TRANSFORMATION TECHNIQUES FOR THE PARAMETER ESTIMATION OF DISCRETE-TIME TRANSFER FUNCTIONS

David Walter Norris

A thesis submitted for the
degree of Doctor of Philosophy

Centre for Computing and Automation,
Imperial College of Science and Technology,
University of London.

August
1969

ABSTRACT

This thesis describes several original advances to the art of estimation of discrete time systems from data records. The theory developed here shows that systems given by rational Z polynomials are better characterised by the roots of the polynomials rather than the coefficients.

The root description allows general expressions to be found for the system response and output autocorrelations for both infinite and finite data lengths. From such expressions, the importance of filter stability in Åström's estimation method can be seen very clearly. A novel transformation is introduced which is used to restrain the filter estimates to the class of stable systems. This leads to the establishment of a new estimation method which describes systems in terms of polynomial roots

The breakdown of estimation methods for limited data sets is shown to be due to a relation between the pole 'strength' and the length of the data record. As a result criteria are developed which enable judgements to be made about the length of data required in order to expect a satisfactory estimate to be achieved. Åström's proofs of consistency etc., are shown to hold for the new root description approach. Several examples are given to illustrate the practical benefits of the new estimation method.

CONTENTS

CHAPTER 6 EXAMPLES AND RESULTS

CHAPTER 7 CONCLUSIONS AND EXTENSIONS

NOTATION

$A(z)$	n^{th} order polynomial in z
A_i	i^{th} coefficient matrix of polynomial matrix $A(z)$
a	Continuous time pole in S plane
$(a+jb),(a-jb)$	Complex conjugate roots
a_i	i^{th} coefficient of polynomial $A(z)$
$B(z)$	n^{th} order polynomial in z
B_i	i^{th} Coefficient matrix of polynomial matrix $B(z)$
b	Continuous time zero in S plane
b_i	i^{th} coefficient of polynomial $B(z)$
C	Contour of integration
$C(z)$	n^{th} order polynomial in z
C_i	i^{th} coefficient matrix of polynomial matrix $C(z)$
c_i	i^{th} coefficient of polynomial $C(z)$
D	m*r Matrix
D'	Region in the complex variable plane
$D(z)$	Denominator polynomial of junior system
d_i	i^{th} coefficient of $D(z)$ polynomial
∂	Partial differentiation operator
$E(.)$	Expectation operator
$E_o(.)$	Expectation with respect to the distribut-ion defined by the true parameters $\underline{\theta}_o$
e_k	k^{th} member of a random sequence
F	System matrix n*n
$F(z)$	General function of a complex variable z

f_{ij}	ijth element of F
f_k	General discrete time signal
$\underline{f}$	Total vector of f_i
$f(X,\theta)$	General probability distribution
G	Control matrix n*r
G_o	Constant Gain term equivalent to b_o
$G(z)$	General rational function of z
g_{ij}	ijth element of G
$g(X_1 \,\ldots\ldots\, X_n)$	Generalised estimator of θ
$g(X_1 \,\ldots\ldots\, X_n,\hat{\theta})$	Joint probability density
$g(\theta)$	Probability density of θ in Bayesian sense
H	Observation matrix m*n
H_i	ith estimate of a second derivative matrix inverse
$H(z)$	General rational function of z
h_k	General discrete time signal
$\underline{h}$	Total vector of h_k
h_{ij}	ijth element of H
I_m	Unit matrix of order m
$I^N(\underline{\theta})$	Information matrix of $\underline{\theta}$ for data record N; see(5.14)
J	n*m matrix
$K(z)$	Matrix polynomial in z
k	Discrete time index
$L(z)$	Matrix polynomial in z
$L(\hat{\theta})$	Likelihood function of $\hat{\theta}$
L'	Logarithm of $L(\theta)$
$L'^N(Y,\hat{\underline{\theta}})$	Logarithm of the likelihood function of $\hat{\underline{\theta}}$ based on data Y

$L'(\underline{\hat{\theta}},\underline{\theta}_o)$		$\lim\limits_{N\to\infty} \frac{1}{N}.E_o\ L'^N(Y,\underline{\hat{\theta}})$
$L'^N_{\theta\theta}(Y,\underline{\hat{\theta}}^N)$		Second derivative matrix of $L'^N(Y,\underline{\hat{\theta}}^N)$
$L'_{\theta\theta}(\underline{\hat{\theta}},\underline{\theta}_o)$		$\lim\limits_{N\to\infty} \frac{1}{N}.E_o\ L'^N_{\theta\theta}(Y,\underline{\hat{\theta}})$
i,j,l		General running indicies
M		n.n matrix
	or	
		Total matrix of $\underline{m}_k$ sequence products
m		Order of observation vector
$\underline{m}_k$		Measurement vector at time k
N		Length of data record
N(z)		Numerator polynomial of junior system
n		Order of system state variable
n_i		i^{th} coefficient of N(z) polynomial
m_i		Order of pole ρ_i
P		Estimate of covariance matrix
P_i		Pole in the z plane
p		Controllability or observability index
	or	
		A complex variable
Q		Matrix $\frac{\partial^2 V(\underline{\theta})}{\partial\theta_i\partial\theta_j}$
q		General variable defined locally
$\mathcal{R}$		Region in Euclidian r space
$R(\underline{\theta})$		Risk function of $\underline{\theta}$
R_i		Residual of a rational polynomial at the i^{th} pole
R_x, R_z		Radii in the X space and z plane
r	or	Order of control vector General integer constant

r_i	i^{th} coefficient of a z polynomial
T	Transformation matrix $n*n$ or Sampling period of discrete time
$T^N(Y,\hat{\underline{\theta}}^N)$	Matrix of orthogonal eigenvectors of $\frac{1}{N}.L'^{N}_{\theta\theta}(Y,\hat{\underline{\theta}}^N)$
t (superscript)	Transpose of a matrix
$\mathcal{S}_o$	Set in $\mathcal{R}$
s	Laplacian complex variable in S plane
$\underline{s}$	Sum vector of $\underline{f}$ and $\underline{h}$
V	Total vector of v_k sequence
$V(\hat{\underline{\theta}})$	Estimation cost associated with $\hat{\underline{\theta}}$
$V(k)$	Total scalar disturbance sequence at time k
$v(k), v_k$	k^{th} member of disturbance sequence v
U	Total N vector of control sequence u_k
$u(k), u_k$	k^{th} member of the control sequence
$w(k), w_k$	k^{th} member of disturbance sequence w
X , X'	n^{th} order transformation with saturating characteristics
X_1 ..... X_N	Generalised data set
$\underline{x}(k)$	n^{th} order state vector at time k
Y	Total N vector of observation sequence y_k
y_k , $y(k)$	k^{th} member of observation sequence
Z	Complex variable plane
Z_i	Zero in the Z plane
z	Unit time advance operator
α^o	Hill climbing correction factor
α_i	i^{th} root of polynomial A

β_i	i^{th} root of polynomial B
Γ	Disturbance input matrix n*n
γ_i	i^{th} of polynomial C root
$\delta(j)$	Defined as 1.0 for j=0 ; otherwise 0.0
δ_i	i^{th} root of polynomial D(z)
ϵ	Napierian base
ϵ_k	k^{th} member of a random sequence
$\zeta(\theta \mid x_1 \,.....\, x_N)$	Conditional probability density
η_i	i^{th} root of polynomial N(z)
$\underline{\theta}$	Vector of parameters
$\underline{\theta}_o$	Vector of true parameters
$\hat{\underline{\theta}}_j$	Estimate of $\underline{\theta}$ at j^{th} iteration
$\hat{\underline{\theta}}^N$	Estimate of $\underline{\theta}$ from data of length N
θ_i	i^{th} element of _
χ, χ'	Constant bias term in a signal
Λ	Covariance matrix of a disturbance signal
$\Lambda^N(Y, \hat{\underline{\theta}}^N)$	Diagonal matrix of eigenvectors of $\frac{1}{N}.L'^N_{\theta\theta}(Y, \hat{\underline{\theta}}^N)$
λ	Scalar variance of e_k
μ	Mean of a probability distribution
$\prod_{i=1}^{l}(q)$	Defined as $q_1 * q_2 * q_3 \,.....\, * q_l$
π	Projection operator
ρ_i	pole of F(z)
$\sum_{i=1}^{l}(q)$	Defined as $q_1 + q_2 + q_3 \,.....\, + q_l$
σ_e^2	Variance of a signal e
τ	Discrete time variable

Φ	Weighting matrix
$\Phi_{ff}(z)$	Power spectral density of a signal f
ϕ_r	Serial autocorrelation at delay r
ϕ_r^N	Sample estimate of ϕ_r over data record N
ϕ_r (systems)$\cdot\sigma^2$	Cross correlation between two signals produced by the named systems. see(4.51)
Ψ	Covariance matrix of $\tilde{\theta}$
$\hat{}$	Sign denoting estimated value
$\sim$	Denotes error between true and estimation values
$*$	Sign denoting a specially filtered value or A term which is not present in all cases
' (prime) " (double prime)	Used to denote a transformed variable in some sense
$^{\Delta}$ (superscript)	Implies deletion of a term
$\triangleq$	Defining equality
$\lVert \cdot \rVert$	General Norm
$\oint_c$	Integral around a closed Contour
Prob. (.)	Probability of a variable (.)

CHAPTER 1

THE STRUCTURE OF THE PROCESS

1.1 Introduction.

A problem which often arises in the areas of commissioning and running a process control system, is that of estimating the parameters of the plant. These estimates are used for deciding on controller settings for the various digital control loops within an on line computer. Alternatively we may be interested in the synthesis of a control system and its further study, and need a mathematical model of the process and its environment.

Knowledge is frequently lacking about industrial processes, and the basic equations are often dubiously known. Even if the structure of the equations governing the process can be found, the parameters of these equations are often unknown. In this thesis, we will present a technique for numerical identification of a process using measurements of the relevant input and output variables. This technique attempts to represent the observed system as a single input, single output, linear dynamical system with stationary normal disturbances having rational spectra. Such a system can be described by a transfer function with a finite number of parameters. Once a structure and its order has been chosen, the identification problem can then be regarded as a parameter estimation problem.

The assumptions made about the system are restrictive in that estimates are obtained for a linear and time invariant model. It is also assumed that the process is sampled at a fixed sampling rate.

This enables the modelling to be done in discrete time form, which is ideal for direct digital control by a process control digital computer.

The assumptions mentioned above imply a finite dimensional parameter space which is essential for the algorithm presented later based on a hill climbing procedure. Relaxations could be permitted in the assumptions as long as a given structure, for example a non-linearity, can be decided on. The mathematical formulation of the present method would not necessarily hold in this case, although engineering judgement could be exercised in this respect. Similarily the method should produce acceptable engineering results for non-gaussian disturbances, since it attempts to produce "white" or independent residual prediction errors.

One cannot expect an exact model can ever be obtained in practise from a data record of a plant. We are obliged to propose a suitable model structure and then use an estimation algorithm to assign numerical values to the parameters. It will be shown later that various models of the same order may be transformed into each other as convenient after the estimation process is finished. Thus it would be quite valid to choose a structure for estimation, which we knew was well suited to some algorithm, and later transform the model into any other desired form.

Naturally there is a risk in pre-deciding a model structure and an estimation procedure, and it remains very necessary to exercise judgement as an essential part of such a scheme. There is no advantage in proposing a complicated model or estimation procedure unless the results can be used in practise, Thus the complexity of the model or estimation will depend largely on its later intended practical use.

The algorithm will be seen to be extendable to the multidimensional input - output case without severe difficultly now that Rowe[16] has developed a suitable canonical form. Much of the work and methods shown in this thesis are similar to those of Åström[10,11,12,37] . However the claim is made that the parameter set chosen here has considerable advantages in that the algorithm has faster convergence, and valid decisions can be easily made about continuing the climbing process, or about specifying the length of data record required.

The work presented here is devoted to obtaining the most accurate estimates in the quickest manner and is not concerned with controlling plant using those estimates. A control engineer may not be interested in obtaining estimates which would give him less than say 1% of plant running cost improvement. The schemes proposed have the advantage of giving simple "rules of thumb" which can be used to make decisions about the quantities of data required to achieve an acceptable result.

1.2 Outline of the Thesis.

The rest of this chapter is devoted to a development of the model which will be used for estimating the process. It is assumed that all systems will have a state-space description, together with conditions on controllability and observability. This description is transformed to a transfer function description between measurable input and output variables.

Various estimation methods as used by previous workers are outlined in Chapter 2 together with the failure areas of their algorithms. A full description is given of the maximum likelihood algorithm as used by Åström, careful consideration shows that an alternative parameter set is better suited for estimation purposes.

A closed form solution is given in Chapter 3 for the variance of the output of a discrete time rational transfer function, whose input is white noise. Various contour plots are shown of the variance as the poles and zeros of a simple discrete time filter are moved on the z transform plane. A non linear transformation is introduced which is used to confine the pole positions in the z plane within the stable region, while allowing a hill climbing procedure to work in an unconstrained space. It is further shown that the proposed method will also cover non-minimum phase plants which have z transform zeros outside the z plane unit circle.

Chapter 4 develop~~e~~s the first and second differentials of the maximum likelihood function in the chosen parameter space. Expressions are also obtained for the variance of the second differentials and the bias arising from the finiteness of the data recorded from the original process. This leads to some simple criteria for either stopping the estimation procedure, or for pre-deciding how much data is required to be recorded from the plant. These criteria are then related to other criteria derived from more intuitive ideas, and shown to be similar in result.

In Chapter 5, the necessary statistical proofs of consistency, efficiency and unbiasedness follow similar lines to those of Åström, but are derived in the new parameter space.

Several computed examples are given in Chapter 6 to demonstrate the useful~~l~~ness of the new estimation method. The improved convergence rate of the new method is shown in comparison with Åström's method, which has been taken as the most effective method known in the literature to date. The examples have been chosen to demonstrate the progress made in areas where estimation is known to be difficult, for example, where the system disturbance is by correlated noise.

The final chapter summarises the work of the thesis, and mentions a number of areas in which more work can be done. It is shown that the new estimation scheme can be extended to the multivariable situation where both the system inputs and outputs are vector quantities. A further transformation of Rowe's canonical form[16] is required. The essential principle remains that climbing efficiency can be greatly improved by having constrained control over the eigenvalues of the dynamic system used during maximum likelihood estimation, thus ensuring system stability.

1.3 Contributions of the Thesis.

The principal contributions of this thesis, which are believed to be advances in the state of the art of estimation, are summarised below.

A fully described technique has been developed to calculate the output variance and any autocorrelation term of a discrete time system fed with a 'white' sequence, It is now clear that the roots of the defining polynomials, rather than the coefficients, are the most distinctive features of a system. This means that the above results can be given by general expressions for systems of any order. All the auto-correlation calculations have been repeated for the case of a data sequence which has a finite history. ~~and which is therefore strictly non-stationary.~~ Variance contour plots have been given for various pole-zero configurations and these have been shown to change to some degree for the finite data situation. It has been shown that zeros lying outside the unit circle in the Z plane are strongly related to continous time non-minimum phase systems, and that such systems can be satisfactorily estimated without difficulty.

The importance of filter stability has been realised when using Åström's estimation method. A novel transformation method has therefore been developed to restrain these filters to the class of stable systems. Any hill climbing procedure used in the estimation process can now work in an unconstrained space and yet ensure convergence through being able to define stable estimates only. This approach has been used in a new and practical estimation method which describes a system in terms of the polynomial roots. It is shown that the new scheme is as equally easy as Åström's, since the estimation cost and its derivatives can be calculated with a similar efficiency. The new method has been shown to satisfy Åström's theorems and proofs for consistency etc., without any great modification.

The effect of finite data lengths on the estimation procedure has been studied. Valid new criteria have been developed, which relate the length of a data sequence to the 'strength' of the poles. These criteria, which originate from the non-stationarity of filtered data, have been seen to be similar to those given by more heuristic reasoning. Several examples have shown that the new estimation method has a faster convergence in difficult situations than Åström's, and this is aided by using the above criteria for stopping tests.

1.4 Model Structure.

It is assumed that the process is described by a general discrete time state variable description given by (1.1). This model is linear, time invariant with stochastic disturbances and is assumed to be stable, controllable, and observable.

$$\begin{aligned} \underline{x}(k+1) &= F\underline{x}(k) + Gu(k) + \Gamma w(k) \\ y(k) &= Hx(k) + du(k) + v(k) + \varkappa \qquad k=1,2,\ldots\ldots \end{aligned} \tag{1.1}$$

In equation (1.1) $\underline{x}$ is an n vector of state variables, and in general there are r controls u, and m observations y. Thus matric~~i~~es H and G are m.n and n.r respectively. Since in this thesis we are principally considering a single input single output process, then the control u, observation y and disturbances v, w are scalars, with m=1, r=1.

Matric~~i~~es H and G reduce to 1.n and n.1 respectively, while F is square n.n and Γ is n.1. In general there will be a constant bias term $\varkappa$ in the observations y(k) representing a bias level in the measuring instrument. Both v and w are scalar random noise variables, each drawn at time k from a univariate normal distribution and have the following statistical characteristics.

$$E(w(k)) = 0. \tag{1.2}$$

$$E(w(k).w(k-i)) = \sigma_w^2.\delta(k-i) \tag{1.3}$$

$$E(v(k)) = 0. \tag{1.4}$$

$$E(v(k).v(k-i)) = \sigma_v^2.\delta(k-i) \tag{1.5}$$

$$E(v(k).w(k-i)) = \sigma_{vw}^2.\delta(k-i) \tag{1.6}$$

where E(.) = expected value of (.)

σ_v^2, σ_w^2 are the variances of the v and w sources.

$$\delta(k-i) \triangleq 1.0 \;;\; k-i = 0$$
$$\triangleq 0.0 \;;\; k-i \neq 0$$

If disturbances w(k) and v(k) were in fact zero, the system of (1.1) would be deterministic. Some of the following sections initially require this condition when developing the required transforms.

Frequently we might expect F to be partioned and G to have some zero elements such that there are some states in $\underline{x}$ which belong to a measurement noise or disturbance process which is distinct from the plant controlled by u(k). These extra states are then regarded as the noise states used to describe some coloured noise disturbance. Thus the assumptions are extended in that the state vector x is taken to be controllable from the inputs u(k) or from w(k) or from both.

The concepts of observability and controlability introduced by Kalman,[2,3,50,51] and others will be defined in the following manner. Assume the system of equation (1.1) is noise free and therefore deterministic. For a ~~sequence u(k),k=1, ... n and a~~ given initial condition on $\underline{x}$ at time k=1, the system is said to be controllable if state $\underline{x}$ can be changed from any initial condition to the origin of the state space of $\underline{x}$ in a finite time by applying input u(k),k=1,...n over this period n as shown in equations (1.7) to (1.10).

$$\underline{x}(2) = F\underline{x}(1) + Gu(1) \tag{1.7}$$

$$\underline{x}(3) = F\underline{x}(2) + Gu(2) = F^2\underline{x}(1)+FGu(1)+Gu(2) \tag{1.8}$$

$$\vdots$$

then $$\underline{x}(n+1)-F^n\underline{x}(1) = F^{n-1}Gu(1)+F^{n-2}Gu(2) \;\ldots\ldots\; +Gu(n) \tag{1.9}$$

$$= (G,FG, \;\ldots\ldots\; F^{n-1}G) \begin{bmatrix} u(n) \\ \vdots \\ u(1) \end{bmatrix} \tag{1.10}$$

As we were given $\underline{x}(1)$ and $\underline{x}(n+1)$ then the controls could be uniquely found only if $(G,FG, \ldots\ldots F^{n-1}G)$ had rank n and was therefore invertible. This result is more complicated for the case when $r>1$ as demonstrated by Luenberger[51] and by Rowe[16]. The array $(G,FG, \ldots\ldots F^{p-1}G)$ must have rank n, where p is a controllability index[51], $p \leqslant \min(n_m, n-r+1)$ and n_m is the degree of the minimal polynomial of matrix F. Since we are considering only single input single output systems at the moment, this consideration does not apply.

