03	0.5792E-01
31	0.3575E+00	-0.1225E+01	0.1000E+01	0.5560E-01	0.7669E-01	-0.1638E-03	0.4996E-01
32	0.3577E+00	-0.1228E+01	0.1000E+01	0.5557E-01	0.7670E-01	-0.1106E-03	0.4887E-01
33	0.3633E+00	-0.1231E+01	0.1000E+01	0.5485E-01	0.7668E-01	0.4222E-04	0.2269E-01
34	0.3678E+00	-0.1237E+01	0.1000E+01	0.5450E-01	0.7600E-01	0.5612E-04	0.1124E-02
35	0.3680E+00	-0.1237E+01	0.1000E+01	0.5448E-01	0.7601E-01	0.2922E-04	0.2600E-03
36	0.3680E+00	-0.1237E+01	0.1000E+01	0.5447E-01	0.7600E-01	0.1350E-04	0.6209E-04
37	0.3679E+00	-0.1237E+01	0.1000E+01	0.5446E-01	0.7601E-01	0.1278E-05	0.2603E-05
38	0.3679E+00	-0.1237E+01	0.1000E+01	0.5446E-01	0.7601E-01	0.1075E-05	0.8880E-06

ESTIMATED PARAMETERS ARE:
P(1)= 0.3679423338E+00
P(2)= -0.1237048507E+01
P(3)= 0.1000000000E+01
P(4)= 0.5446491763E-01
P(5)= 0.7600706320E-01
P(6)= 0.1075081228E-06

See figures (6.2.11a) and (6.2.11b)