By an analogous approach to the controllability condition, a similar dual condition applies for the observability of the system. The state $\underline{x}$ at time k=1 can be determined uniquely by observing y(k) for a finite time if

$$(H^t,(HF)^t, \;\ldots\ldots\; (HF^{n-1})^t)^t \tag{1.11}$$

has rank n, and is therefore nonsingular. Again, for the situation where $m>1$, the power of F runs to p-1 and $p \leqslant \min(n_m, n-m+1)$ is now an index of observability similar to the case above.

1.5 Transformation to companion form.

A transform T, an n.n non-singular matrix with constant elements, can be used as an equivalence transform to change the co-ordinate set of the deterministic system to $\underline{x}'=T\underline{x}$, where the prime symbols refer to the new system. The system matric~~i~~es are then given by (1.12).

$$F' = TFT^{-1} \qquad G' = TG \qquad H' = HT^{-1} \qquad (1.12)$$

An equivalence transform has the property that for the same set of inputs u, the original system and the new system will both give the same outputs y, as in (1.13), for appropriate initial conditions. This has been more formally given by Athans and Falb[53].

$$y' = H'\underline{x}' = HT^{-1}T\underline{x} = y \qquad (1.13)$$

If T is defined as $(q^t,(qF)^t \;.....\; (qF^{n-1})^t)$, where q is an arbitrary vector which satisfies T having rank n, then the system F' can be seen to reduce to the normal companion form[51], with the states x_i referred to as phase variables[20].

$$F' = \begin{bmatrix} q \\ qF \\ \cdot \\ \cdot \\ \cdot \\ qF^{n-1} \end{bmatrix} F \begin{bmatrix} q \\ qF \\ \cdot \\ \cdot \\ \cdot \\ qF^{n-1} \end{bmatrix}^{-1} = \begin{bmatrix} q \\ qF \\ \cdot \\ \cdot \\ qF^{n-2} \\ qF^{n-1} \end{bmatrix} \begin{bmatrix} qF^{-1} \\ q \\ \cdot \\ \cdot \\ \cdot \\ qF^{n-2} \end{bmatrix}^{-1} \triangleq (M)(J)^{-1} \qquad (1.14)$$

Because of the particular structure, i^{th} row of M equal to the $(i+1)^{th}$ row of J, and J non-singular, it can be shown[16] by postmultiplying both sides of (1.14) by J, that the $(i+1,i)^{th}$ element of F' is 1.0, while the rest of the F_{ij} elements are zero, for i=1, ... n, j=1, ... n-1. Thus F' takes the form of (1.15), which is the normal companion form.

$$F' = \begin{bmatrix} 0 & 1 & 0 & 0 & .. & .. \\ 0 & 0 & 1 & 0 & . & . \\ 0 & 0 & 0 & 1 & & \\ . & . & . & & & \\ . & . & . & 0 & 0 & 1 \\ -a_n & -a_{n-1} & \cdots & \cdots & \cdots & -a_1 \end{bmatrix} \qquad (1.15)$$

The a_i terms, i=1, ... n, have yet to be determined but will satisfy the characteristic polynomial of both F and F' since this is a criterion for their similarity[52]. The terms are therefore unique and independent of the choice of the n-vector q.

When q is chosen to be the vector H for the scalar observation case studied, the transforming matrix T is the observability matrix in equation (1.11). The new system matrix H',1.n, also has a particular form when T is chosen in this manner, and is given by (1.16). Equation (1.16) may be compared with (1.14) and it can be seen that M' is now 1.n, while J' is still n.n and non-singular.

$$H' = HT^{-1} = H \begin{bmatrix} H \\ HF \\ \cdot \\ \cdot \\ \cdot \\ HF^{n-1} \end{bmatrix}^{-1} = HF^{-1} \begin{bmatrix} HF^{-1} \\ H \\ \cdot \\ \cdot \\ \cdot \\ HF^{n-2} \end{bmatrix}^{-1} = (M')(J')^{-1} \qquad (1.16)$$

By a very similar argument to (1.15) H' reduces to (1, 0, 0, ..., 0)

(1.17)

The new matrix G' = TG does not show any special form. A precisely dual transformation derived from controllability conditions can also be used on F to give the transposed canonical form for F, and a simple form for G'.

1.6 Transfer function description.

We propose here a transfer function description of the system as a rational Z polynomial between the input variables u(k) and the output variables y(k).

$$y(k)/u(k) = B(z)/A(z) \qquad (1.18)$$

$$\text{or } y(k)+a_1y(k-1).....+a_ny(k-n) = b_1u(k-1)+b_2u(k-2).....b_nu(k-n)$$

(1.19)

$$\text{Thus } A(z^{-1}) = 1.0 + a_1z^{-1} + a_2z^{-2}..... a_nz^{-n} \qquad (1.20)$$

$$B(z^{-1}) = b_1z^{-1} + b_2z^{-2}..... b_nz^{-n} \qquad (1.21)$$

The direct control term $b_ou(k)$ and the bias term χ in y(k) have not been included at this point. The operator z is the unit advance operator in discrete time and (1.18) describes the transfer function as a rational z polynomial. We will now show that the description (1.18) can be transformed to give the companion form description of equations (1.15) and (1.17), and that the two descriptions are therefore equivalent to the deterministic version of (1.1) with d=0.0. From (1.1) and using the companion form of F as in (1.15)

$$\begin{aligned} x_1(k) &= x_2(k-1)+g_1u(k-1) \\ x_2(k) &= x_3(k-1)+g_2u(k-1) \\ &\vdots \\ x_n(k) &= -a_nx_1(k-1) \; \ldots\ldots \; -a_1x_n(k-1)+g_nu(k-1) \end{aligned} \qquad (1.22)$$

where g_i are the elements of G' defined by (1.12)

Due to the form of (1.17)

$$y(k) = x_1(k) \qquad (1.23)$$

The set of equations in (1.22) can be re-arranged and the identity of (1.23) used to give the set (1.24)

$$\begin{aligned} x_2(k-1) &= x_1(k)-g_1u(k-1) = y(k)-g_1u(k-1) \\ x_3(k-1) &= x_2(k)-g_2u(k-1) = y(k+1)-g_1u(k)-g_2u(k-1) \\ &\vdots \\ x_n(k-1) &= x_{n-1}(k)-g_{n-1}u(k-1) = y(k+n-2)-g_1u(k+n-3)\ldots\ldots-g_{n-1}u(k-1) \end{aligned}$$

(1.24)

It follows by changing the time index k to k+1 that

$$x_n(k) = y(k+n-1)-g_1u(k+n-2) \;\;.....\;\; -g_{n-1}u(k) \tag{1.25}$$

Equation (1.25) can now be used in (1.22) to give

$$\begin{aligned} y(k+n-1)&-g_1u(k+n-2)-g_2u(k+n-3) \;\;.....\;\; -g_{n-1}u(k) \\ &= -a_nx_1(k-1)-a_{n-1}x_2(k-1) \;\;.....\;\; -a_1x_{n-1}(k-1)+g_nu(k-1) \\ &= -a_ny(k-1)-a_{n-1}\left[y(k)-g_1u(k-1)\right] \;\;..... \\ &\qquad -a_1\left[y(k+n-3)-g_1u(k+n-4) \;\;.....\;\; -g_{n-1}u(k-2)\right]+g_nu(k-1) \end{aligned} \tag{1.26}$$

Equation (1.26) can be re-arranged and the time index shifted again to give (1.27)

$$\begin{aligned} y(k)&+a_1y(k-1)+ \;\;.....\;\; a_{n-1}y(k-n+1)+a_ny(k-n) \\ &= g_1u(k-1)+g_2u(k-2) \;\;.....\;\; g_nu(k-n) \\ &\quad +a_{n-1}g_1u(k-n)+ \;\;.....\;\; +a_1\left[g_1u(k-2)+g_2u(k-3) \;\;.....\;\; g_{n-1}u(k-n)\right] \end{aligned} \tag{1.27}$$

$$= b_1u(k-1)+b_2u(k-2)+ \;\;.....\;\; b_nu(k-2) \tag{1.28}$$

By equating co-efficients in (1.27) and (1.28) then

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \,.. & 0 & 0 \\ a_1 & 1 & 0 \,.. & 0 & 0 \\ a_2 & a_1 & 1 \,.. & 0 & 0 \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ a_{n-1} & a_{n-2} & .. & a_1 & 1 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \\ \cdot \\ \cdot \\ \cdot \\ g_n \end{bmatrix} \qquad (1.29)$$

The equivalance of the transfer function description to the companion form can now be seen since equation (1.27) is identical to (1.19). There are also n initial conditions to be set on the y(k) sequence, $k = -n+1, -n+2, \ldots\ldots, -1, 0$ before the system is released at time $k=1$, and correspond to the n initial conditions on the state vector x at $k=1$.

The transfer function description is only slightly changed for the case where the co-efficient d in (1.1) is non-zero. This allows a direct connection between system input u and output y. When the co-ordinates of the state x are changed, as in (1.12), the term $d_*u(k)$ appears without modification in equations (1.22) to (1.29) which can be reworked to give (1.30).

$$\begin{bmatrix} b_o \\ b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \,\ldots\ldots & 0 & 0 \\ a_1 & 1 & 0 \,\ldots\ldots & 0 & 0 \\ a_2 & a_1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ a_n & a_{n-1} & a_{n-2}\ldots & a_1 & 1 \end{bmatrix} \begin{bmatrix} d \\ g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix} \qquad (1.30)$$

Thus the polynomial B(z) has been extended by the term b_oz^o and therefore now has a total of (n+1) coefficients. Such a system may well exist in discrete time, and yet have no physical meaning in a continuous time system. This is because most physical plant will take at least some small time to respond to a control input u(k).

1.7 Number of parameters.

Rowe[16] has developed the arguments of the previous two sections to the multivariate input output case where $m>1$ and $r>1$. The transformation matrix T corresponding to that in section 1.5 is then not unique as H is no longer a vector. The requirement for T to have rank n can be met by several selections of n linearly independant rows from the rows of the complete array HF^i, i=0,1, ... p; $p\leq\min(n_m,n-m+1$ Thus there are as many difference equation transfer function description like (1.28) as there are ways in selection the rows to make T. However, for the case m=1, r=1 the matrix F must have $n_m=n$ in order to be observable and controllable from the scalar input and output. The transformation T becomes then unique as (1.11), and the transfer function description is also unique.

The original system (1.1), without direct control via d, could be specified by n^2 parameters in F, nm in H, and nr in G. By employing the transformation T derived from the observability array, a new system can be obtained with F' having n_m, and G having nr parameters. Thus the total number of parameters is n(m+r) and this can be shown[50,16] to be the minimum necessary to describe the system. The original system could also have had the number of parameters n(m+r) if it was already the minimal form with some zero elements. There are other transformations T which will provide a minimal form for the system; for example a dual approach using the controllability conditions can be used to derive a T_2 with similar results. An extra number of parameters mr should be included with the minimal form for the case where direct control of the output is allowed i.e. for d in (1.1).

For the case studied m=1, r=1, the minimal number of parameters is 2n = n(m+r) and the transfer function description (1.19), (1.28) is unique and also has 2n parameters. Two extra parameters are required to describe the d or b_o term when present and the constant bias term γ on y_k. There are naturally n initial conditions also to be inserted into the sequence y(k) for complete identity with system (1.1) to be achieved. Thus there is no loss and much to be gained if the transfer function description is easier[10,15] to estimate than other descriptions. Other forms may be derived if required after the estimation procedure has been completed.

1.8 Stochastic system description.

We will now consider the full case of system (1.1) including all the noise terms. As mentioned in section 1.4 only some of the states in the model are under the direct influence of the control u. The estimation procedure can only be concerned with the states that are both observable and controllable from u and — or the disturbance. Rowe[16] shows that it is necessary to assume the system is observable, controllable by the control and noise inputs together, and output controllable in the mean by the control input u. The observability conditions remain as for the deterministic case.

The system is controllable by the inputs u and w if the array

$$((G,\Gamma),F(G,\Gamma),F^2(G,\Gamma), \ldots\ldots ,F^{p-1}(G,\Gamma)) \qquad (1.31)$$

has rank n, where the controllability index $p \leqslant \min(n_m, n-r-j+1)$ and Γ is (n.j). For the system studied in this thesis w(k) is a scalar and thus j=1. This statement is similar to (1.10), and arises because the image of the vector space of dimension (r+j) spans the space of $\underline{x}$ when the array (1.31) is of rank n. If the system is output controllabl in the mean sense, it implies that the expected value of y(k),k≥p, can be reached from arbitary initial conditions. This condition requires[16] array (1.32) to have rank m for the least positive integer $p \leqslant \min(n_m, n-r+1)$

$$(HG,HFG, \ldots\ldots ,HF^{p-1}G,D) \qquad (1.32)$$

1.9 Stochastic Difference equation.

The transformation methods used to derive a difference equation in the deterministic case can also be applied to the full stochastic system (1.1). The difference equation takes the form of (1.33), where in general A,B,K and L are matrix polynominals, but reduce to scalar polynomials for this study.

$$A(z^{-1})y(k) = B(z^{-1})u(k)+K(z^{-1})w(k)+L(z^{-1})v(k) \qquad (1.33)$$

The procedure for obtaining A and B given system (1.1) is not changed by the inclusion of stochastic disturbances[16]. The co-efficients of $K(z^{-1})$ can be found in the same way as (1.29).

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \\ \cdot \\ \cdot \\ \cdot \\ K_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \cdots\cdots & 0 \\ a_1 & 1 & 0 \cdots\cdots & 0 \\ a_2 & a_1 & 1 & \cdot \\ & & & \cdot \\ & & & \cdot \\ & & & \cdot \\ a_{n-1} & a_{n-2} & \cdots\cdots a_1 & 1 \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \Gamma_3 \\ \cdot \\ \cdot \\ \cdot \\ \Gamma_n \end{bmatrix}$$

where $K(z^{-1}) = K_1z^{-1}+K_2z^{-2} \cdots\cdots K_nz^{-n}$

It can be shown[16] that in fact $L(z^{-1})$ is equal to $A(z^{-1})$. The total scalar disturbance V(k) to the difference equation is the sum of the extra components in (1.33).

$$V(k) = K(z^{-1})w(k)+L(z^{-1})v(k)$$
$$= (0\ 1)\begin{bmatrix} w(k) \\ v(k) \end{bmatrix} + (K_1\ a_1)\begin{bmatrix} w(k-1) \\ v(k-1) \end{bmatrix} \ \ldots\ldots\ (K_n a_n)\begin{bmatrix} w(k-n) \\ v(k-n) \end{bmatrix} \qquad (1.34)$$

The serial auto-correlations ϕ_τ of V(k) are defined by (1.35), and will be scalar for the m=1 case.

$$\phi_\tau \triangleq E\ (V(k).V(k-\tau)\ )\ , \quad \tau=0,\pm1,\pm2\ \ldots\ldots \qquad (1.35)$$

There are in the general case m(m+1)/2 parameters in ϕ_o, and m^2 in ϕ_τ, $\tau=\pm1,\pm2$, ... $\pm n$ since ϕ_o is symetric. When m=1 there are n+1 parameters in total needed to describe ϕ_τ. Since the sequences v(k) and w(k) have the independance properties as expressed in (1.2) to (1.6), ϕ_τ is zero for $/\tau/>n$.

1.10 Statistical Equivalence of processes.

The sequence V(k) defined by (1.34) has a normal distribution because it is formed as a linear sum of the random variables w(k) and v(k) which are themselves normally distributed. Thus the Gaussian process V(k) is defined completely by its zero first moment and the (n+1) second moments from (1.35).

Another scalar process V'(k) is proposed by the definition in (1.36), which has the same number of degrees of freedom as V(k) in the choice of its parameters.

$$V'(k) \triangleq (1.0+c_1z^{-1}+c_2z^{-2} \ldots\ldots c_nz^{-n})\lambda e(k) = C(z^{-1})\lambda e(k) \qquad (1.36)$$

where $E(e(k)) = 0$.

$E(e(k).e(k-i)) = 1.0\delta(k-i)$, and $e(k)$ is normally distributed.

This new process $V'(k)$ is also Gaussian with zero first moment, and $(n+1)$ non-zero second moments ϕ'_τ given in (1.37).

$$\phi'_\tau = E\ (V'(k).V'(k-\tau)); \tau=0,\pm1,\pm2,\ \ldots\ldots\ \pm n \qquad (1.37)$$

For the general case[16], each of the co-efficients in (1.36) is an m.n matrix and λ is replaced by a symetric matrix $\Lambda^{\frac{1}{2}}$ with $m(m+1)/2$ parameters. It is claimed that the two processes $V(k)$ and $V'(k)$ can be statistically equivalent when $E(V(k)) = E(V'(k))$, and $\phi_\tau = \phi'_\tau$, $\tau=0,\pm1,\pm2,\ \ldots\ldots\ \pm n$. Later work in this thesis will be concerned with representing the process (1.1) by the difference equation (1.38) which employs the $C(z^{-1})$ description in place of that used in (1.33).

$$A(z^{-1})y(k) = B(z^{-1})u(k)+C(z^{-1})\lambda e(k)+\mathcal{H}'$$

where $A(z^{-1}) = 1 + a_1z^{-1}\ \ldots\ldots\ a_nz^{-n}$;

$B(z^{-1}) = b_o+b_1z^{-1}\ \ldots\ldots\ b_nz^{-n}$;

$C(z^{-1}) = 1 +c_1z^{-1}\ \ldots\ldots\ c_nz^{-n}$;

$\mathcal{H}' = \mathcal{H}*\ (1+a_1+a_2\ \ldots\ldots + a_n)$

for convenience in notation the polynomials can be multiplied throughout by z^n to give

$$A(z)y(k) = B(z)u(k) + C(z)\lambda e(k) + \mathcal{H}' \qquad (1.38)$$

The problem of finding the parameter values for the V'(k) process given only ϕ'_τ equal to ϕ_τ is the problem of spectral factorization. This is not easy as there are n simultaneous equations, $m(m+1)/2+pm^2$ in the general case obtained by equating moments, which are non-linear and their solution is not unique. Some solutions will have unstable roots and must be discarded. Other workers have studied these problems[16,54-58], and given more formal and general accounts. The conditions under which factorization is possible are given in (1.39)

$\phi(z) = \sum_{i=-\infty}^{+\infty} \phi_i z^i$ is a rational function in z with

$$\begin{aligned} &\text{i)} \quad \phi(z) = \phi(z^{-1}) \\ &\text{ii)} \quad \phi_0 \geqslant 0. \end{aligned} \tag{1.39}$$

We maintain that it is easier to estimate a process with the description (1.38), with the minimal number of parameters, than the original model (1.1). This has been noted by Åström[37] and Mayne[15] in that it is far simpler to construct a likelihood function which avoids introducing unknown state variables.

There are 4n+3 parameters in (1.38) needed to define the system, which is the same as the minimal number required to describe (1.1). These are made up from n in A(z), n+1 in B(z), n in C(z), with two more used for λ and μ'. The other n parameters belong to initial conditions. It will be seen later that for long data sequences y(k), u(k) it is possible to ignore the initial conditions y(k),k= -n+1, -n+2, ... ,-1,0 on the model at time k=1, which correspond to the initial conditions on the state vector at k=1. Any initial conditions are regarded as having decayed within the data length to be insignificant compared to the stochastic disturbances. This leaves only 3n+3 parameters to be considered for estimation.

Once an estimation procedure has been completed using description in (1.38), it is possible to obtain a st·.te representation by reversing the transformations described. This has been studied at length by Kalman[39], Dewey[40], and Mayne[38] among others, for the minimal realisation condition. Such a condition is required to make the solution unique, because there are a large number of possible state realisations. For example, the minimal realisation of the deterministic scalar input-output system of (1.1) has only 2n parameters in (1.15) and (1.17). The original description required (n^2+2n) parameters.

It should be clear that having estimated a transfer function as (1.38), we cannot derive a state variable description that includes both v(k) and w(k). This follows from the above section in that we are unable to distinguish between two independent Gaussian noise sources and must classify them as one. Such a state variable description is given by (1.40), and in more detail by Rowe[16].

$$
\begin{aligned}
\underline{x}(k+1) &= F'\underline{x}(k)+G'u(k)+\Gamma'e(k) \\
y(k) &= H'x(k)+Du(k)+\lambda e(k)
\end{aligned}
\qquad (1.40)
$$

1.11 Deciding the order of the model.

If we were provided with a data record $y(k),u(k);k=1, \ldots\ldots N$ of a plant, then an estimate of its parameters could be obtained only after the structure (1.1) or (1.38) had been chosen. In particular we would have to select a value of n, the state dimensionality, consistent with any prior knowledge of the plant and later usage of the model. This is naturally a higher level procedure than the simple parameter estimation problem. Many people have studied this area[13,16,12,43,44] without any definitive answer being found for determining the model order n. The best engineering approach seems to be that of starting with n=1 and increasing it by one after having estimated the corresponding parameters. A close watch is kept on some index of performance and the confidence which can be placed on individual estimates. The whole procedure appears to be akin to hypothesis testing and can only provide results in a probabalistic sense. It is expected that some lower plateau of performance would be reached for values of n equal to or greater than a value n^*, and that poorer confidence levels would be associated with the estimates for $n>n^*$. The value of n^* and the parameter estimates for the n^{*th} order model are then adopted as a satisfactory solution to the whole estimation problem. An example of this procedure is given in Chapter 6.

CHAPTER 2

PARAMETER ESTIMATION

2.1 Properties of Estimators.

It is well known that field trials with physical plant are arduous, and that data collection from such a system is expensive. We should therefore be interested in making optimal use of the data when estimating plant parameters. These estimates would probably be used to effect a low mean square prediction error of the plant output by being used in a stochastic regulator. This prediction error may be more important in practice than the individual parameter errors obtained in estimation. A situation where this occurs is shown in section 2.2.

Statisticians[21,23] can present an argument based on experience as far back as K.F. Gauss, 1809, that a quadratic loss function of the form (2.1) is quite realistic for many parameter estimation problems, and is often chosen for mathematical convenience.

Loss = Function $\left[(\theta-\hat{\theta})^t(\theta-\hat{\theta})\right]$

where $\underline{\hat{\theta}}$ is the estimate of parameters $\underline{\theta}$ and satisfies (i) Loss$\geqslant$0. for all permitted $\underline{\theta}$ and $\underline{\hat{\theta}}$,(ii) There is one $\underline{\hat{\theta}}$ for each $\underline{\theta}$ value for which Loss = 0. (2.1)

The value of risk $R(\underline{\theta})$, defined as the expectation of the random loss (2.1) for a given estimation method, can be compared for various methods. The estimation method that gives the minimum risk $R_{min}(\underline{\theta})$, will minimise $E\left[(\underline{\theta}-\underline{\hat{\theta}})^t(\underline{\theta}-\underline{\hat{\theta}})\right]$, the mean squared error. However this method may only produce a minimum mean squared error for some values of $\underline{\theta}$. Thus since $\underline{\theta}$ is unknown, the choice of estimation method may well contain an arbitary element.

Other than the above criteria, estimation methods can be classified broadly according to their following basic properties.

Unbiased estimate: Expected value of the estimate $\hat{\underline{\theta}}$ is equal to the true value $\underline{\theta}$. $E(\hat{\underline{\theta}}) = \underline{\theta}$ (2.2)

Consistent estimate: Let $\hat{\underline{\theta}}_1 \,.....\, \hat{\underline{\theta}}_n$ be a sequence of estimates of $\underline{\theta}$ for increasingly larger sets of data. This sequence is then a squared error consistent[23] estimate of $\underline{\theta}$ if

$$Lt_{n\to\infty}\, E((\underline{\theta}-\hat{\underline{\theta}}_n)^t(\underline{\theta}-\hat{\underline{\theta}}_n)) = 0. \quad \text{for all } \underline{\theta} \qquad (2.3)$$

Since $R_{min}(\underline{\theta}) = E((\underline{\theta}-\hat{\underline{\theta}}_n)^t(\theta-\underline{\theta}_n))$ it follows that the above condition implies that both the bias and variance of $\hat{\underline{\theta}}_n$ approach zero.

Efficient estimate: If $\hat{\underline{\theta}}$ is an unbiased estimate of $\underline{\theta}$ and has a finite loss, and no other unbiased estimate has a smaller loss, then $\hat{\underline{\theta}}$ is an efficient estimate of $\underline{\theta}$. (2.4)

Mood and Greybill[23] point out that there are estimators which are only efficient in a limiting sense for very large data sets. i.e. asymptotically efficient. Also since estimators with minimum mean-squared error rarely exist for all $\underline{\theta}$ values, a reasonable procedure is to restrict the class of estimating functions to unbiased estimators and see if an estimator with minimum mean squared error can be found. This occurs much more frequently than the existance of a general minimum mean-square estimator.

Minimum variance unbiased estimate: If $\hat{\underline{\theta}} = g(X_1, \ldots\ldots X_n)$ is an estimate of $\underline{\theta}$ based on data $X_1 \ldots\ldots X_n$ drawn from a distribution $f(X,\underline{\theta})$, then $\hat{\underline{\theta}}$ is a minimum variance estimate, provided that:

(i) $E\,(\hat{\underline{\theta}}) = \underline{\theta}$ (2.5)

(ii) Covariance $(\hat{\underline{\theta}})$ is less than the covariance of any other estimate satisfying condition (i) (2.6)

2.2 Least squares estimator.

This method of estimation is based on theory presented in a large number of text books, but which was first introduced by Gauss, and has since had very little development. The transfer function description (1.38) is re-arranged to give a mixed autoregressive-moving average model (2.7). As will be seen later the bias term $\varkappa'$ can be dropped from the model as it can be separately estimated. The notation y(k) has been changed here, and in succeeding sections to y_k, similarly u(k) becomes u_k.

$$y_k = -a_1 y_{k-1} \;\cdots\cdots\; -a_n y_{k-n} + b_o u_k \;\cdots\cdots\; + b_n u_{k-n} + e_k + c_1 e_{k-1} \;\cdots\cdots\; + c_n e_{k-n}$$

$$= m_{1,k}\theta_1 + m_{2,k}\theta_2 \;\cdots\cdots\; m_{q,n}\theta_q + v_k$$

$$= \underline{m}_k^t . \underline{\theta} + v_k \quad ; \quad k=1, \;\cdots\cdots\; N$$

$$\text{where } q = 2n+1 \quad ; \underline{\theta}^t \triangleq (-a_1,-a_2 \;\cdots\cdots\; -a_n,b_o,b_1, \;\cdots\cdots\; b_n) \;;\; 1.q$$

$$\underline{m}_k^t \triangleq (y_{k-1},y_{k-2}, \;\cdots\cdots\; y_{k-n},u_k, \;\cdots\cdots\; u_{k-n}); \; 1.q \qquad (2.7)$$

u_k is bounded and has finite 1st and 2nd moments and

v_k is a zero mean coloured noise sequence defined by

$$v_k = e_k + c_1 e_{k-1} \;\cdots\cdots\; c_n e_{k-n} \qquad (2.8)$$

A *loss* function (2.9) is now defined and can be regarded for this work as quadratically costing the prediction error v_k between y_k and the predicted output $\underline{m}_k^t \underline{\theta}$ for an estimate $\underline{\theta}$ of $\underline{\theta}_o$.

$$R(\underline{\theta}) = \sum_{k=1}^{N} (y_k - \underline{m}_k^t \underline{\theta})^2 = (Y-M\underline{\theta})^t(Y-M\underline{\theta})$$

where N is the length of the data sequence y_k,u_k

$$\text{and } Y^t \triangleq (y_1,y_2, \;\cdots\cdots\; y_N) \;;\; 1.N \;;\; M^t \triangleq (\underline{m}_1,\underline{m}_2, \;\cdots\cdots\; \underline{m}_N) \;;\; q.N \qquad (2.9)$$

$R(\hat{\underline{\theta}})$ ~~does not have the expectation form defined in section 2.1; but it~~ can ~~still~~ be minimised by setting the derivatives with respect to $\hat{\underline{\theta}}$ equal to zero to discover the resulting conditions.

$$\frac{\partial R(\underline{\theta})}{\partial \underline{\theta}} = -2M^tY+2M^tM\underline{\theta} = 0. \tag{2.10}$$

The least squares estimate is thus given by

$$\hat{\underline{\theta}} = (M^tM)^{-1}M^tY \quad ; \quad q.1 \tag{2.11}$$

The risk $R(\underline{\theta})$ can be written as a squared norm of a vector as in(2.12)

$$R(\underline{\theta}) = \|Y-M\theta\|^2 = \|Y\|^2-Y^tM\underline{\theta}-\underline{\theta}^tM^tY + \underline{\theta}^tM^tM\underline{\theta} \tag{2.12}$$

If$\overset{t}{M}Y$ were defined from (2.11) as $\quad \overset{t}{M}Y = \overset{t}{M}M\hat{\underline{\theta}}$ (2.13)

Then (2.12) becomes $R(\underline{\theta}) = \|Y\|^2+(\underline{\theta}-\hat{\underline{\theta}})^tM^tM(\underline{\theta}-\hat{\underline{\theta}})-\hat{\underline{\theta}}^tM^tM\hat{\underline{\theta}}$

$$= \|Y\|^2 + \|M(\underline{\theta}-\hat{\underline{\theta}})\|^2-\|M\hat{\underline{\theta}}\|^2 \tag{2.14}$$

$$= \|Y-M\theta\|^2 \geqslant \|Y\|^2-\|M\hat{\theta}\|^2 \tag{2.15}$$

since each term (2.14) is positive

Now $R(\underline{\theta})$ in (2.14) will be minimum iff $\underline{\theta} = \hat{\underline{\theta}}$, where $\hat{\underline{\theta}}$ is given by(2.11). Also (2.13) is another form of (2.11), and hence it has been shown[21] that (2.11) gives an absolute minimum for risk $R(\underline{\theta})$ given the structure (2.7) and (2.9).

For the model as given by (2.7), the estimate $\hat{\underline{\theta}}$ can be shown to be biased.

$$E(\hat{\underline{\theta}}-\underline{\theta}) = E((M^tM)^{-1}M^tY-\underline{\theta}) = E((M^tM)^{-1}M^tM\underline{\theta}+(M^tM)^{-1}M^tV-\underline{\theta})$$

$$= E((M^tM)^{-1}M^tV) \tag{2.16}$$

If the terms M and V in (2.16) were independant then (2.16) would reduce to $E((M^tM)^{-1}M^t) .E(V)$ which is zero since V has zero mean. However the matrix M includes $\underline{m}_k^t$ as the k^{th} row and $\underline{m}_k^t$ contains elements $y_{k-1}, y_{k-2}, \ldots\ldots y_{k-n}$, and these elements must be related to $v_{k-1}, v_{k-2}, \ldots\ldots$ through (2.7), since v_k is correlated over n delays in definition (2.8). Thus the bias term (2.16) is not zero, and the least squares estimator is biased for estimates of the co-efficients of $y_{k-1}, \ldots\ldots y_{k-n}$, but unbiased for $u_k, \ldots\ldots u_{k-n}$ co-efficients if the u_k and v_k sequences are uncorrelated.

The biased estimate might be used in practise to predict y_k in a controller. The value of the sum of the prediction error squared, i.e. the value of $R(\underline{\theta})$, can be obtained by substituting (2.11) in (2.9), and (2.7) for the value of Y. The value of the bias on $\underline{\hat{\theta}}$ is taken from (2.16)

$$\begin{aligned} R(\underline{\hat{\theta}}_{biased}) &= (M\underline{\theta}_{exact}+V-M\underline{\hat{\theta}}_{biased})^t(M\underline{\theta}_{exact}+V-M\underline{\hat{\theta}}_{biased}) \quad (2.17)\\ &= (M\underline{\theta}_{exact}+V-M\underline{\theta}_{exact}-M(M^tM)^{-1}M^tV)^t \\ &\qquad (M\underline{\theta}_{exact}+V-M\underline{\theta}_{exact}-M(M^tM)^{-1}M^tV) \end{aligned}$$

$$\therefore R(\underline{\hat{\theta}}_{biased}) = V^t(I-M(M^tM)^{-1}M)^t(I-M(M^tM)^{-1}M)V \qquad (2.18)$$

If instead of the biased $\underline{\hat{\theta}}$ we had used the exact value of $\underline{\theta}$, then the value of $R(\underline{\theta})$ would have been

$$R(\underline{\theta}_{exact})=(M\underline{\theta}_{exact}+V-M\hat{\underline{\theta}}_{exact})^t(M\underline{\theta}_{exact}+V-M\underline{\theta}_{exact}) = V^tV \tag{2.19}$$

By the argument of (2.12) to (2.15), $R(\hat{\underline{\theta}}_{biased})$ is less than $R(\hat{\underline{\theta}}_{exact})$ and we would expect the prediction performance to be improved by using the biased estimate. This is because the biased estimates have partially absorbed the effect of the c_i co-efficients and thus offset some of the colouration in v_k. Should the case arise that c_i,i=1, ... n is zero in (2.8), then v_k would be an independant sequence given by e_k. The data y_k, u_k at each time k would now be independant of all v_k and the bias term (2.16) would be zero. There would then be no distinction in performance between (2.18) and (2.19).

The co-variance matrix of the least squared estimate can now be calculated in (2.20)

$$\begin{aligned}
\text{cov.}(\tilde{\underline{\theta}})=\Psi &= E((\underline{\theta}-\hat{\underline{\theta}})(\underline{\theta}-\hat{\underline{\theta}})^t) \qquad (2.20)\\
&= E(((M^tM)^{-1}M^tY-\underline{\theta})((M^tM)^{-1}M^tY-\underline{\theta})^t)\\
&= E((M^tM)^{-1}M^tM\underline{\theta}+(M^tM)^{-1}M^tV-\underline{\theta})((M^tM)^{-1}M^tM\underline{\theta}+(M^tM)^{-1}M^tV-\underline{\theta})^t\\
&= E\left[(M^tM)^{-1}M^tV)((M^tM)^{-1}M^tV)^t\right]
\end{aligned}$$

$$\therefore \Psi=(M^tM)^{-1}M^tE\left[VV^t\right]M(M^tM)^{-1} \tag{2.21}$$

where $V^t \triangleq (v_1,v_2, \ldots\ldots v_N)$; 1.N

$$\text{If} \quad E(VV^t) = \sigma_e^2 * I_N \tag{2.22}$$

then equation (2.21) be simplified to give the minimum variance property[23], described in section 2.4.

$$\mathrm{Cov}\ (\hat{\underline{\theta}}) = \Psi = \sigma_e^2.(M^tM)^{-1} \qquad (2.23)$$

This again requires the sequence v_k to be independant and thus all $c_i, i=1, \ldots\ldots n$ to be zero, i.e. the disturbance on the model (2.7) to be white.

R.C.K. Lee[19] describes the least squares algorithm above in a recursive manner. Assume a solution to (2.11) has been obtained for a data set $k=1, \ldots\ldots N$; $N>q$ and more measurements are taken at N+1 to give $\underline{m}_{N+1}$. The matrix M defined in (2.9) will now have $\underline{m}^t_{N+1}$ as an extra row, and the vector Y an extra element y_{N+1}. The new solution to (2.11) is now given by (2.24)

$$\hat{\underline{\theta}}_{N+1} = (M^t_{N+1}\ M_{N+1})^{-1} M^t_{N+1}\ Y_{N+1} \qquad (2.24)$$

The inverse of the matrix required in (2.24) can be efficiently obtained using the matrix inversion lemma[19] as shown in (2.25)

$$P_{N+1} = (M^t_{N+1}\ M_{N+1})^{-1} = (M^t_N M_N + \underline{m}_{N+1}\underline{m}^t_{N+1})^{-1}$$

$$\therefore\ P^{-1}_{N+1} = P^{-1}_N + \underline{m}_{N+1}\ \underline{m}^t_{N+1}$$

then

$$P_{N+1} = P_n - P_N\underline{m}_{N+1}\ (\underline{m}^t_{N+1}\ P_n\underline{m}_{N+1} + 1)^{-1}\ \underline{m}^t_{N+1}P_N \qquad (2.25)$$

The only inversion now required is a scalar which is computationally useful if n is large. The formula for least squares recursive estimation can now be given by (2.26)

$$\underline{\hat{\theta}}_{N+1} = \underline{\hat{\theta}}_N + P_N\underline{m}(\underline{m}^t P_N \underline{m}+1)^{-1}(y_{N+1} - \underline{m}^t \underline{\hat{\theta}}_N) \qquad (2.26)$$

where $\underline{m} = \underline{m}_{N+1}$

The old estimate for $\underline{\theta}$ has been updated by a correction term based on the new data and the old value of P_N. The algorithm,using equations (2.25) and (2.26) recursively, after a minimal data set $N = q$ has been obtained, is a form of the Kalman sequential estimating method[2,3].

Any estimate obtained from these equations will be identical for the same data length N to that of (2.11), and will therefore share the same faults. Thus the estimate $\underline{\hat{\theta}}$ will still be biased due to the correlation of y_k and v_k. Lee[19] shows that the matrix P_N always decreases and in the limit approaches the null matrix when $N \to \infty$, independant of the assumed initial conditions. He concludes that for the condition v_k is white, and P_N is therefore proportional to the covariance matrix (2.21), the estimate $\underline{\hat{\theta}}$ is statistically consistent. When v_k is coloured as in the general case from (2.8), the correct conclusion must be that the least squares estimate is i) biased; ii) not consistent, due to its bias; iii) not efficient, since there do exist more efficient estimators[23,21] as shown later in section 2.4.

2.3 Methods of avoiding bias.

Various devices have been suggested to remove the bias on the estimate of $\underline{\hat{\theta}}$ in (2.11). Lee updates the algorithm (2.25),(2.26) only every n data points, and thus trys to avoid the bias effect, since $\underline{m}_{k-n-1}$ cannot be correlated with v_k which contains terms only as far back as e_{k-n}. Some bias remains as y_{k-1}, y_{k-2} is affected by all the past values of e_k. However the matrix M^tM in (2.11) tends to be nearly singular and difficult to invert. There is also a large wastage of data with this approach which might be used in a more optimal manner. Thus the covariance of the estimates with Lee's method must be worse than the simple method of (2.11).

Mayne[26] and Tzafestas[27] have used an estimate $\underline{m}_k$ in equation (2.11) derived from data at time k-n-1 and before, so that the elements y_{k-1}, ... y_{k-n} are un-correlated with v_k.

The vector $\underline{m}_k$ is provided by a linear regression estimator which may in turn be biased, but of which the only important property is that of prediction. The two estimators, for $\underline{m}_k$ and then for $\underline{\Theta}$ can be updated together in a recursive manner. The estimate for $\underline{\Theta}$ can be shown[26] to be asymptotically unbiased and consistent, but lacks efficiency. Rucker[60] and Levadi[61] have also developed a method which first estimates y_k and m_k assuming a noise free model, and then applies a sequential least squares algorithm. It is claimed that the estimate of $\underline{\Theta}$ is unbiased and consistent but not minimum variance. Rowe[16] has given a "bootstrap estimator" which is similar to the above methods and is also asymptotically unbiased, and consistent but not minimum variance.

Once some estimate of $\underline{\Theta}$ has been obtained, the residuals v_k may be examined[27] to give an estimate of the c_i,i=1, ... n coefficients of C(z) in (1.38) or (2.8).

$$\hat{v}_k = y_k - \underline{m}_k^t \, \hat{\underline{\Theta}} \qquad (2.27)$$

$$v_k \triangleq \epsilon_k + c_1\epsilon_{k-1} + c_2\epsilon_{k-2} \ldots\ldots c_n\epsilon_{k-n} \qquad (2.28)$$

where v_k is defined by (2.8) and $\epsilon_k \triangleq \lambda e_k$

Using the residual sequence v_k we can form the sample auto correlations $\hat{\phi}_i$ (2.29)

$$\hat{\phi}_i = \frac{1}{N-i}\sum_{k=1}^{N-i}\hat{v}_k\hat{v}_{k+i} \text{ ; } i=0,1, \ldots\ldots n \qquad (2.29)$$

From (2.28) the auto correlations ϕ_i of v_k can be expressed in terms of the c_j co-efficients j=i, ... n of C(z)

$$\phi_i = E(v_kv_{k-i}) = c_oc_i + c_1c_{i+1} \ldots\ldots c_{n-i}c_n \qquad (2.30)$$

where $c_o \triangleq \lambda$, and i=0,1, ... n ; $\phi_i \triangleq 0.0$ for i>n due to (2.28)

The set of equations (2.30) can be compared to the set (2.29) derived from the plant data,and a n+1 set of non-linear simultaneous equations obtained in n+1 unknowns c_i,i=0,1 ..... n. This is now the same as the spectral factorization problem mentioned in section 1.10 and requires some iterative routine for its solution. Since the estimate of $\underline{\theta}$ has already been shown to be biased and the effect of the coloured noise sequence v_k partially absorbed, we cannot expect the estimate of the c_j co-efficients to be of statistical utility.