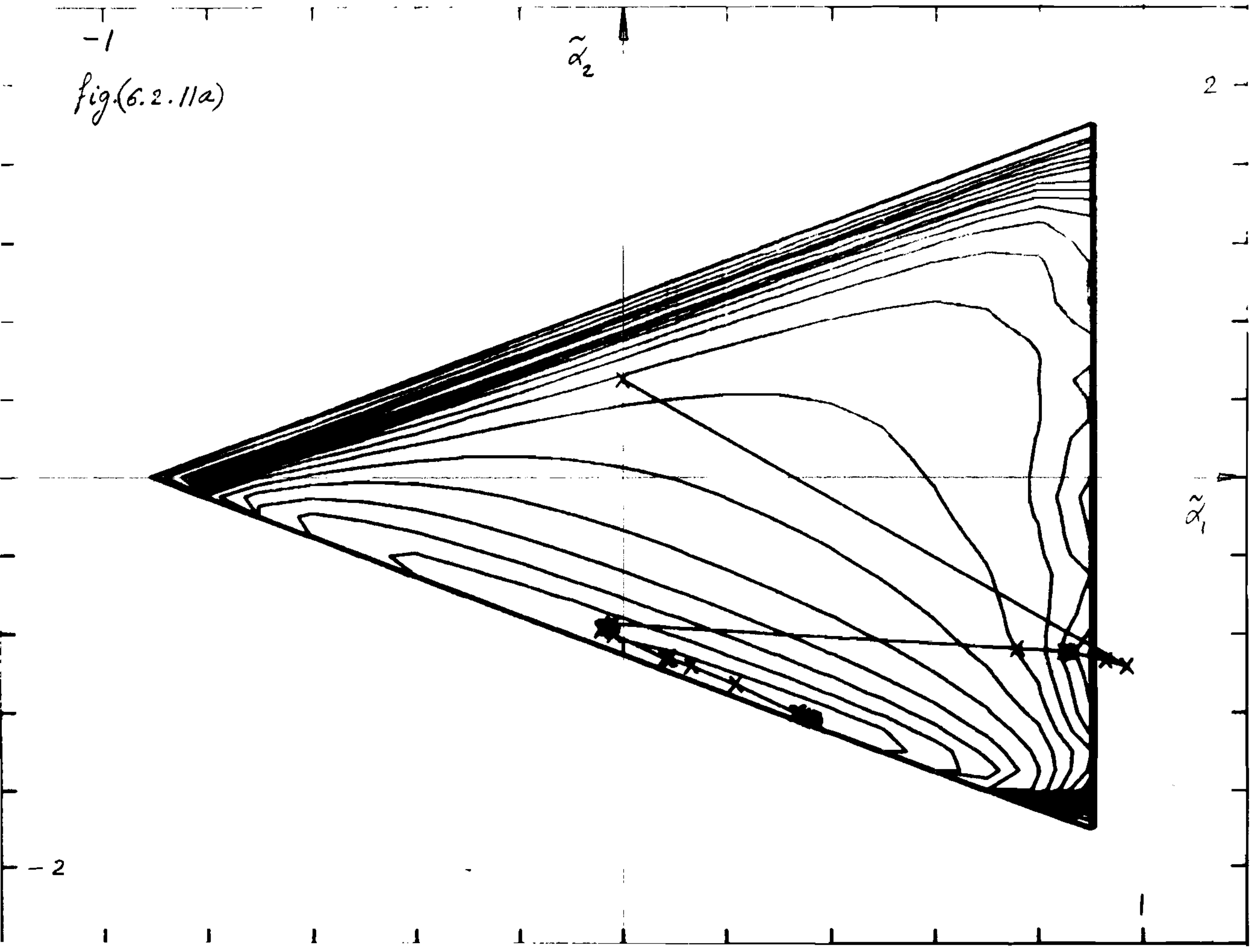
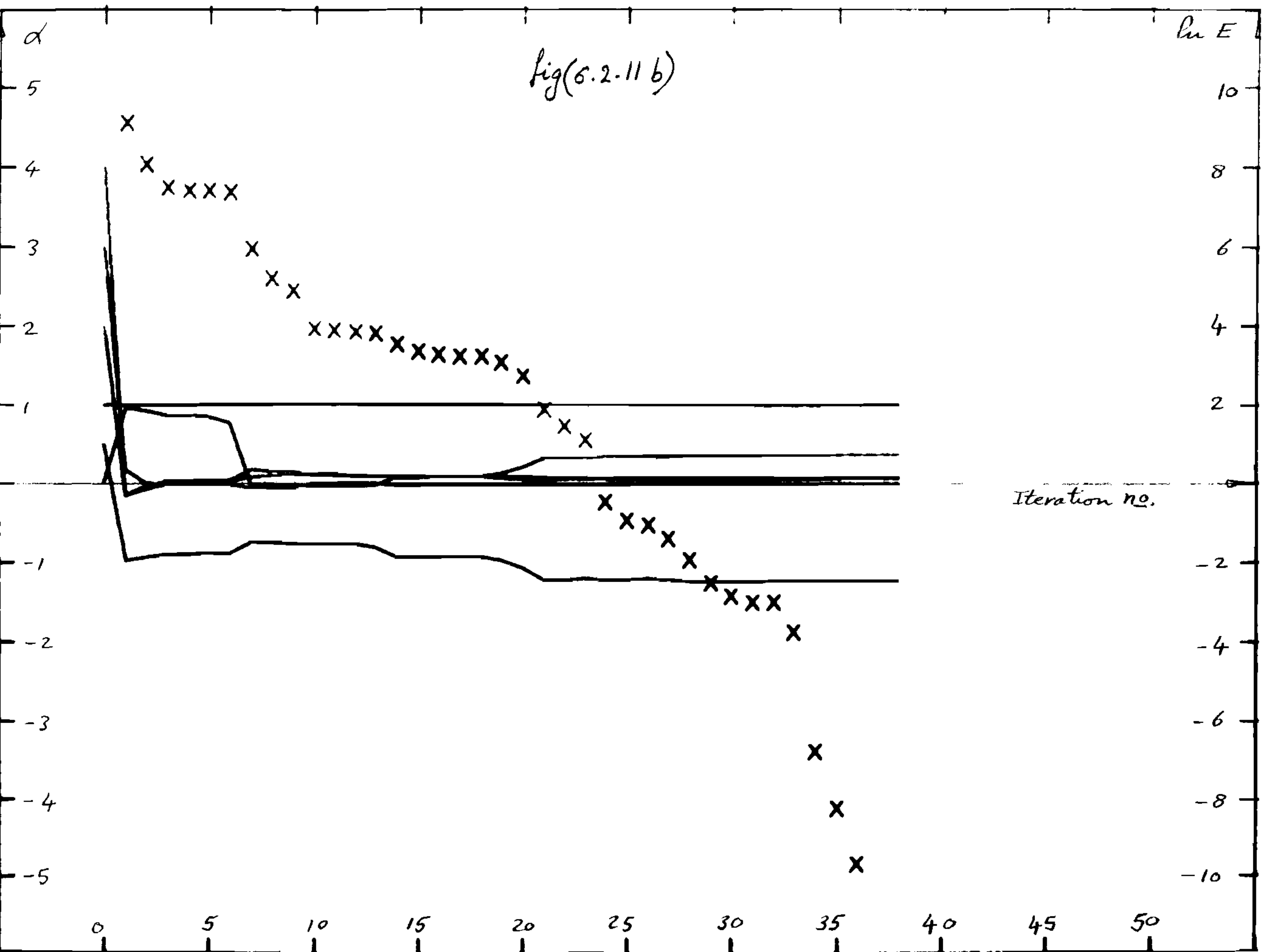


fig.(6.2.11a)



fig(6.2.11 b)