Clarke's[13] approach to estimating C(z) is to invert the model for C(z) into an autoregressive process and then use the data from the residuals $\hat{v}_k$ to estimate the terms of this process, by re-applying the least squares algorithm. This method is extended iteratively and will be re-examined in the next section. The final result is a cascade of autoregressive filters of indefinite number which are said[13] to converge in practise and might then be inverted to give C(z).

2.4 Generalised Least squares estimation.

If we are permitted to assume some knowledge about the v_k sequence of (2.8), then some predetermined weighting could be applied to the components of the risk function (2.9), and a more general risk function could be defined as in (2.31).

$$R(\underline{\hat{\theta}}) \triangleq (Y-M\underline{\hat{\theta}})^t \Phi (Y-M\underline{\hat{\theta}}) \tag{2.31}$$

The weighting matrix Φ need only be considered symmetric, since any skew symmetrical portion will not contribute[21] to the value of R. It also has to be positive definite to make R positive only. The minimisation of section 2.2 can be repeated to give (2.32)

$$\underline{\hat{\theta}} = (M^t\Phi^{-1}M)^{-1} M^t\Phi^{-1}Y \tag{2.32}$$

It can similarly be shown[21] that $\underline{\hat{\theta}}$ gives an absolute minimum for $R(\underline{\hat{\theta}})$ given the model (2.7) and the weighting matrix Φ.

The co-variance matrix of the estimation errors can be obtained as in section 2.2 and becomes (2.33)

$$\Psi = (M^t\Phi^{-1}M)^{-1}M^t\Phi^{-1}E(VV^t)\Phi^{-1}M(M^t\Phi^{-1}M)^{-1} \tag{2.33}$$

Suppose now that Φ is chosen to be equal to $\Lambda \triangleq E(VV^t)$, then the co-variance matrix Ψ becomes Ψ^* as in (2.34) due to the resulting simplification.

$$\Psi^* = (M^t\Phi^{-1}M)^{-1} = (M^t(E(VV^t))M)^{-1} = (M^t\Lambda^{-1}M)^{-1} \tag{2.34}$$

It is possible to demonstrate [21,62] that this choice of Φ yields a minimum error co-variance matrix Ψ^*, and thus the smallest possible value for the risk $R(\hat{\underline{\theta}})$.

$$\text{Thus } \Psi^* \leqslant \Psi \text{ for all choices of } \Phi, \text{ where } \Psi^* \text{ is for } \Phi = \Lambda \tag{2.35}$$

This is intended to mean that the difference $\Psi - \Psi^*$ is non-negative definite, since both Ψ and Ψ^* are positive definite. The estimate for $\underline{\theta}$ obtained from (2.32) under the condition $\Phi = \Lambda$ is thus a minimum variance estimate or Markov estimate. Since in practice Λ will not be known, the minimum variance condition is unattainable, and instead we must employ some Φ which will be close to Λ, or attempt an iterative procedure for which each successive Φ will be closer to Λ. One possibility might be to use the residuals $\hat{v}_k$ from the least squares estimate to evaluate a suitable Φ, for example as in (2.36).

$$\Phi = \hat{V}\hat{V}^t \tag{2.36}$$

However as Clarke[13] shows, Φ in this case has a zero determinant and is therefore non-invertible and cannot therefore be used in (2.32).

The generalised least squares method can be viewed as transforming the data y_k, u_k to another set $y_k^* u_k^*$ for which the transformed noise sequence v_k now has zero autocorrelation $\emptyset_i$ for all $i \neq 0$. For this it is again necessary to have full knowledge of the covariance matrix of the original noise process $\Lambda = E(VV^t)$.

The iterative procedure mentioned above could be performed by choosing an estimate $\hat{C}(z)$ of the polynomial $C(z)$ and then filtering y_k, u_k to give y_k^*, u_k^* in (2.37).

$$y_k^* = \frac{1}{\hat{C}(z)} y_k \; ; \; u_k^* = \frac{1}{\hat{C}(z)} u_k \qquad (2.37)$$

then from (1.38)

$$A(z)\, \hat{C}(z)\, y_k^* = B(z)\, \hat{C}(z)\, u_k^* + \lambda C(z)\, e_k \qquad (2.38)$$

For the scalar input-output case considered, the polynomials in z will commute and we can premultiply (2.38) by $C^{-1}(z)$ to obtain

$$A(z) y_k^* = B(z)\, u_k^* + \hat{C}^{-1}(z) \lambda C(z)\, e_k \qquad (2.39)$$

The model can now be recast into the form of (2.7) and has the same meaning for $\underline{\theta}$. As $\hat{C}(z)$ approaches $C(z)$ the disturbance sequence $v_k^* = \hat{C}^{-1}(z)\, C(z) \lambda e_k$ becomes more independant and un-correlated with its past. The estimate for $\underline{\theta}$ and hence A and B will become less biased and approach the minimum variance situation as $\Phi \rightarrow \Lambda$ in (2.34). The remaining problem is how to choose successive values of $\hat{C}(z)$ to approach C(z) more closely at each iteration.

A number of authors have tried the above approach. Durbin[63], Clarke[13], Tretter[5], and Steiglitz[4] have all suggested the two stage method of searching for the parameters of C(z), and least squares solution for A(z) and B(z). The latter pair of authors represented $\hat{C}(z)$ in terms of its co-efficients and optimised these by Powell's minimisation algorithm[8]. The complete scheme came very close to Åström's method[10,11,12,37] which will be discussed later.

Clarke[13], as mentioned before, represents $\hat{C}(z)$ as an ever increasin cascade of auto regressive filters, one of which can be estimated by the least squares algorithm at each iteration of the process. The scheme is halted when the process appears to have converged. The estimate of $\hat{\underline{\theta}}$ is then approaching the minimum variance condition for the generalised least squares method since the residuals v_k are as white as possible. The $\hat{C}(z)$ polynomial could be recovered in principle by inverting the cascade of auto-regressive filters, but cannot be expected to be a statistically satisfactory estimate. Box and Jenkins[43,44] have described a process in an alien notation which is similar to the methods of this section. In effect $\hat{C}(z)$ is represented as a second order polynomial and the risk contours examined in c_i co-efficient space within stability constraints to find the optimum. As will be shown later the stability constraints on $\hat{C}(z)$ are important. Their presence greatly affects the shape of the hill during the climbing procedure, and determines whether or not the final system estimate would be of practical use.

The advantage of a least squares estimation procedure is that no explicit assumptions have to be made about the statistical properties of the random variables, beyond their boundedness. If we permit ourselves some knowledge or assumption about the probability distribution of the variables, then we can obtain more general estimation methods. It will be seen later that maximum likelihood or Bayesian methods reduce to the least squares case when the disturbances have Gaussean distributions.

2.5 Maximum likelihood Estimation.

The principle of maximum likelihood was introduced by Gauss and developed much later by R.A. Fisher in 1912. This approach is commonly regarded as providing a satisfactory estimation method because it makes the most optimal use of available data, and satisfies asymptotically the properties listed in section 2.1. In return for this benefit, we have to assume more knowledge about the stochastic disturbances, in particular the probability density function of the noise.

Given a set of data X_1 ..... X_N drawn as a random sample from a probability density $f(X,\theta)$ then the joint probability density $g(X_1, \ldots\ldots X_N,\hat{\theta})$ is known as the likelihood function. We want to know from which density this particular set X_1 ..... X_N is most likely to have come. As $\hat{\theta}$ takes different values the density changes and we wish to find the value of $\hat{\theta}$ which maximises $g(X_1, \ldots\ldots X_N,\hat{\theta})$. This value is a function of the data set X_1, ... X_N and is the maximum likelihood estimate (MLE) of θ. The likelihood function $g(X_1, \ldots\ldots X_n,\hat{\theta})$ can be regarded as function $L(\hat{\theta})$ of $\hat{\theta}$ for a given data set, and form of probability density $f(X,\Theta)$.

$$L(\Theta) = f(X_1,\Theta).f(X_2,\Theta). \ldots\ldots .f(X_N,\Theta) = \prod_{i=1}^{N} f(X_i,\Theta) \tag{2.40}$$

If $L(\Theta)$ can satisfy regularity conditions, which is commonly the case, then the maximum likelihood estimate (MLE) can be obtained from

$$\frac{\partial L(\Theta)}{\partial \Theta} = 0. = \frac{\partial}{\partial\Theta}\prod_{i=1}^{N} f(X_i,\Theta) \tag{2.41}$$

Since Log_e L is monotonic in L and attains its maximum when L is a maximum, equation (2.42) is often easier to handle.

$$\frac{\partial \mathrm{Log}_e L}{\partial \Theta} = 0. = \frac{\partial}{\partial\Theta}\sum_{i=1}^{N} \mathrm{Log}_e f(X_i,\Theta) \tag{2.42}$$

For the case when Θ is in fact a vector $\underline{\Theta}$ then (2.42) becomes a vector set of equations.

It has been pointed[23] out that it is unwise to rely on the differentiation process to locate the minimum. The function $L(\Theta)$ might have cusps or other discontinuities on the first derivative. Equation (2.41) will also locate minima and other stationary points than maxima, unless the form of $L(\Theta)$ is well known or the result is checked

Under fairly general conditions, Fisher has shown that $L(\Theta)$ approaches a normal distribution for large data sets. In fact a maximum likelihood estimate[21,23,64] is asymptotically normal, asymptotically efficient and asymptotically consistent. According to the literature[21] little can be stated about the properties of a MLE for small sample sizes. This will be further discussed and some conditions suggested in Chapter 4. A further property of a maximum likelihood estimate which can be demonstrated[23] is that of invariance. Thus if $\hat{\underline{\theta}}$ is a MLE of $\underline{\theta}$ in the density $f(X,\underline{\theta})$, and $g(\underline{\theta})$ is a function of $\underline{\theta}$ with a single-valued inverse, then the MLE of $g(\underline{\theta})$ is $g(\hat{\underline{\theta}})$.

We are going to assume that all the noise disturbances on the system (1.1) or (1.38) are normally distributed. Thus $f(X,\theta)$ is a normal or Gaussian probability density function . This assumption appears reasonable in practical situations, indeed it is possible to show[48] by the central limit theorem that the distribution of the sum of a large number of independent variates with different distributions, will appear normal. ~~all distributions will appear normal for large data sets.~~ The probability of a value X_i being drawn from a normal distribution mean μ, variance σ^2 is given by (2.43)

$$\text{Prob.} \quad (X_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp \left(-\frac{1}{2\sigma^2}(X_i-\mu)^2\right) \tag{2.43}$$

The joint probability of a sequence X_i ..... X_N being drawn is

$$\begin{aligned}\text{Prob.} \quad (X_1, \,.....\, X_N) &= \prod_{i=1}^{N} (\text{Prob.}(X_i)) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i-\mu)^2\right)\end{aligned} \tag{2.44}$$

The logarithm L' of the likelihood function defined by (2.44) is

$$L' = -\frac{N}{2}\,\text{Log}_e\, 2\pi - \frac{N}{2}\,\text{Log}_e\, \hat{\sigma}^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i-\hat{\mu})^2 \tag{2.45}$$

The maxima of L' with respect to $\hat{\sigma}^2$ and $\hat{\mu}$ are given by the solutions of (2.46)

$$\frac{\partial L'}{\partial \hat{\mu}} = \frac{1}{\hat{\sigma}^2} \sum_{i=1}^{N} (X_i - \hat{\mu}) = 0.$$
$$\frac{\partial L'}{\partial \hat{\sigma}^2} = -\frac{N}{2} \cdot \frac{1}{\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^{N} (X_i - \hat{\mu})^2 = 0. \tag{2.46}$$

Thus the MLE of μ and σ^2 is given by (2.47)

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} (X_i - \hat{\mu})^2 \tag{2.47}$$

The estimate $\hat{\mu}$ is unbiased, but that of σ^2 is biased by $\frac{N}{N-1}$. This illustrates that the MLE may in general be biased, but can frequently be simply adjusted[64] to produce an unbiased estimate (2.48)

$$\hat{\sigma}^2_{\text{unbiased}} = \frac{N}{N-1} \cdot \hat{\sigma}^2 \tag{2.48}$$

Åström[10,11,12,37] has extended the maximum likelihood procedure to the model (1.38) by finding an expression for $\hat{\epsilon}_k$ in terms of the measured variables.

$$\hat{\epsilon}_k = \frac{1}{\hat{C}(z)} \left[\hat{A}(z) y_k - \hat{B}(z) u_k \right] ; \; k=1, \ldots\ldots N \tag{2.49}$$

where $\hat{\epsilon}_k$ is defined as $\hat{\lambda} \hat{e}_k$ and taken to be normally distributed with zero mean and variance $\hat{\lambda}^2$.

It is also assumed that y_k, u_k are available from time k=1 with zero initial conditions in the plant. The joint probability of a sequence $\hat{\epsilon}_1, \ldots\ldots, \hat{\epsilon}_k, \ldots\ldots, \hat{\epsilon}_N$ is now a likelihood function $L(\hat{\underline{\theta}})$ dependant on the parameter set $\hat{\underline{\theta}}$ which is defined by (2.50). Strictly $\underline{\theta}$ should also include the n initial conditions on equation (1.38) corrosponding to those of the plant (1.1), and the bias component $\varkappa'$, and the value of λ. For brevity these have not been included here but are considered later.

$$\underline{\theta}^t = (-a_1, \ldots\ldots -a_n, b_o, b_1, \ldots\ldots b_n, c_1, \ldots\ldots c_n) \;;\; 1.(3n+1) \qquad (2.50)$$

$$L(\hat{\underline{\theta}}) = \text{Prob.}(\hat{\epsilon}_1, \ldots\ldots \hat{\epsilon}_N) = \left(\frac{1}{2\pi\hat{\lambda}^2}\right)^{N/2} \exp\left(-\frac{1}{2\hat{\lambda}^2}\sum_{k=1}^{N} \hat{\epsilon}_k^2\right) \qquad (2.51)$$

The logarithm of the likelihood function now becomes

$$L'(\hat{\underline{\theta}}) = \text{Log}_e L(\hat{\underline{\theta}}) = -\frac{1}{2\hat{\lambda}^2}\sum_{k=1}^{N}\hat{\epsilon}_k^2 - \frac{N}{2}\log_e \hat{\lambda}^2 - \frac{N}{2}\text{Log}_e 2\pi \qquad (2.52)$$

Since only the first term of (2.52) is a function of $\hat{\underline{\theta}}$ defined in terms of A,B, and C polynomials, the conditions that minimise (2.52) with respect to $\hat{\underline{\theta}}$ are the same that minimise the cost function $V(\hat{\underline{\theta}})$(2.53

$$V(\hat{\underline{\theta}}) = \tfrac{1}{2}\sum_{k=1}^{N}\hat{\epsilon}_k^2 \qquad (2.53)$$

This implies that the MLE method and the least squares method of sections 2,2 and 2.4 are strongly related since (2.53) is very similar to (2.9) and (2.31). Indeed for the assumption of Gaussian disturbances the two methods are essentially identical and differ only in their philosophy.

As in (2.46), equation (2.52) can be differentiated to give

$$\frac{\partial L'}{\partial \underline{\theta}} = \frac{-2}{2\hat{\lambda}^2}\sum_{k=1}^{N}\hat{\epsilon}_k \cdot \frac{\partial \hat{\epsilon}_k}{\partial \underline{\theta}} \qquad (2.54)$$

The MLE estimate $\hat{\underline{\theta}}$ would then be that value for which the vector set of equations (2.54) were equal to zero and thus a stationary point in $\hat{\underline{\theta}}$ space. Åström points out that $L'(\hat{\underline{\theta}})$ is quadratic in the A and B terms in $\underline{\theta}$, but is non-quadratic in the C terms in $\underline{\theta}$. Thus an analytic solution cannot be found for equations (2.54), and it is necessary to resort to some non-linear programming method of solution. This can be described as hill climbing in $\hat{\underline{\theta}}$ space.

The function $V(\hat{\underline{\theta}})$ of (2.53) is treated for simplicity as returning the altitude of the hill for each $\hat{\underline{\theta}}$ value. Notice that strictly a minimum of $V(\hat{\underline{\theta}})$ is required corrosponding to a maximum likelihood. For ease the procedure will still be referred to as hill climbing, the negative sign being understood.

Having optimised $L'(\hat{\underline{\theta}})$ via $V(\hat{\underline{\theta}})$ by this method, it is necessary to consider the estimate of λ in (2.52). Differentiating L' with respect to $\hat{\lambda}$ gives (2.55)

$$\frac{\partial L'}{\partial \hat{\lambda}} = \frac{1}{\hat{\lambda}^3}\sum_{k=1}^{N}\hat{\epsilon}_k^2 - \frac{N}{\hat{\lambda}} \qquad (2.55)$$

If we take $\hat{\lambda}^2$ to be given by $\frac{1}{N}\sum_{k=1}^{N}\hat{\epsilon}_k^2$, where $\hat{\epsilon}_k$ is the residual sequence left when $\hat{\underline{\theta}}$ is applied to the data y_k, u_k then this value of $\hat{\lambda}^2$ will make (2.55) identically zero. Thus optimisation in λ is in a sense meaningless. This should not be surprising, since from (2.49) we are only measuring a constant gain factor λ^2, the variance of ϵ_k, relative to a base of 1.0, the variance of e_k in (1.38).

The term $\varkappa'$ describing the bias level on the measurement y_k in (1.38) can also be estimated separately by measuring the means of signals, and will be shown in detail later, Thus the complete set of (3n+3) parameters in (1.38) has been dealt with. The implied assumption in this development is that the length N of the data set is such that the n initial system conditions have decayed to an insignificant level compared to the stochastic signals. This assumption is explained more fully in Chapter 5, where it is shown that initial condition estimates are always inconsistent.

It might be thought that there would exist a recursive form of the MLE method parallel to that of section 2.2. Such concepts appear to be alien to the maximum likelihood method of solution described in this section. Searching for the optimum value of $\hat{\underline{\theta}}$ requires running over the whole data set to evaluate $L'(\underline{\theta})$. This process is in effect summarised in a few matrix and vector terms for the A and B coefficients in the least squares estimate. Thus for estimates which contribute quadratically to the cost we might expect the easy addition of the knowledge of an extra data point. Since $L'(\underline{\theta})$ is non-quadratic in C(z) the above does not apply. It might be assumed that the estimate $\hat{C}(z)$ would not change a great deal from that for a data set of size N to that for a set of size N+i, where i is an integer. We might then evolve some heuristic scheme of estimating the A and B coefficients recursively by least squares and re-estimating $\hat{C}$ each i time steps. For $i \rightarrow \infty$, i.e. $\hat{C}(z)$ fixed, we would have the generalised least squares method of section 2.3, since we would have decided Φ as some estimate of the covariance matrix of v_k.

2.6 Feedback Control.

Åström's[10,11,12,37] application of the MLE to the system model (1.38) will be developed in greater detail in section 2.8. It is necessary to repeatedly run over the data set y_k, u_k using equation (2.49 and the estimates of A,B, and C polynomials. The prediction error ϵ_k is evaluated at each time step k. If it is worked out in detail, this is exactly the same as applying one step feedback control of the system in a stochastic regulator problem[65]. The object then is to remove from the output y_k all possible predictable disturbances leaving only the value ϵ_k which is not predictable, since by definition it is a random sequence independant of all other signals. Thus it is argued that in solving the MLE equations, the stochastic one step regulator problem has also been solved as a by product and will not be considered at length in this thesis. It should be clear that the estimated polynomial $1/\hat{C}(z)$ in equation (2.49) must be stable. This would be vital for a control scheme based on these estimates, and will be shown to be very important during the estimation process itself. The poles of $1/\hat{C}(z)$ must be restricted to lie within the unit circle on the complex Z plane to ensure the discrete time stability of the estimate and to ensure a satisfactory hill climbing procedure. We will later show that climbing in the space of the coefficients af $\hat{C}(z)$ is most unsound. A transformation is used to convert an unconstrained climbing space into the space of all roots of $\hat{C}(z)$ lying within the unit circle. The resulting hill climbing in the new unconstrained space then shows considerable improvements over previous methods. It was assumed in section 1.4 that the system, and hence the polynomials A and B in (1.38) represented a stable system, and this would also be required for their estimated values.

2.7 Bayesian Estimation.

Given a set of data $X_1 \ldots\ldots X_N$, we proposed in section 2.1 to find a best (by some criterion) estimate of a parameter θ assuming that the parent distribution $f(X,\theta)$ was deterministic in θ. This is known as the classical approach, and the object is to find an estimation method for the random variable $\hat{\theta}$, which satisfies some of the properties of section 2.1 relevant to the problem considered.

The Bayesian approach assumes that θ is known to vary randomly, and has a known probability density function $g(\theta)$. This supposition may not be realistic and will be discussed later. Bayes theory indicates that a good estimate would be based on the a posteriori conditional probability density function $\zeta(\theta|X_1 \ldots\ldots X_N)$, since it contains all the statistical information[48]. Having ζ we could adopt any suitable criterion to obtain a 'best' estimate. For example we might choose a loss function, Loss $(\theta-\hat{\theta})$, as described in section 2.1 and desire to minimise the risk $R(\hat{\theta})$.

$$R(\hat{\Theta}) = E_{\text{all }\Theta}(\text{Loss }(\Theta-\hat{\Theta})) \qquad (2.56)$$

$$= \int_{-\infty}^{\infty} \text{Loss }(\Theta-\hat{\Theta})\ \zeta(\Theta|X_1 \,\ldots\ldots\, X_N)d\Theta \qquad (2.57)$$

The 'best' estimate of Θ would then be the mean of the conditional density $\zeta(\Theta|X_1 \,\ldots\ldots\, X_N)$. Alternative 'best' estimates by other criteria could be derived from the mode or median of this distribution.

In general $\zeta(\Theta|X_1 \,\ldots\ldots\, X_N)$ is evaluated with the aid of Bayes rule. This rule has been badly used in history[64], but its usage will be taken as valid here.

$$\zeta(\Theta|X_1 \ldots X_N) = \frac{\text{Prob. }(\Theta, X_1 \ldots X_N)}{\text{Prob. }(X_1 \ldots\ldots X_N)}$$

$$= \frac{\text{Prob. }(X_1 \ldots X_N|_\Theta)\ .\ \text{Prob. }(\Theta)}{\text{Prob. }(X_1 \ldots\ldots X_N)}$$

where Prob.(Θ) = g(Θ), the a proiri probability density function of Θ.

The density Prob.($X_1, \ldots\ldots, X_N|_\Theta$) is often regarded as a likelihood function L"($\Theta|X_1 \,\ldots\ldots\, X_N$) which indicates more correctly that the data points $X_1 \,\ldots\ldots\, X_N$ have particular values and that the parameter Θ is to be estimated. Recursive schemes may be easily constructed since a posteriori knowledge ζ at time N can be used for a priori knowledge g(Θ) at time N+1. The initial information g(Θ) decays in importance as more information is accumulated. Thus it is possible to start with a density g(Θ) which is poorly known and yet obtain a Bayesian estimate after a long data sequence.

The validity of the Bayesian estimation method has been questioned by some statisticians[21,23,64], since a priori density $g(\theta)$ must be assumed and its existence is in many problems doubtful. Even if it exists, often the form of $g(\theta)$ is unknown, and the Bayes solution cannot be explicitly calculated. Without any a proiri knowledge, we might assume $g(\theta)$ to be uniform, i.e. all values of θ to be equally likely. Lee[19] shows in this case that the Bayes estimate reduces to the most probable estimate which lies as an abscissa to the maximum of $\zeta(\theta|x_1 \ldots\ldots x_N)$. This is also the same estimate that would be obtained by the classical maximum likelihood approach of section 2.5. We would expect this anyway if the density functions were unimodal and symmetric. Aoki points out that working with density functions directly, involves storing the whole function as a table of points which is rather unhandy for calculation. For the above reasons we have adopted the maximum likelihood doctrine for the purposes of this thesis.

2.8 Åström's method in detail.

This maximum likelihood estimation scheme was introduced in section 2.5 for the model of (1.38), and will now be given in greater detail. It was shown that the method could be reduced to minimising (2.53) with the sequence $\hat{\epsilon}_k$ given by (2.49). Values designated by the sign ^ are those of the approximate model, which are estimated values of the true variables or parameters. The derivative of $V(\hat{\underline{\theta}})$ with respect to the parameter set $\hat{\underline{\theta}}$ is given by (2.58)

$$\frac{\partial V(\underline{\hat{\theta}})}{\partial \underline{\hat{\theta}}} = \sum_{k=1}^{N} \hat{\epsilon}_k \cdot \frac{\partial \epsilon_k}{\partial \underline{\hat{\theta}}} \qquad (2.58)$$

The optimum $\underline{\hat{\theta}}$ to make equations (2.58) equal to zero is obviously the same solution as for equation (2.54). The necessary derivatives $\frac{\partial \hat{\epsilon}_k}{\partial \underline{\hat{\theta}}}$ can be obtained by differentiating the difference equation (2.49).

$$\frac{\partial \hat{\epsilon}_k}{\partial \hat{a}_i} = \frac{1}{\hat{C}(z^{-1})} \cdot z^{-i} y_k \qquad i=1, \ldots\ldots n \qquad (2.59)$$

$$\frac{\partial \hat{\epsilon}_k}{\partial \hat{b}_i} = \frac{1}{\hat{C}(z^{-1})} \cdot z^{-i} u_k \qquad i=0,1, \ldots\ldots n \qquad (2.60)$$

$$\frac{\partial \hat{\epsilon}_k}{\partial \hat{c}_i} = \frac{1}{\hat{C}(z^{-1})} \cdot z^{-i} \hat{\epsilon}_k \qquad i=1, \ldots\ldots n \qquad (2.61)$$

The evaluation of the derivatives of $V(\underline{\hat{\theta}})$ (2.58) can now be seen as a simple run over the data set 1, ... N at each stage multiplying $\hat{\epsilon}_k$ with a signal y_k, u_k or $\hat{\epsilon}_k$ passed through a filter $1/\hat{C}(z)$. The same filter is used in each of (2.59) to (2.61), and each i^{th} differential can be obtained by a simple shifting process before multiplying. This naturally leads to considerable simplifications of the computations. The second differentials of $V(\underline{\hat{\theta}})$ can also be derived from (2.58) as shown in (2.62);

$$\frac{\partial^2 V(\underline{\hat{\theta}})}{\partial\hat{\theta}_i \partial\hat{\theta}_j} = \sum_{k=1}^{N} \frac{\partial\hat{\epsilon}_k}{\partial\hat{\theta}_j} \cdot \frac{\partial\hat{\epsilon}_k}{\partial\hat{\theta}_j} + \sum_{k=1}^{N} \hat{\epsilon}_k \cdot \frac{\partial^2\hat{\epsilon}_k}{\partial\hat{\theta}_i \partial\hat{\theta}_j} \qquad (2.62)$$

And from (2.59) to (2.61)

$$\frac{\partial^2\hat{\epsilon}_k}{\partial\hat{a}_i \partial\hat{c}_j} = \frac{1}{\hat{C}(z^{-1})} \cdot -z^{-j} \frac{\partial\hat{\epsilon}_k}{\partial\hat{a}_i} = \frac{1}{\hat{C}^2(z^{-1})} \cdot -z^{-i-j} y_k \qquad (2.63)$$

$i=1, \ldots\ldots n$
$j=1, \ldots\ldots n$

$$\frac{\partial^2\hat{\epsilon}_k}{\partial\hat{b}_i \partial\hat{c}_j} = \frac{1}{\hat{C}(z^{-1})} \cdot -z^{-j} \frac{\partial\hat{\epsilon}_k}{\partial\hat{b}_i} = \frac{1}{\hat{C}^2(z^{-1})} \cdot +z^{-i-j} u_k \qquad (2.64)$$

$i=0, 1, \ldots\ldots n$
$j=1, \ldots\ldots n$

$$\frac{\partial^2\hat{\epsilon}_k}{\partial\hat{c}_i \partial\hat{c}_j} = \frac{-1}{\hat{C}(z^{-1})} \cdot z^{-j} \frac{\partial\hat{\epsilon}_k}{\partial\hat{c}_i} + \frac{1}{\hat{C}(z^{-1})} \cdot -z^{-i} \frac{\partial\hat{\epsilon}_k}{\partial\hat{c}_j} \qquad \begin{matrix} i=1, \ldots\ldots n \\ j=1, \ldots\ldots n \end{matrix}$$

$$= \frac{1}{\hat{C}^2(z^{-1})} \cdot +z^{-j-i}\hat{\epsilon}_k + \frac{1}{\hat{C}^2(z^{-1})} +z^{-i-j}\hat{\epsilon}_k = 2 \cdot \frac{1}{\hat{C}^2(z^{-1})} \cdot z^{-i-j}\hat{\epsilon}_k \qquad (2.65)$$

$$\frac{\partial^2\hat{\epsilon}_k}{\partial\hat{a}_i \partial\hat{a}_j} = \frac{\partial^2\hat{\epsilon}_k}{\partial\hat{b}_i \partial\hat{b}_j} = \frac{\partial^2\hat{\epsilon}_k}{\partial\hat{a}_i \partial\hat{b}_j} = 0. \qquad (2.66)$$

If an exact match $\underline{\hat{\theta}}$ with $\underline{\theta}$ had been achieved, then we would expect $\hat{\epsilon}_k$ to be an independent sequence. Thus the 2nd term of (2.62) would go to zero at the exact match, as well as the terms arising from (2.66). There would be some justification for ignoring this 2nd term altogether, even under mismatch conditions. The second derivative matrix of $V(\underline{\hat{\theta}})$ would then be positive semi-definite due to the 1st term in (2.62), being effectively a correlation matrix of a set of signals of length N. The full expression might give a non-positive definite matrix, which would be difficult to handle with the Newton-Raphson hill climbing routine used by Åström. There would be a loss of climbing efficiency due to using the approximate matrix, but this could be offset by the easier computation. The 1st term also requires only the sum product of the first derivative terms which have already been generated.

The Newton-Raphson algorithm used to optimise $V(\hat{\underline{\theta}})$ is given by (2.67)

$$\hat{\underline{\theta}}_{i+1} = \hat{\underline{\theta}}_i - \alpha \left\{ \frac{\partial^2 V(\hat{\underline{\theta}})}{\partial \hat{\underline{\theta}} \partial \hat{\underline{\theta}}} \right\}_i^{-1} \cdot \frac{\partial V(\hat{\underline{\theta}})}{\partial \hat{\underline{\theta}}_i} \quad \text{at the } i^{th} \text{ iteration} \quad (2.67)$$

At the first iteration i=1, and Åström sets the $\hat{C}(z)$ coefficients $\hat{c}_j$, j=1 ..... n to be zero. This gives a simple least squares solution of A and B coefficients for $\hat{\underline{\theta}}_2$, which will be biased since the colouration due to C(z) is not accounted for. Ideally the factor α is 1.0 on quadratic hills and gives one step convergence, but is commonly set $0.<\alpha<1.0$ according to the ease of climbing. Thus for difficult hills α may be quite small, and even negative to produce some climbing action if the practical matrix is not positive definite.

We give some examples in Chapter 5 to illustrate this effect and it is suggested that another algorithm due to Fletcher and Powell[9] could be useful here. The iterative scheme of (2.67) is replaced by (2.68), where the second derivative matrix has been replaced by an estimated matrix H_i.

$$\underline{\hat{\Theta}}_{i+1} = \underline{\hat{\Theta}}_i - \alpha . H_i \frac{\partial V(\hat{\Theta})}{\partial \underline{\hat{\Theta}}_i} \qquad \text{at the } i^{th} \text{ iteration where } H_i = I \text{ initially} \qquad (2.68)$$

Here H_i is forced to be positive definite, and updated with gradient information at each iteration, and tends to the inverse of the second derivative matrix at the optimum. The factor α is determined by searching for a minimum along the line defined by $H_i * \frac{\partial V(\hat{\Theta})}{\partial \underline{\hat{\Theta}}_i}$ so that the $(i+1)^{th}$ gradient is orthogonal to the line. For a quadratic surface the algorithm of (2.68) takes q iterations, where q is the dimension of $\underline{\Theta}$. For more general surfaces the Fletcher-Powell algorithm frequently shows improvement in convergence over the Newton-Raphson, and thus is beneficial for the problem studied.

2.9 Filter stability.

The method of section 2.8 requires several runs over the data set y_k, u_k with a filter $\frac{1}{\hat{C}(z^{-1})}$ which is defined in terms of coefficients $\hat{c}_i$, i=1, ... n as in (1.38). So far there has been no restriction on the values of $\hat{c}_i$ and the filter could easily be unstable with its poles outside the unit circle in the z plane. The output of such a filter will not remain bounded over a finite interval when excited with bounded arbitrary sequence. This can cause a great deal of trouble with a hill climbing routine, especially when the roots of $\hat{C}(z)$ lie near the unit circle. A small change in one of the $\hat{c}_i$ coefficients can give an extremely large change in the cost function value. The natural response is to use a much smaller value of α in (2.67) or (2.68). This effect can slow down the whole convergence, as it is equivalent to striking a constraint, the unit circle, which the climbing algorithms given cannot handle effectively. The alternative is to check analytically the stability of $\frac{1}{\hat{C}(z^{-1})}$ before being used to evaluate the function. This in general requires finding the n roots of the polynomial when given the coefficients, and there is no easy method when n is larger than 4. Resort must then be made instead to say Jury's[29] test for stability before proceeding. Some method of constraining the roots of $\hat{C}(z)$, or at least of detecting sensitive conditions, has to be found to cover the situation, and avoid human intervention in the climbing process which has been required on occasion with Åström's method.

CHAPTER 3

THE ARCHITECTURE OF THE ESTIMATION HILL

3.1 The variance of a signal.

The maximum likelihood method introduced in sections 2.5 and 2.8 is prone to difficulties due to sensitivity and unstable filtering as explained in section 2.9. To analyse these troubles we would like to be able to predict the cost $V(\underline{\hat{\theta}})$ in equation (2.53) for various systems and parameters, before the estimation procedure is started. We require to calculate the sample variance of the output signal of a discrete time system expressed as a rational z polynomial, and then examine the result for possible sensitive regions.

Consider a simple discrete time filter as in equation (3.1), which as a rational z polynomial is a mixture of moving average and auto-regressive representations[32].

$$v_k = \frac{1 + n_1 z^{-1} + n_2 z^{-2} \;\ldots\ldots\; n_m z^{-m}}{1 + d_1 z^{-1} + d_2 z^{-2} \;\ldots\ldots\; d^l z^{-l}} \; e_k = \frac{N(z)}{D(z)} \; e_k \qquad (3.1)$$

where $N(z) = z^l + n_1 z^{l-1} + n_2 z^{l-2} + \ldots\ldots + n_m z^{l-m}$

$D(z) = z^l + d_1 z^{l-1} + d_2 z^{l-2} \quad \ldots\ldots\; d_l z^o \quad ; \quad l \geqslant m$

e_k is a random independent sequence, $E(e_k) = 0.$

$E(e_k . e_{k-i}) = \sigma_e^2 * \delta(i)$

The system as defined gives no prior response to an input e_k and is therefore physically realisable for all positive values of m and l.

A physical plant may well give no output at time k due to an input at time k, and this would require delay terms in (3.1).

The variance of the output v_k can be calculated, for example for l=1,m=0, by expanding (3.1) as (3.2).

$$v_k + d_1 v_{k-1} = e_k$$

$$\therefore\ v_k = e_k - d_1 v_{k-1} \tag{3.2}$$

Equation (3.2) is now squared and expectations taken to give (3.3)

$$E(v_k^2) = E(e_k^2) + d_1^2 E(v_{k-1}^2) - 2d_1 E(e_k v_{k-1}) \tag{3.3}$$

$$= \sigma_e^2 + d_1^2 E(v_{k-1}^2) \tag{3.4}$$

The last term of (3.3) is zero since e_k and v_{k-1} are independent; also we can take $E(v_{k-1}^2) = E(v_k^2)$ for a stationary process. Thus the variance of the signal y_k is $\sigma_e^2/(1-d_1^2)$, which is a simple result.

For larger values of m and l it is easier to invert the polynomial D(z) by synthetic long division[32] to give a moving average representation. The squared terms can then be summed as shown in (3.5)

$$v_k = N(z).\left[D(z)\right].^{-1}\ e_k$$

$$= (r_o + r_1 z^{-1} + r_2 z^{-2} \ \ldots\ldots\ r_j z^{-j}) e_k \ ;\ j \rightarrow \infty$$

$$\therefore E(v_k^2) = (r_o^2 + r_1^2 + r_2^2 \ \ldots\ldots\ r_j^2)\,\sigma_e^2 \ ;\ \ j \rightarrow \infty \tag{3.5}$$

where r_o = 1.0 for N and D defined as in (3.1)

The cross product terms $E(e_k . e_{k-i})$, $|i|>0$ are all zero due to the independence of the e_k sequence. The inversion process in general produces a chain of coefficients, r_k as shown, which converge asymptotically to zero magnitude when $j \rightarrow \infty$ if $\frac{1}{D(z)}$ is a stable polynomial.

It is easy to show[45] that the inversion process to obtain r_o, r_1, r_2 ..... is equivalent to exciting a system $\frac{1}{D(z)}$ starting at zero initial conditions with deterministic pulses from $k=1$ to $k=m+1$. The successive amplitudes of these pulses are given by the successive coefficients of $N(z)$. After $k=m+1$ the system is allowed to run free until $k=j+1$; $j \rightarrow \infty$. It is obvious that any finite limit on j would be an arbitrary one decided by numerical convergence. Only by knowing the dynamic modes of such a system, i.e. the roots of $D(z)$[17], can we easily obtain a closed form solution to the sum shown in (3.5). As one might expect, the simple structure for the variance in terms of the coefficients of $D(z)$, as exemplified by (3.4), is not repeated for higher orders. Many text books[29,30,31] give tables of variances expressed in terms of the coefficients of N and D, and these demonstrate all too clearly that no structure exists as the polynomial order is increased. We claim that only by working in the roots of N and D can the structure of the variance calculation be made plain, and the analysis mentioned above really be attempted.

3.2 Approach via complex variable theory.

The variance calculation considered in the provious section can be approached more rigorously using complex variable theory[21,34,49]. A complex variable z is defined to have real and imaginary components in the complex plane. A general function of z, F(z) is regular in a region D' if it is single valued and its differential exists at all points within D'. Then the partial derivatives at each point have to satisfy the Cauchy-Riemann differential equations. A singularity is any finite point z_o where F(z) ceases to be regular, but is regular in the neighbourhood of z_o. Cauchy's integration theorem, as derived from the Cauchy-Riemann equations, states that the integral of F(z) round a closed contour C surrounding region D' is zero, provided that F(z) is regular in D' and on C.

$$\oint_C F(z)dz = 0 \text{ for } F(z) \text{ analytic on and inside C} \qquad (3.6)$$

If F(z) has a singularity at $z=\rho_i$ within C, then the integral becomes

$$\oint_C F(z)dz = 2\pi j \,.(\text{residue at the singularity } Z=\rho_i) \qquad (3.7)$$

When F(z) is a rational function the only singularities are poles of finite order. A pole is defined by (3.8)

$$F(z) \text{ has a pole } \rho_i \text{ of order } m_i \text{ if } F(z) = \frac{F'(z)}{(z-\rho_i)^{m_i}} \qquad (3.8)$$

where F'(z) is regular within C and non-zero at $z=\rho_i$

The residues in (3.7) can then be calculated by equation (3.9)

Residue of F(z) at a pole ρ_i of order m_i

$$= \frac{1}{(m_i-1)!} \cdot \frac{\partial^{m_i-1}}{\partial z^{m_i-1}} \left((z-\rho_i)^{m_i} F(z)\right)\Bigg|_{z=\rho_i} \qquad (3.9)$$

Jury[21] defines a z transform as (3.10), which can be regarded as a Laurent series in z.

$$F(z) \triangleq \sum_{k=0}^{\infty} f_k z^{-k} = f_0 + f_1 z^{-1} + f_2 z^{-2} + f_3 z^{-3} \; \ldots\ldots \qquad (3.10)$$

where f_k is the value of a signal f at discrete time k.