6.3. Conclusions and suggestions:

6.3.1. Quadratic error criterion:

- Analytic explicit solutions can be derived if the number of independent observations is equal to or more than the number of unknown parameters.
- In case of noisy observations beter estimates can be obtained by increasing the total number of observations used for estimation
- Constant levels of the error criterion function form concentric ellipsoids whose centre is the true parameter values in the parameter space
- The input signal spectrum has no effect on the shape of the ellipsoids of the error criterion function; it may affect the steepness of the ellipsoids due to changes in the signal amplitude.
- If iterative "hill climbing" techniques are used to approach the solution the steepness of the error criterion ellipsoids can be affected by increasing the number of independent observations used for estimation and/or increasing the input signal amplitude level.
- If quadratically convergent methods are applied, the exact minimum is located in a certain number of iterations. The conjugate gradient methods perform similarly and they converge in as many iterations as the number of unknown parameters.
- If the ellipsoids are well conditioned the steepest descent method converges in a limited number of iterations, but it may have a very slow convergence and an oscillatory behaviour if the ellipses are very steep along one direction but not the others.
- More investigations can be done on the input signal parameters which affects the character of the ellipsoids of the error criterion function.

6.3.2. Nonquadratic error criterion:

- Iterative "hill climbing" techniques are used to approach the solution and the convergence to the minimum is highly dependent on the shape of the error criterion function in the parameter space.
- The shape of the error criterion function is dependent on the input signal spectrum as well as the location of the true process parameters in the parameter space.
- The input signal can be suitably chosen to affect, desirably, the shape of the error criterion function in the parameter space. Many fruitful applications and extentions of this investigation should be considered

in further work.

- The steepest descent method may converge in a limited number of iterations if the true process parameters are suitably located in the feasible region of the parameters space. It behaves very poorly in most of the practical situations.
- No remarkable difference has been detected between the performance of the conjugate gradient and the modified conjugate gradient methods; they behave almost similarly in many cases. They are much superior to the steepest descent method in that they converge fast to the exact minimum from any point in the neighbourhood of the minimum where the criterion function can be well approximated by a quadratic; thus, also in cases of very steep ellipsoids.
- It seems reasonable to suggest an algorithm which starts with the steepest descent method and switches to the conjugate gradient method once the minimum is approached or the convergence becomes slow. This is profitable since the steepest descent is computationaly simpler.
- If the number of the unknown parameters increases, more iterations are needed for convergence. It is of much importance to mention that the convergence in the case of letting all the parameters to be adjustable is better than the convergence in the case of fixing one of the parameters to its true value and adjusting the other parameters; this is an interesting result which may be explained in further investigations.

Literature.

1. Bard, Y.,
Nonlinear parameter estimation,
Academic Press Inc., New York 1974.

2. Barret, J.F.,
Lecture notes on statistical estimation and filtering applied to stochastic processes,
Dept. of Elec. Eng., University of Technology, Eindhoven 1978.

3. Box, M.J., D. Davies and W.H. Swann,
Non-linear optimization techniques,
Oliver & Boyd for Imperial Chemical Industries Ltd., London 1969.

4. Davidon, W.C.,
Variable metric method for minimization,
Argonne National Laboratory report ANL-5990, December 1959.

5. Eykhoff, P.,
System identification; parameter and state estimation,
John Wiley & Sons Ltd., London 1974.

6. Fletcher, R. and M.J.D. Powell,
A rapidly convergent descent method for minimization,
Computer J., Vol. 6, No. 2, pp. 163-168, July 1963.

7. Fletcher, R., and C.M. Reeves,
Function minimization by conjugate gradients,
Computer J., Vol. 7, pp. 149-154, 1964.

8. Harbers, J.W.L.,
Parameter estimation with a self adjusting analog model,
M. Sc. thesis, University of Technology, Eindhoven 1975.

9. Jennrich, R.I.,
Asymptotic properties of nonlinear least squares estimators,
The Annals. of Mathematical Statistics, Vol. 40, No. 2, pp. 633-643, 1969.

10. Jong, J.L. de,
Lecture notes on numerical algorithms for nonlinear optimization problems,
Dept. of Mathematics, University of Technology, Eindhoven 1976.

11. Papoulis, A.,
The Fourier integral and its applications,
McGraw-Hill Book Company Inc., New York 1962.

12. Ragazzini, J.R. and G.F. Franklin,
Sampled-data control systems,
McGraw-Hill Book Company Inc., New York 1958.

13. Rosko, J.S.,
Digital simulation of physical systems,
Addison-Wesley Publishing Company, 1972.

14. Sorenson, H.W.,
Comparison of some conjugate direction procedures for function minimization,
Journal of The Franklin Institute, Vol. 288, No. 6, pp. 421-441, 1969.

15. Timmers, W.,
Parameter estimation of an analog process with a model nonlinear-in-the-parameters (in Dutch),
M. Sc. thesis, University of Technology, Eindhoven 1977.

16. Tou, J.T.,
Digital and sampled-data control systems,
McGraw-Hill Book Company Inc., New York 1959.