Given F(z) we can obtain f_n by multiplying the series (3.10) by z^{n+1} and integrating around a contour C enclosing any singularities of F(z). The only surviving term from the integration is given by (3.11)

$$\oint_C F(z) z^{n-1} dz = 2\pi j \; .(\text{Residue of } F(z) z^{n-1} \text{ at } z=0.) \qquad (3.11)$$

$$= 2\pi j \cdot f_n \qquad (3.12)$$

The analysis can also be extended to a two sided z transform (3.13)

$$\text{if } F(z) = \sum_{k=-\infty}^{k=\infty} f_k z^{-k} \qquad (3.13)$$

$$\text{then } f_n = \frac{1}{2\pi j} \oint_C F(z) z^{n-1} dz \qquad (3.14)$$

The contour C must lie in the ring of convergence of F(z) in the z plane with the point z=0. in its interior. If f_n is bounded, the unit circle /z/=1.0 will belong to the ring and can be used for C. The two sided z transform is often used to describe the power spectral density $\Phi_{ff}(z)$, (3.16), of a discrete time signal f_k in terms of its auto-correlation function ϕ_n as defined in (3.15)

$$\phi_n = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} f_k f_{k-n} \text{ ; where n is an integer} \qquad (3.15)$$

$$\Phi_{ff}(z) \triangleq \sum_{n=-\infty}^{n=\infty} \phi_n z^{-n} \qquad (3.16)$$

Since ϕ_n is even and bounded if f_k is bounded it can be recovered from $\Phi_{ff}(z)$ in the same manner as (3.11). Contour C can again be the unit circle.

$$\phi_n = \frac{1}{2\pi j} \oint_C \Phi_{ff}(z) z^{n-1} dz \qquad (3.17)$$

Now the variance of the signal f_k is defined as ϕ_o in (3.15) and can therefore be obtained for n=0 from (3.17) if $\Phi_{ff}(z)$ is given. The spectrum $\Phi_{ff}(z)$ can be found from equations (3.13),(3.15) and (3.16) and can be written as $F(z)F(z^{-1})$, since it is symmetric. The variance σ_f^2 of the signal f_k is then given by (3.18)

$$\text{variance } (f_k) = \sigma_f^2 = \phi_o = \frac{1}{2\pi j}\int_C F(z)F(z^{-1})z^{-1}dz \qquad (3.18)$$

If the signal f_k were passed through a system with transfer G(z), then the variance of the output would be given by

$$\text{variance of output signal} = \frac{1}{2\pi j}\int_C G(z)G(z^{-1})F(z)F(z^{-1})z^{-1}dz$$

$$= \frac{1}{2\pi j}\int_C G(z)G(z^{-1})\Phi_{ff}(z)z^{-1}dz \qquad (3.19)$$

The contour C in (3.18) can be conveniently chosen as the unit circle, as this would always separate the singularities of F(z)/z from those of F(z^{-1}). It should now be clear that since we have to evaluate all the residues of F(z)/z given that F(z) is a rational function of z, it is most desirable to know the roots of F(z), explicitly. Otherwise it is necessary to solve for the roots of F(z), or else integrate (3.18) numerically as a definite integral around the unit circle. This latter method is quite easy, but does not shed any light on the structure of the variance result. Similarly, the method of Nekolny[46,47], although having the advantage of indicating filter stability, again fails to supply us with any insight into the structure behind the sensitivity of parameters.

3.3 Calculation of signal variance.

We are going to represent the simple filter (3.1) in terms of its roots as (3.20) in order to gain the advantages seen in the previous section.

$$\frac{N(z)}{D(z)} = \frac{\prod_{i=1}^{m}(z-\eta_i)}{\prod_{i=1}^{l}(z-\delta_i)} \tag{3.20}$$

where η_i, δ_i are the roots of N(z),D(z) respectively as defined in (3.1)

For the case m<l in (3.1), the root form of (3.20) has an extra factor z^{l-m} which can be considered as l-m extra terms. Each of these terms can be taken to have η_i equal to zero. Thus the strict equivalent system to (3.1) is given as (3.21)

$$\frac{N(z)}{D(z)} = \frac{\prod_{i=1}^{l}(z-\eta_i)}{\prod_{i=1}^{l}(z-\delta_i)} \quad ; \text{ where } \eta_i = 0.0 \;;\; i=m+1, \ldots\ldots l \tag{3.21}$$

The residues R_i at the poles δ_i will be required when (3.21) is broken down into partial fractions.

$$R_i \triangleq \frac{\prod_{j=1}^{l}(\delta_i-\eta_j)}{\prod_{j=1}^{l}(\delta_i-\delta_j)} \tag{3.22}$$

The system in (3.21) can be reduced to a proper fraction by long division. This can then be expanded in partial fractions to give (3.23)

$$\frac{N(z)}{D(z)} = 1.0 + \sum_{i=1}^{l} \frac{1}{z-\delta_i} \cdot R_i \qquad (3.23)$$

The roots η_i and δ_i of the polynomials are in general complex, but have been taken for ease to be distinct. A similar analysis may be repeated for multiple roots, but has not been given here.

To calculate the variance of the v_k sequence of (3.1) we can consider it in the sense of (3.19) by substituting N(z)/D(z) for G(z). The power spectral density $\Phi_{ee}(z)$ of the signal e_k is a constant $\sigma_e^2 * z^o$ since by the definition of e_k in (3.1), all the auto-correlations ϕ_i are zero, $-\infty < i < \infty$, except ϕ_o which is σ_e^2. This means (3.24) can be simplified as shown.

$$\text{variance } (v_k) = \frac{1}{2\pi j}\int_C \frac{N(z)}{D(z)} \cdot \frac{N(z^{-1})}{D(z^{-1})} \Phi_{ee}(z) z^{-1} dz$$

$$= \frac{\sigma_e^2}{2\pi j}\int_C \frac{N(z)}{D(z)} \cdot \frac{N(z^{-1})}{D(z^{-1})} z^{-1} dz$$

where C is the unit circle $/z/ = 1.0$ (3.24)

The partial fraction forms of N and D can be substituted in to (3.24) from (3.21) or (3.23). The same expansions will hold for $F(z^{-1})$, since we can derive the partial fractions in terms of z^{-1}.

$$\text{variance } (v_k) \triangleq \sigma_v^2 = \frac{\sigma_e^2}{2\pi j}\int_C \frac{1}{z}\cdot \left[1.0 + \sum_{i=1}^{l} \frac{R_i}{z-\delta_i}\right]\cdot\left[1.0 + \sum_{\nu=1}^{l} \frac{R_\nu}{z^{-1}-\delta_\nu}\right] dz \tag{3.25}$$

This can be split into four separate integrations i) to iv):

i) $\frac{\sigma_e^2}{2\pi j}\int_C \frac{1}{z} \cdot 1.0 \, dz$ Residue at z=0.0 is σ_e^2

ii) $\frac{\sigma_e^2}{2\pi j}\int_C \frac{1}{z} \, 1.0 \sum_{\nu=1}^{l} \frac{R_\nu Z}{1-\delta_\nu Z} \, dz$ Residue at z=0.0 is 0.0

iii) $\frac{\sigma_e^2}{2\pi j}\int_C \frac{1}{z}\cdot\sum_{i=1}^{l} \frac{R_i}{z-\delta_i} \cdot 1.0 \, dz$ Residue at z=0.0 is $\sum_{i=1}^{l} \frac{R_i}{-\delta_i} \quad \sigma_e^2$

Residue at $z=\delta_i$ is $\frac{R_i}{\delta_i} \; \sigma_e^2$

$\therefore$ Sub total $= \sum_{i=1}^{l} \frac{R_i}{\delta_i} - \sum_{i=1}^{l} \frac{R_i}{\delta_i} = 0.0$

iv) $\frac{\sigma_e^2}{2\pi j}\int_C \frac{1}{z}\cdot\sum_{i=1}^{l} \frac{R_i}{z-\delta_i} \cdot\sum_{\nu=1}^{l} \frac{R_\nu Z}{1-\delta_\nu z} \, dz$ Residue at z=0. is 0.0

Residue at $z=\delta_i$ is $\frac{R_i}{\delta_i} \, \frac{R_\nu \delta_i}{1-\delta_\nu\delta_i} \, \sigma_e^2$;

$\therefore$ Sub total $\sum_{i=1}^{l} \sum_{\nu=1}^{l} \frac{R_i R_\nu}{1-\delta_\nu\delta_i} \quad \sigma_e^2$

$$\therefore \sigma_v^2 = \text{total of residues} = \left[1.0 \quad + \sum_{i=1}^{l}\sum_{\nu=1}^{l} \frac{R_i R_\nu}{1-\delta_i\delta_\nu}\right] \sigma_e^2 \tag{3.26}$$

Slightly different forms can be obtained for (3.26) by factorising in different ways. For example, for the case of m=1, we can evaluate the initial response v_1 to a unit pulse input e_1 by using (3.11) and this gives (3.27)

$$v_1 = \frac{\prod_{i=1}^{m}(\eta_i)}{\prod_{i=1}^{l}(\delta_i)} + \sum_{i=1}^{l} \frac{R_i}{\delta_i}$$

$$= 1.0 \text{ by long division of (3.20) or (3.1)}$$

$$\text{for } e_1 = 1.0 \qquad (3.27)$$

The identity (3.27) can be substituted in (3.26) to give the other possible forms (3.28),(3.29), and (3.30)

$$\sigma_v^2 = \left[\frac{\prod_{i=1}^{m}(\eta_i)}{\prod_{i=1}^{l}(\delta_i)} + \sum_{i=1}^{l} \frac{R_i}{\delta_i} + \sum_{i=1}^{l}\sum_{\nu=1}^{l} \frac{R_i R_\nu}{1-\delta_i\delta_\nu}\right] \sigma_e^2 \qquad (3.28)$$

$$\sigma_v^2 = \left[\frac{\prod_{i=1}^{m}(\eta_i)}{\prod_{i=1}^{l}(\delta_i)}\left\{2.0 - \frac{\prod_{i=1}^{m}(\eta_i)}{\prod_{i=1}^{l}(\delta_i)}\right\} + \sum_{i=1}^{l}\sum_{\nu=1}^{l} \frac{R_i R_\nu}{\delta_i\delta_\nu} + \sum_{i=1}^{l}\sum_{\nu=1}^{l} \frac{R_i R_\nu}{1-\delta_i\delta_\nu}\right] \sigma_e^2 \qquad (3.29)$$

$$\sigma_v^2 = \left[\frac{\prod_{i=1}^{m}(\eta_i)}{\prod_{i=1}^{l}(\delta_i)}\left\{1.0 + \sum_{i=1}^{l} \frac{R_i}{\delta_i}\right\} + \sum_{i=1}^{l}\sum_{\nu=1}^{l} \frac{R_i R_\nu}{\delta_i\,\delta_\nu(1-\delta_i\delta_\nu)}\right] \sigma_e^2 \qquad (3.30)$$

The form of (3.26) will be adopted here as being the simplest and can be used for m=l or for m<l. Some further simplifications can be made if the δ_i poles include complex pairs. Then certain terms containing $\delta_i\delta_j$ pairs, where δ_j is a conjugate of δ_i, will combine algebraically. The structure of the variance calculation can now be easily seen from (3.28) for any number l of poles δ_i and any number m of zeros η_i, for the system (3.1) or (3.20). This is in distinct contrast to the variance expressions given in

terms of coefficients only[29,30,31]. The way is now open to making some more definite statements about the sensitivity of the estimation procedures given in Chapter 2. It will be made clear later that one of the results of the estimation method will be to reduce the variance σ_v^2 of such a simple system as (3.20) to as small as possible. For m=1 this is achieved by matching all the poles δ_i and zeros η_i. All the residues R_i then go to zero, and $\sigma_y^2 = \sigma_e^2$. When m<1, complete matching of all η_i and δ_i cannot occur due to the lack of sufficient zeros.

3.4 Sample variance of a finite data set.

The variance σ_v^2 of a signal v_k can frequently only be ~~calculated~~ estimated from a sample, k=1, ... N . Then $\hat{\sigma}_v^2$ is defined as $\frac{1}{N-1}\sum_{k=1}^{N}(v_k-\bar{v})^2$, where $\bar{v}$ is the mean value of v_k and is given by $\frac{1}{N}\sum_{k=1}^{N} v_k$. The $\frac{1}{N-1}$ factor occurs due to the loss of one degree of freedom. If the signal v_k is known to have zero mean, i.e. it originates from a bias free source as in (3.1), then σ_v^2 can be calculated as $\frac{1}{N}\sum_{k=1}^{N} v_k^2$.

For the case when v_k is formed from an infinite past history ϵ_k ; k=1,0, ... $-\infty$ the signal is stationary in the statistical sense. Then $E(v_1^2) = E(v_2^2) = ... = E(v_N^2)$ and the expectation of the sample variance is equal to that for the infinite data case.

During the estimation procedure to be used later, a signal v_k ; k=1, ... N can only be generated as in (3.1) from a finite data

set e_k ; k=1, ... N . The filter N(z)/D(z) therefore has a growing memory property and is strictly non-stationary. Thus
$E(v_1^2) \neq E(v_2^2) \ldots\ldots \neq E(v_N^2)$.

To be able to calculate the expectation of the sample variance above, $E(v_k^2)$ can be expanded as $E(v'_o e_k + v'_1 e_{k-1} \ldots\ldots + v'_{k-1} e_1)^2$. Here v'_i is defined as the impulse response at delay i of the system (3.1) when excited by an input $e'_1 = 1.0$ and thereafter $e'_k = 0.0$, k>0. The term $E(v_k^2)$ can now be expressed as $\sigma_e^2 * \sum_{i=0}^{k-1} v_i'^2$, since cross terms $e_k e_{k-j}$; j≠0 have an expectation of zero. We now require a closed form expression for $\sum_{i=0}^{k-1} v_i'^2$, for any value of k, to be able to calculate $E[\hat{\sigma}_v^2$ in the finite history situation.

Consider two sequences f_k and h_k whose z transforms as defined by (3.10) are F(z) and H(z). Jury[21] shows, by arguing from the Laplace transform definition, that the product $f_k h_k$ sequence can be represented as a z transform G(z) by means of the convolution integral (3.31).

$$G(z) \triangleq \frac{1}{2\pi j}\int_C p^{-1}F(p)H(z/p)dp \triangleq f_o h_o + f_1 h_1 z^{-1} + f_2 h_2 z^{-2} \ldots\ldots$$

$$= \sum \text{residues of } p^{-1}F(p) = -\sum \text{residues of } H(z/p) \qquad (3.31)$$

where p is a complex variable, and contour c encloses all the singularities of F(p)/p, but excludes those of H(z/p).

By setting z=1.0 in (3.31), we can obtain the sum $\sum_{k=0}^{\infty} f_k h_k$ as in (3.32)

$$\sum_{k=0}^{\infty} f_k h_k = f_o h_o + f_1 h_1 + f_2 h_2 \ldots\ldots = \frac{1}{2\pi j}\int_C F(p)/_p \cdot H(p^{-1})dp \qquad (3.32)$$

Suppose the sequence f_k was defined to be equal to a sequence (v'_k) , then (3.32) would give $\sum_{k=0}^{\infty} v'^{2}_{k} h_k$. We will now define the sequence h_k to be 1.0 for $o \leqslant k < N'$ and zero for any other value of k. Thus the z transform H(z) of the h_k sequence can be defined by (3.33) as from(3.10)

$$h_k = 1.0 \text{ for } 0 \leqslant k < N' \quad ; 0.0 \text{ otherwise}$$

$$= (1.0 \text{ for } 0 \leqslant k \leqslant \infty) - (1.0 \text{ for } N' \leqslant k \leqslant \infty)$$

$$\therefore H(z) = \frac{1}{1-z^{-1}} - \frac{z^{-N'}}{1-z^{-1}} = \frac{z}{z-1}(1-z^{-N'}) \qquad (3.33)$$

The sum product $\sum_{k=0}^{\infty} f_k h_k$ in (3.32) in this case will now be equal to the sum of v'^{2}_{k} for $o \leqslant k < N'$, i.e. $\sum_{k=0}^{N'-1} v'^{2}_{k}$. Thus we have obtained the sum of v'^{2}_{k} for a finite length N'.

$$\sum_{k=0}^{N'-1} v'^{2}_{k} = \sum_{k=0}^{\infty} f_k h_k = \frac{1}{2\pi j}\int_C \frac{1}{p} F(p) H(p^{-1}) dp \qquad (3.35)$$

where H(z) is defined as in (3.33)

$$F(z) \text{ is the z transform } \triangleq \sum_{k=0}^{\infty} v'^{2}_{k} z^{-k} \qquad (3.36)$$

Now F(z) defined by (3.36) can be calculated using the convolution integral (3.31)

$$F(z) = \frac{1}{2\pi j}\int_C \frac{N(p)}{D(p)} \cdot \frac{1}{p}\frac{N(z/p)}{D(z/p)} \cdot dp \qquad (3.37)$$

This can be expanded as shown before in (3.25)

$$F(z) = \frac{1}{2\pi j}\int_C \frac{1}{p}\left(1.0 + \sum_{i=1}^{l} \frac{R_i}{p-\delta_i}\right)\left(1.0 + \sum_{\eta=1}^{l} \frac{R_\eta}{zp^{-1} - \delta_\eta}\right) dp \qquad (3.38)$$

Then in a similar way to that of (3.25) to (3.28) we can show

$$F(z) = 1.0 \ + \sum_{i=1}^{l} \sum_{\gamma=1}^{l} \frac{R_i R_\gamma}{z-\delta_i\delta_\gamma} \qquad (3.39)$$

Now (3.35) can be calculated explicitly using F(z) from (3.39) and H(z) from (3.33)

$$\sum_{k=0}^{N-1} v'^{2}_{k} = \frac{1}{2\pi j}\int_C \frac{1}{p}\left[1.0 + \sum_{i=1}^{l}\sum_{\gamma=1}^{l} \frac{R_i R_\gamma}{p-\delta_i\delta_\gamma}\right] \frac{p^{-1}(1-p^{N'})}{p^{-1}-1}dp$$

The 1st term in $[\cdot]$ gives: $\frac{1}{2\pi j}\int_C \frac{1}{p} \cdot \frac{1-p^{N'}}{1-p} .dp$

residue at p=0. is σ_e^2 only.

The i,γ th term of $[\cdot]$ gives $\frac{1}{2\pi j}\int_C \frac{1}{p} \cdot \frac{R_i R_\gamma}{p-\delta_i\delta_\gamma} \cdot \frac{1-p^{N'}}{1-p} .dp$

residue at p=0.0 is $-\frac{R_i R_\gamma}{\delta_i\delta_\gamma}$

residue at p= $\delta_i\delta_\gamma$ is $\frac{R_i R_\gamma}{\delta_i\delta_\gamma} \cdot \frac{1-(\delta_i\delta_\gamma)^{N'}}{1-\ \delta_i\delta_\gamma}$

these two residues combine algebraically to give $R_i R_\gamma \cdot \frac{1-(\delta_i\delta_\gamma)^{N'-1}}{1-\delta_i\delta_\gamma}$

The sum of all the above residues gives ~~the expectation of the sample variance as~~

$$\sum_{k=0}^{N'-1} v'^2_k = 1.0 + \sum_{\gamma=1}^{l}\sum_{i=1}^{l} R_i R_\gamma \cdot \frac{1-(\delta_i\delta_\gamma)^{N'-1}}{1-\delta_i\delta_\gamma} \tag{3.40}$$

We can now form the *expectation of the* sample variance for a finite data history mentioned before by using a separate term for each $E(v_k^2)$, $k=1, \ldots\ldots N$ and using the relation of (3.40).

Expectation of Sample variance of $v_k = \phi_o^N = \frac{1}{N}\left[E(v_1^2)+E(v_2^2) \;\ldots\ldots\; E(v_N^2)\right]$

$$
\begin{array}{lr}
= \frac{1}{N}\left\{ E(v'_o e_1)^2 \right. & \text{for } E(v_1^2) \\
\quad + E(v'_o e_2 + v'_1 e_1)^2 & \text{for } E(v_2^2) \\
\quad + \ldots\ldots & \ldots\ldots \\
\quad + E(v'_o e_{N-1} + \ldots\ldots + v'_{N-3} e_2 + v'_{N-2} e_1)^2 & \text{for } E(v_{N-1}^2) \\
\quad \left. + E(v'_o e_N + \quad \ldots\ldots + v'_{N-2} e_2 + v'_{N-1} e_1)^2 \right\} & \text{for } E(v_N^2)
\end{array}
$$

$$
\begin{aligned}
= \frac{\sigma_e^2}{N}\Bigg\{ & 1.0 \\
& + 1.0 + \sum_{i=1}^{l}\sum_{\gamma=1}^{l} R_i R_\gamma \\
& + \ldots\ldots \\
& + \ldots\ldots \\
& + 1.0 + \sum_{i=1}^{l}\sum_{\gamma=1}^{l} R_i R_\gamma \; \frac{1-(\delta_i\delta_\gamma)^{N-2}}{1-\delta_i\delta_\gamma} \\
& + 1.0 + \sum_{i=1}^{l}\sum_{\gamma=1}^{l} R_i R_\gamma \; \frac{1-(\delta_i\delta_\gamma)^{N-1}}{1-\delta_i\delta_\gamma} \Bigg\}
\end{aligned}
$$

All these terms add together to give (3.41):

$$
\begin{aligned}
E\hat{\sigma}_v^2 = \phi_o^N &= \sigma_e^2 + \frac{\sigma_e^2}{N}\sum_{i=1}^{l}\sum_{\gamma=1}^{l} \frac{R_i R}{(1.-\delta_i\delta_\gamma)^2}\left[N(1.-\delta_i\delta_\gamma)-1.0 + (\delta_i\delta_\gamma)^N\right] \\
&= \sigma_e^2 + \sigma_e^2\left[\sum_{i=1}^{l}\sum_{\gamma=1}^{l} \frac{R_i R_\gamma}{(1.-\delta_i\delta_\gamma)}\right] * \left[1.0 - \frac{1.-(\delta_i\delta_\gamma)^N}{N(1.-\delta_i\delta_\gamma)}\right]
\end{aligned} \tag{3.41}
$$

Clearly (3.41) is biassed but will converge to the infinite memory stationary case of (3.26) when N is large. If we decide on a particular value for the bias, then N and $(\delta_i \delta_\gamma)$ are clearly related. Such a relation is shown in graph form in figure 20 and will be discussed in Chapter 4 as a criterion for N given $(\delta_i \delta_\gamma)$.

3.5 Auto-correlation function of v_k.

The auto-correlations ϕ_r of the signal v_k, defined by (3.17) can be evaluated in a similar way to the variance ϕ_0 in the previous section. The system z transform $N(z)/D(z)$ can be broken down into the form of (3.23) for convenience and can then be substitu ted into (3.17) in the manner of (3.19).

$$\phi_r = \frac{1}{2\pi j}\int_C \frac{N(z)}{D(z)} \cdot \frac{N(z^{-1})}{D(z^{-1})} \cdot \Phi_{ee}(z) z^{r-1} dz$$

$$= \frac{\sigma_e^2}{2\pi j}\int_C \left(1.0 + \sum_{i=1}^{l} \frac{R_i}{z-\delta_i}\right)\left(1.0 + \sum_{\gamma=1}^{l} \frac{R_\gamma}{z^{-1}-\delta_\gamma}\right) z^{r-1} dz \qquad (3.42)$$

since $\Phi_{ee}(z) = \sigma_e^2$ only, due to the definition of e_k in (3.1)

This integral can again be expressed as the sum of four parts.

i) $\dfrac{\sigma_e^2}{2\pi j}\displaystyle\int_C z^{r-1} .1.0 .1.0 \ dz$ No singularities for $r>0$

ii) $\frac{\sigma_e^2}{2\pi j}\int_C z^{r-1}.1.0 \ .\sum_{\gamma=1}^{l} \frac{R_\gamma z}{1-\delta_\gamma z} .dz$ No singularities for r>0

iii) $\frac{\sigma_e^2}{2\pi j}\int_C z^{r-1} .\sum_{i=1}^{l} \frac{R_i}{z-\delta_i} .1.0 \, dz$ Residue at $z=\delta_i$ is $\sigma_e^2 R_i \ \delta_i^{\,r-1}$.

Sub. total$=\sigma_e^2\sum_{i=1}^{l} R_i \delta_i^{\,r-1}$

iv) $\frac{\sigma_e^2}{2\pi j}\int_C z^{r-1}\sum_{i=1}^{l} \frac{R_i}{z-\delta_i} .\sum_{\gamma=1}^{l} \frac{R_\gamma z}{1-\delta_\gamma z} . \, dz$

Residue at $z=\delta_i$ is $\sigma_e^2 R_i \delta_i^{\,r-1}\sum_{\gamma=1}^{l} \frac{R_\gamma}{1-\delta_\gamma\delta_i}$;

sub. total $\sum_{i=1}^{l}\sum_{\gamma=1}^{l} \frac{R_i R_\gamma \delta_i^{\,r}}{1-\delta_\gamma\delta_i}\sigma_e^2$

$$\phi_r = \left[\sum_{i=1}^{l} R_i\delta_i^{\,r-1} + \sum_{i=1}^{l}\sum_{\gamma=1}^{l} \frac{R_i R_\gamma \delta_i^{\,r}}{1-\delta_\gamma\delta_i}\right]\sigma_e^2 \qquad (3.43)$$

By repeating this derivation in other ways,equivalent forms for ϕ_r can be found.

A similar result can be derived for the cross correlation ϕ''_r between the outputs of a system N(z)/D(z) and another system N'(z)/D'(z) each with the same input e_k. Unless the systems were the same the auto-correlation function would not be symmetric i.e. $\phi''_{-r} \neq \phi''_r$. The two possible orderings of the z transforms in (3.42) would then be significant. Equation (3.43) would have either R_γ and δ_γ derived from N'(z)/D'(z) or R_i and δ_i derived from N'(z)/D'(z), depending on the sign of r.

As in section 3.4, $\hat{\not{0}}_r^N$ for a finite data set v_k is defined as $\frac{1}{N-r}\sum_{k=r+1}^{N} v_k v_{k-r}$ and will be equal to the infinite data case if v_k is a stationary sequence derived from an infinite input history. If v_k is generated from onły a finite length input e_k, k=1, ... N , and is therefore non-stationary, then $E(v_k v_{k-r}) \neq E(v_N v_{N-r})$, k=r+1, ... N-1

To calculate $E\hat{\not{0}}_r^N$, $E(v_k v_{k-r})$ can be expanded as $E(v'_o e_k + v'_1 e_{k-1} \;.....\; v'_{k-1} e_1) * (v'_o e_{k-r} + ... v'_{k-r-1} e_1)$. The various terms $E(v_k v_{k-r})$ can be expressed as $\sigma_e^2 \sum_{i=0}^{k-r-1} v'_{i+r} v'_i$. Again cross products of e_k have a zero expectation. The necessary expressions will now be developed by first defining F(z) in a similar way to that before.

$$F(z) \triangleq \sum_{k=0}^{\infty} v'_k v'_{k+r} z^{-k}$$

, where r is the index of $\not{0}_r^N$ and v'_k is the impulse response.

(3.44)

Then using (3.32) we can derive (3.45) in the same say as for (3.35)

$$\sum_{i=0}^{N'-1-r} v'_i v'_{i+r} = \frac{1}{2\pi j}\int_C F(z)H(z^{-1})z^{-1}dz \qquad (3.45)$$

where F(z) is defined above.

H(z) is defined as (3.33), but with the limit N' replaced by N'-r

From the definition of F(z), (3.44) and from the convolution integral (3.31) we can derive the following in the same way as (3.37) to (3.39), and (3.42), (3.43).

$$F(z) = \frac{1}{2\pi j}\int_C p^{r-1}\,\frac{N(p)}{D(p)}\cdot\frac{N(z/p)}{D(z/p)}\,.dp \qquad (3.46)$$

$$= \frac{1}{2\pi j}\int_C p^{r-1}\left(1.0 + \sum_{i=1}^{l}\frac{R_i}{p-\delta_i}\right)\left(1.0 + \sum_{\gamma=1}^{l}\frac{R_\gamma}{zp^{-1}-\delta_\gamma}\right)dp \qquad (3.47)$$

$$F(z) = \sum_{i=1}^{l} R_i\,\delta_i^{r-1} + \sum_{i=1}^{l}\sum_{\gamma=1}^{l}\frac{R_i R_\gamma \delta_i^{r}}{z-\delta_i\delta_\gamma} \qquad (3.48)$$

we can now substitute this in (3.45)

$$\sum_{i=0}^{N'-1-r} v_i'\, v_{i+r}' = \frac{1}{2\pi j}\int_C \frac{1}{z}\cdot\sum_{i=1}^{l} R_i\,\delta_i^{r-1}\cdot\frac{1-z^{N-r}}{1-z}dz + \frac{1}{2\pi j}\int_C \frac{1}{z}\cdot\sum_{i=1}^{l}\sum_{\gamma=1}^{l}\frac{R_i R_\gamma \delta_i^{r}}{z-\delta_i\delta_\gamma}\cdot\frac{1-z^{N'-r}}{1-z}\,.dz \qquad (3.49)$$

1st integral: Residue at z=0. is $\sum_{i=1}^{l} R_i\,\delta_i^{r-1}$

2nd integral: Residue at z=0. is $\sum_{i=1}^{l}\sum_{\gamma=1}^{l}\frac{R_i R_\gamma \delta_i^{r}}{-\delta_i\delta_\gamma}$

Residue at $z=\delta_i\delta_\gamma$ is $\frac{R_i R_\gamma \delta_i^{r}}{\delta_i\delta_\gamma}\cdot\frac{(1-(\delta_i\delta_\gamma)^{N'-r})}{(1-\delta_i\delta_\gamma)}$

Sub. total is $\sum_{i=1}^{l}\sum_{\gamma=1}^{l}\frac{R_i R_\gamma \delta_i^{r}}{\delta_i\delta_\gamma}\;\frac{(1-(\delta_i\delta_\gamma)^{N'-r})}{(1-\delta_i\delta_\gamma)}$

These last two residues add algebraically to give a total:

$$\sum_{i=0}^{N'-1-r} v_i'\, v_{i+r}' = \sum_{i=1}^{l} R_i\,\delta_i^{r-1} + \sum_{i=1}^{l}\sum_{\gamma=1}^{l} R_i R_\gamma \delta_i^{r}\cdot\frac{(1-(\delta_i\delta_\gamma)^{N'-r-1})}{(1-\delta_i\delta_\gamma)}$$

where $r>0$, $N \geqslant r+2$ (3.50)

As before $E\hat{\phi}_r^N$ can be formed by using (3.50) for each $E(v_k v_{k-r})$, k=r+1, ... N and finding the mean.

$$E\hat{\phi}_r^N = \frac{1}{N-r}\left\{E(v_{r+1}v_1) + E(v_{r+2}v_2) \;\cdots\cdots\; +E(v_N v_{N-r})\right\}$$

$$\begin{aligned}
= \frac{\sigma_e^2}{N-r}\Big\{ & v_o' v_r' \\
& +v_o' v_r' + v_1' v_{r+1}' \\
& + \cdots\cdots \\
& +v_o' v_r' + v_1' v_{r+1}' + \cdots\cdots \qquad v_{N-r-1}' \; v_{N-1}' \Big\}
\end{aligned}$$

$$\begin{aligned}
= \frac{\sigma_e^2}{N-r}\Big\{ & \sum_{i=1}^{l} R_i \delta_i^{\,r-1} \\
& + \sum_{i=1}^{l} R_i \delta_i^{\,r-1} \qquad + \sum_{i=1}^{l}\sum_{\gamma=1}^{l} R_i R_\gamma \delta_\gamma^{r} * \frac{1-(\delta_i\delta_\gamma)^1}{1-\delta_i\delta_\gamma} \\
& + \cdots\cdots \\
& + \sum_{i=1}^{l} R_i \delta_i^{r-1} \qquad + \sum_{i=1}^{l}\sum_{\gamma=1}^{l} R_i R_\gamma \delta_\gamma^{r} * \frac{1-(\delta_i\delta_\gamma)^{N-r-1}}{1-\delta_i\delta_\gamma} \Big\}
\end{aligned}$$

Again these terms can be added to give

$$E\hat{\phi}_r^N = \sigma_e^2 \sum_{i=1}^{l} R_i \delta_i^{r-1} + \frac{\sigma_e^2}{N-r}\left[\sum_{i=1}^{l}\sum_{\gamma=1}^{l} \frac{R_i R_\gamma \delta_\gamma^{\,r}}{(1-\delta_i\delta_\gamma)^2} * \left\{(N-r-1)(1-\delta_i\delta_\gamma) - \delta_i\delta_\gamma(1-(\delta_i\delta_\gamma)^{N-r-1})\right\}\right]$$

$$\therefore E\hat{\phi}_r^N = \sigma_e^2 \sum_{i=1}^{l} R_i \delta_i^{\,r-1} + \sigma_e^2 \sum_{i=1}^{l}\sum_{\gamma=1}^{l} \frac{R_i R_\gamma \delta_\gamma^{\,r}}{(1-\delta_i\delta_\gamma)} * \left\{1.0 - \frac{1}{N-r} * \frac{(1-(\delta_i\delta_\gamma)^{N-r})}{(1-\delta_i\delta_\gamma)}\right\}$$

Thus$E\hat{\phi}_r^N$ is biassed as a function of r under the given conditions, but converges to the stationary case (3.43) when N is large.

3.6 Response of system N(z)/D(z).

The relations obtained in the last few sections will be completed by an expression for the response of the system of (3.1) to a unit pulse input at time k=0. We can substitute N(z)/D(z) as in (3.23) into (3.14).
N/D

$$v_o' = \frac{1}{2\pi j}\int_C z^{-1}(1.0 \quad +\sum_{i=1}^{l} \frac{R_i}{z-\delta_i} \quad)dz \qquad (3.51)$$

This again can be treated as the sum of two integrals as before.

1st integral: Residue at z=0. is 1.0 only

2nd integral: Residue at z=0. is $\sum_{i=1}^{l} \frac{R_i}{-\delta_i}$

Residue at $z=\delta_i$ is $\frac{R_i}{\delta_i}$, total residue $\sum_{i=1}^{l} \frac{R_i}{\delta_i}$

These last two terms cancel leaving $v_o' = 1.0$ only

The response $v'_r, r>0$ can be obtained similarly, but there are now no residues at z=0. Thus the full response becomes

$$v'_o \; ; \; v'_1 ; \; \ldots\ldots = 1.0 \quad ; \sum_{i=1}^{l} R_i \; ; \sum_{i=1}^{l} R_i \delta_i \; ; \sum_{i=1}^{l} R_i \delta_i^2 \; \ldots\ldots \tag{3.52}$$

The outputs shown in (3.52) can be summed in various ways as considered in equation (3.34) to give all the relations shown in sections 3.3 to 3.5.

3.7 Representation in terms of roots.

The results of the previous sections demonstrate very clearly the value of expressing N(z)/D(z) of (3.1) in terms of their roots. The structure of the variance relation (3.25), among others, is now clear and simple for all orders of the polynomials. This is in distinct contrast to the text book relations[29-32] given in terms of the coefficients of the polynomials. It will be shown that such a root formulation will be of great value in deriving a measure of the sensitivity of the estimation procedure. Since we can always derive the polynomial coefficients knowing the roots, we will estimate the system (1.38) in terms of its roots, and thereby gain the benefit of being able to predict difficult or unrewarding areas of the estimation process. Even if the original problem was posed in the sense of finding the coefficients, we have chosen a root formulation to make the process easy for ourselves.

As mentioned before there is a need to ensure the stability of the estimated polynomials for later practical use, and to ensure correct bounded filtering of the data set during estimation. A check on the stability of a system when given the polynomial coefficients is not easy and would have to be done at each step of the hill climbing described in section 2.5. Stability is relatively easy to ensure once the roots are known; a simple check will determine whether they lie inside or outside the unit circle in the z plane. We will develop a transformation in a later section which will confine the roots of $\hat{C}(z)$ to lie within the unit circle, and yet allow the hill climbing procedure to operate in an unconstrained space.

3.8 A canonical state representation.

The correspondence between the representation in root form and a canonical or normal form of the state representation of (1.1) can be easily demonstrated. The minimal parameter deterministic version of (1.1), for m=r=1, is given here as (3.53) with F in diagonal form.

$$\underline{x}(k+1) = Fx(k) + Gu(k)$$
$$y(k) = Hx(k) + Du(k) + \mathcal{K} \qquad (3.53)$$

where F is diagonal n.n with n non-zero parameters f_{ii}, i=1, ... n

G or H contain n non-arbitary elements, and are n vectors.

D is 1.1 in this case ; D.C. component $\mathcal{K}$ taken as zero

Equations (3.53) can be expressed in transfer function form[16] (3.54) which can be also obtained by the parallel programming approach[20].

$$y(k) = \left[H(zI-F)^{-1}G + D\right] u(k) \tag{3.54}$$

$$= \left[H \begin{bmatrix} z-f_{11} & & & 0 \\ & z-f_{22} & & \\ & & \ddots & \\ 0 & & & z-f_{nn} \end{bmatrix}^{-1} G + D\right] u(k)$$

$$y(k) = \left[z^{-1}H \begin{bmatrix} \frac{1}{1-f_{11}z^{-1}} & & & 0 \\ & \frac{1}{1-f_{22}z^{-1}} & & \\ & & \ddots & \\ 0 & & & \frac{1}{1-f_{nn}z^{-1}} \end{bmatrix} G + D\right] u(k) \tag{3.55}$$

$$\therefore\ y_k = \left[\frac{h_{11}g_{11}}{1-f_{11}z^{-1}} + \frac{h_{12}g_{21}}{1-f_{22}z^{-1}} \cdots\cdots + \frac{h_{1n}g_{n1}}{1-f_{nn}z^{-1}}\right] z^{-1}u_k + du_k \tag{3.56}$$

$$\therefore\ y_k = \frac{\sum_{i=1}^{n} h_{1i}g_{1i}\left[\prod_{\substack{j=1\\ j\neq i}}^{n}(1-f_{jj}z^{-1})\right]}{\prod_{i=1}^{n}(1-f_{ii}z^{-1})}\ z^{-1}u(k) + du(k) \tag{3.57}$$

We can now compare (3.57) with the root form in D(z), coefficient form in N(z), of the system in (3.1), and by comparing the coefficients of u_k:-

Roots δi of D(z) are given by the f_{ii} diagonal elements of F (3.58)

Coefficients n_i of N(z) are related to H,G and D of (3.53) by:

$$d_o z^o + \left[\sum_{i=1}^{n} h_{1i} g_{i1} . -d \; {}^{n}C_1(f_{ii})\right] z^{-1} + \left[-\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} f_j h_{1i} g_{i1} + d \; {}^{n}C_2(f_{ii})\right] z^{-2}$$

$$+ \left[\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{l=1 \\ l\neq i \\ l\neq j}}^{n} f_{jj} f_{ll} h_{1i} g_{i1} - d \; {}^{n}C_3(f_{ii})\right] z^{-3} \; \ldots\ldots \qquad (3.59)$$

(where h_{1i} or g_{i1} ; i=1, ... n may be chosen to be 1.0 for simplicity and ${}^{n}C_k(f_{ii})$ are all the combinations of f_{ii} taken in groups of k)

Equivalent to

$$n_o z^o + n_1 z^{-1} + n_2 z^{-2} + n_3 z^{-3} + \ldots\ldots \qquad (3.60)$$

As Lindorff[30] has shown, we cannot expect a direct relation for the zeros of N(z). The root form will be retained however in view of the simple forms shown in the previous sections. This canonical normal form of F shows the states of (3.53) in an uncoupled form. This can be useful since the discrete time system could be represented as if it were a sampled continous time system of uncoupled differential equations[20,30], whose eigenvalues are individually related to the f_{ii} elements in F.

3.9 Contour plots of the variance expression.

It is of interest to consider the contour plots of the variance relation (3.26) for an infinite data set, as the poles and zeros of the system (3.1) move over the z plane. For example the contours of equal variance σ_y^2 are shown in figure 1 for the case of a complex pair of poles δ_i, δ_j fixed at z=(+0.2,±j 0.876), radius 0.90 from

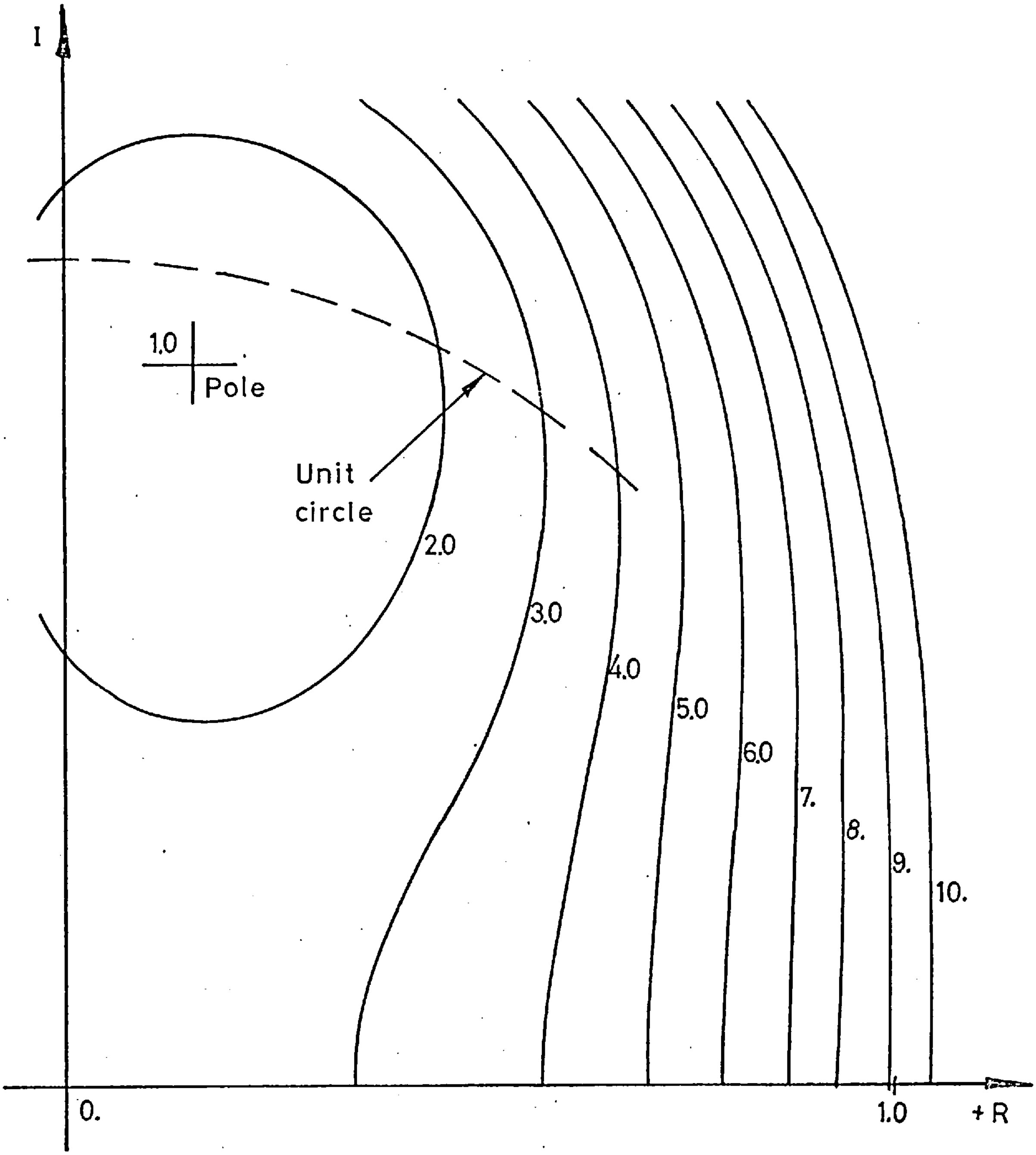


FIG. 1 Isovars for a pair of poles (0.2, ±j0.876)

the origin. A corresponding pair of zeros $\eta_i \eta_j$ are allowed to scan over the portion of the z plane shown. The variance contours will be referred to as "ISOVARS". This term has been derived from a mixture of Greek and Latin roots, and is more concise than the fully Greek version "isoataktos".

The isovars of figure 1 are symmetric about the real axis. Similar patterns would arise if the fixed poles were in the negative half of the z plane. The contours are smooth and appear quadratic about the pole positions with a saddle point on the real axis. It should be noted that the zeros can be outside the unit circle without introducing discontinuities. The figure on each isovar is the value of the variance σ_y^2 on that contour to a base of $\sigma_e^2 = 1.0$, where σ_e^2 is the variance of the input signal e_k in (3.1). Naturally when the zeros coincide with the poles, complete matching has occurred leaving the term $1.0\,\sigma_e^2$ in (3.26). If we now move the fixed poles closer to the unit circle, the contours close in towards the matching point. This is due to the effect of the $\frac{1}{1-\delta_i\delta_y}$ terms in (3.26); thus as $/\delta_i/ \rightarrow 1.0$, any pole-zero mis-match components left in R_i and R_y are multiplied by larger factors. We can qualititatively assign a "strength" to a pole, depending how close to the unit circle it is.

The inverse case to figure 1 is shown in figure 2. Here a complex pair of zeros are fixed at z=(0.20, ± 0.876), and a pair of poles allowed to scan over the z plane. The poles cannot lie outside the unit circle without giving an unstable system, and an infinite variance σ_y^2 for an infinite length data set. Consequently

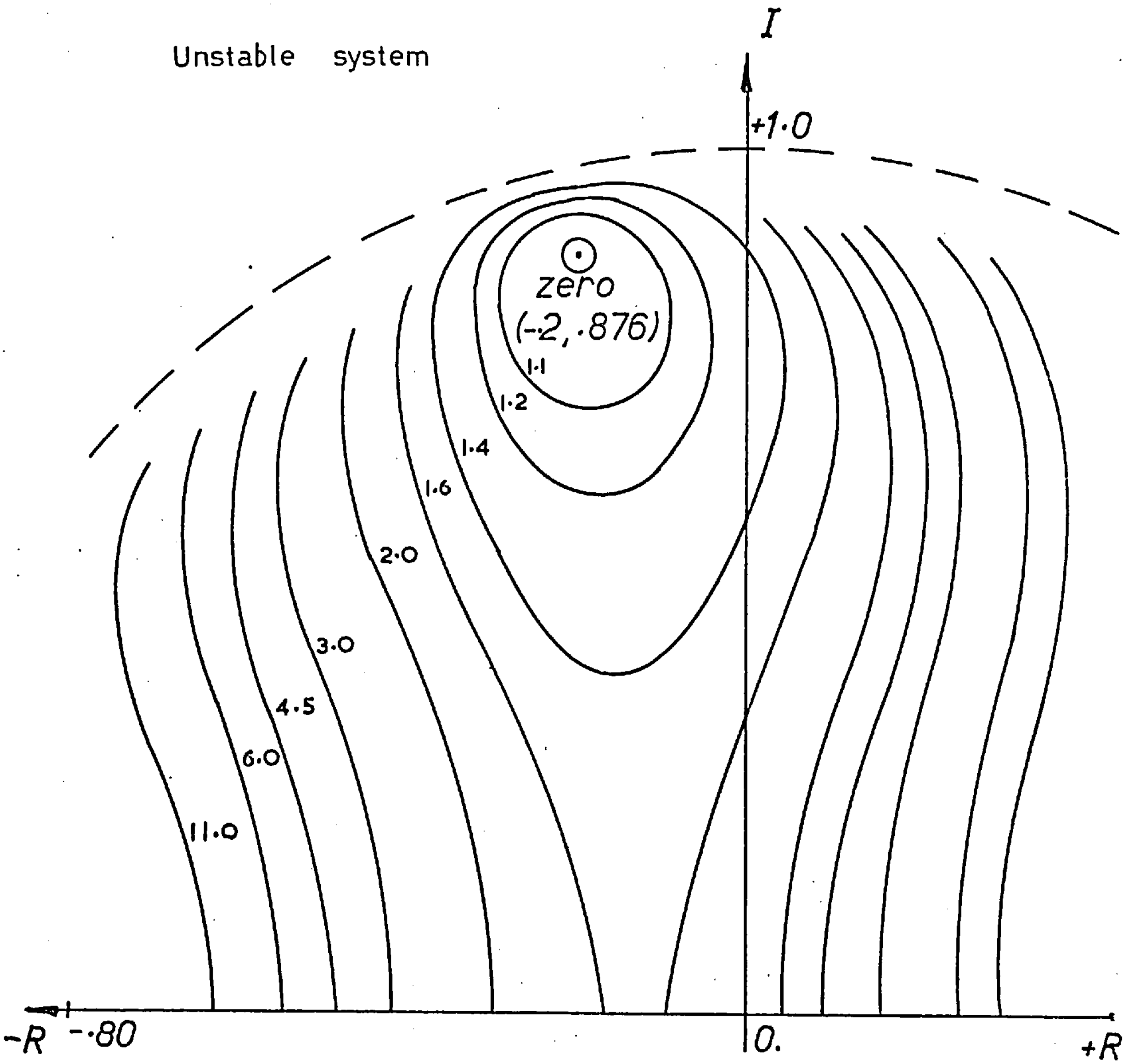


FIG. 2 Pair of fixed Zeros

all the isovars of all values, $1.0<\sigma_v^2<\infty$, must pass between the position of the zero and the unit circle. Thus the gradient in this region is large and quickly changing. The hill surface probably cannot be easily matched by an analytic surface, such as a quadratic, except in very small areas. We could not expect a normal hill climbing routine to work efficiently under these conditions. The situation becomes increasingly worse if the zeros of the system lie closer to the unit circle.

It is possible, in special cases only, to constrain the poles within the unit circle by restraining the coefficients for low order polynomials. The constraints may appear linear in certain terms[43,44] and allow simplifications to be made. These methods are clearly not universal, and cannot be applied to higher order systems. Unless the inherent pole structure as shown in (3.26) is accepted and used, we cannot hope to solve more than a few special cases of restricted interest

In the sense of section 2.8 we will consider the estimation process as one of selecting filter dynamics, i.e. poles and zeros, to give a minimum cost $V(\underline{\hat{\theta}})$ of the e_k sequence from equation (2.49). The estimation process thus reduces in this chapter to matching the poles and zeros of a filter to those of a fixed system. As a result we are concerned with the sensitivity of $V(\underline{\hat{\theta}})$, or the variance of the filter output, to the variable terms describing the dynamics of the filter.

3.10 Transformation to restrain poles.

A number of efficient hill climbing routines, such as Newton-Raphson or Fletcher-Powell, do not easily handle constraints. It would therefore be of advantage to us, if the optimising routine could operate in an unconstrained 2 dimensional space X, each point of which would relate to a point on the Z plane constrained within the unit circle. The transformation suggested is that the radius R_x of a point in X space would be operated on by a saturation function to produce a point at a radius R_z in the Z plane whose maximum value would be 1.0. The corrosponding angles θ_x and θ_z could have the same value. Any point in the infinite domain of X transforms according to these rules, into the finite range in the Z plane represented by the unit circle. If we consider only the set of points inside the unit circle in the Z plane, then the transformation is one to one. The isovars of figure 2, when expressed in the new space X, would be stretched out to cover the whole extent of X space.

A suitable saturation function has to be selected. Some of the known ones are shown in figure 3a. The curves have all been normalised to saturate at 1.0, and to pass through the point R_x,R_z = (1.00,0.7616). The 1st,2nd and 3rd derivatives are also given in figures 3b, 4a and 4b, where it can be seen that the function Tan^{-1} is the least smooth of the set. The function $R_z = \text{Tanh}(R_x)$ was arbitrarily chosen as the transformation to be used although the alternative Erf or Exp. functions would be equally valid. The Tanh function was also used later for a similar purpose by Shaw and Robinson[14] but with a different intention.

Figure 5 shows that the isovars drawn in the new space X now

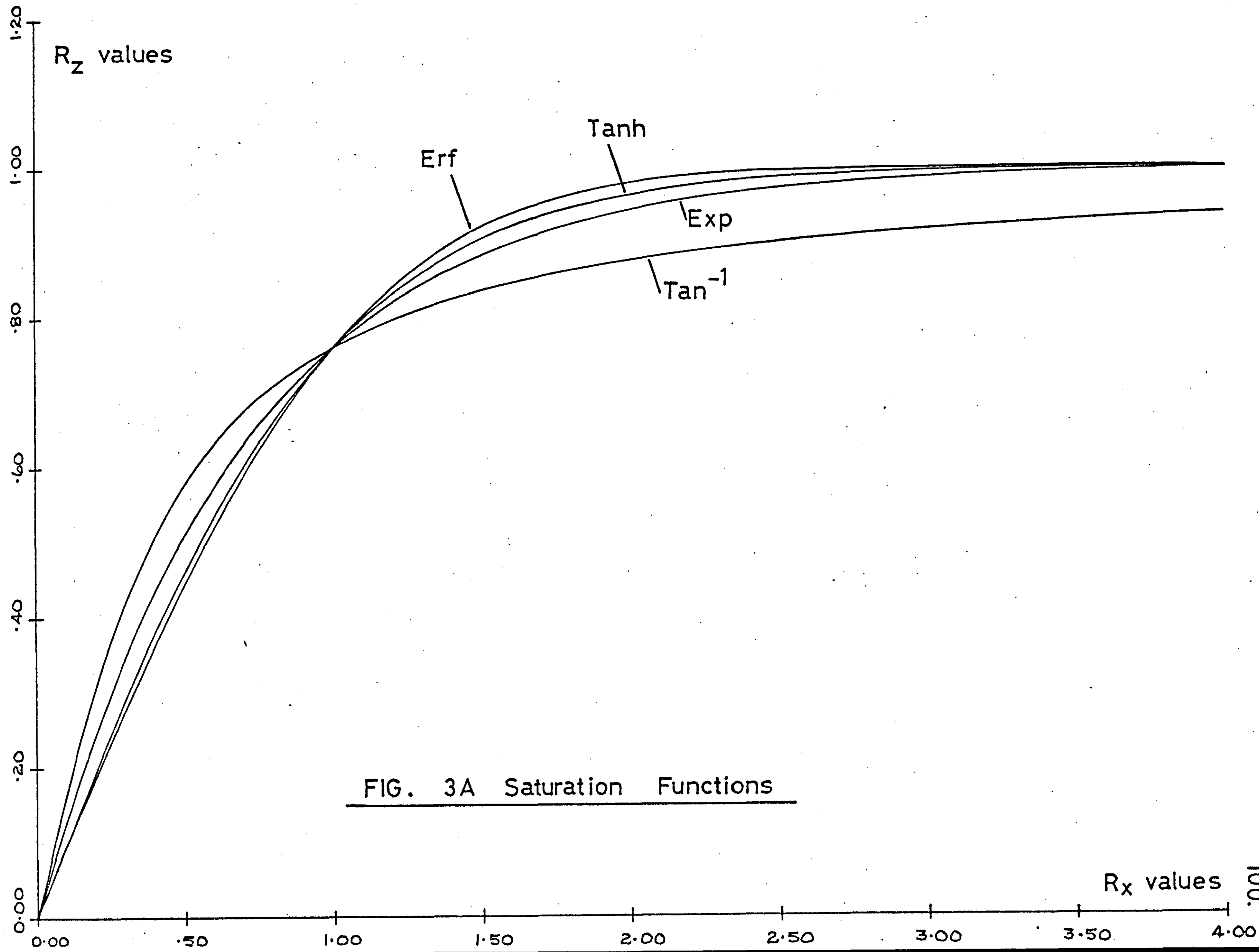


FIG. 3A Saturation Functions

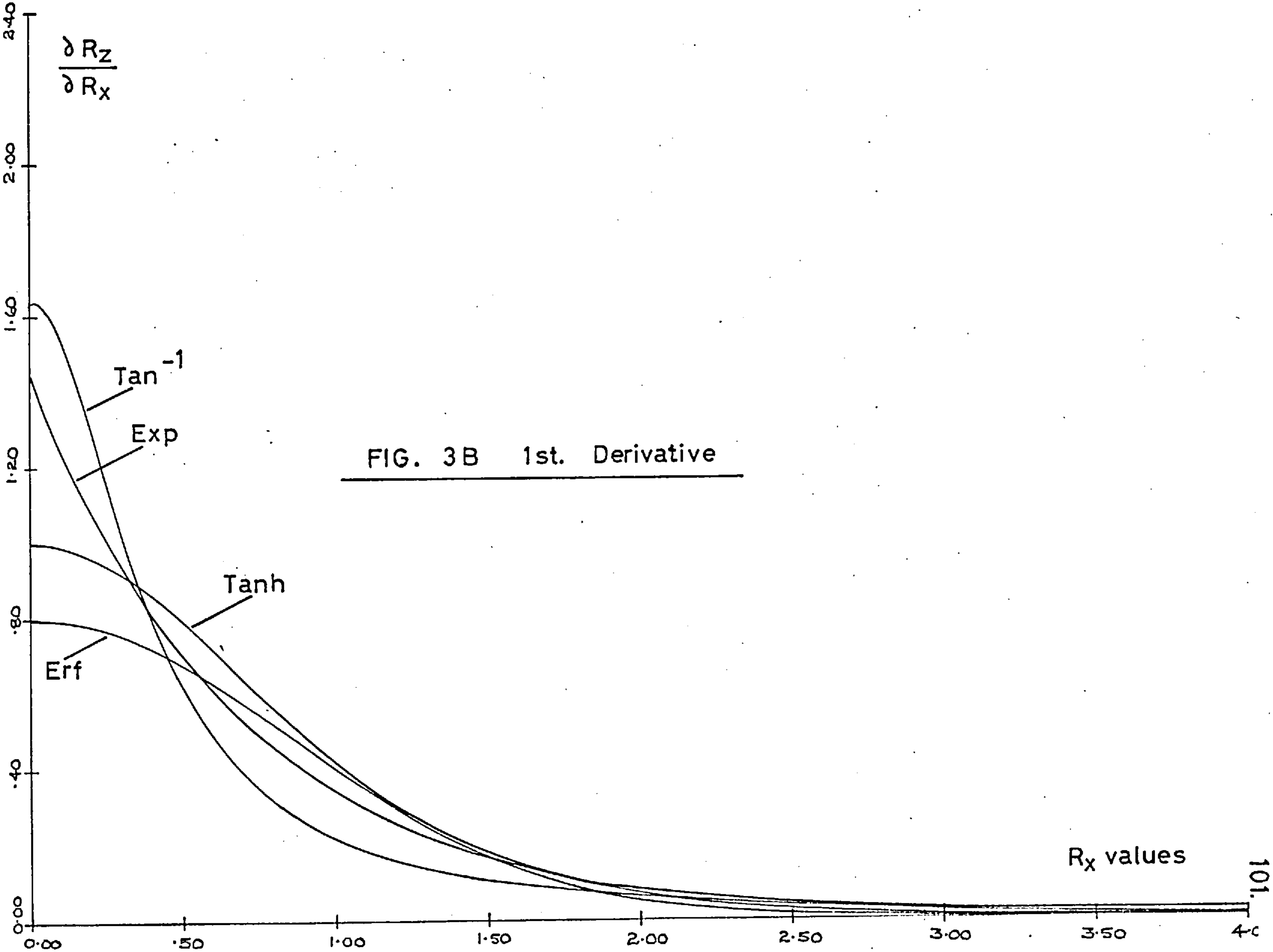


FIG. 3B 1st. Derivative

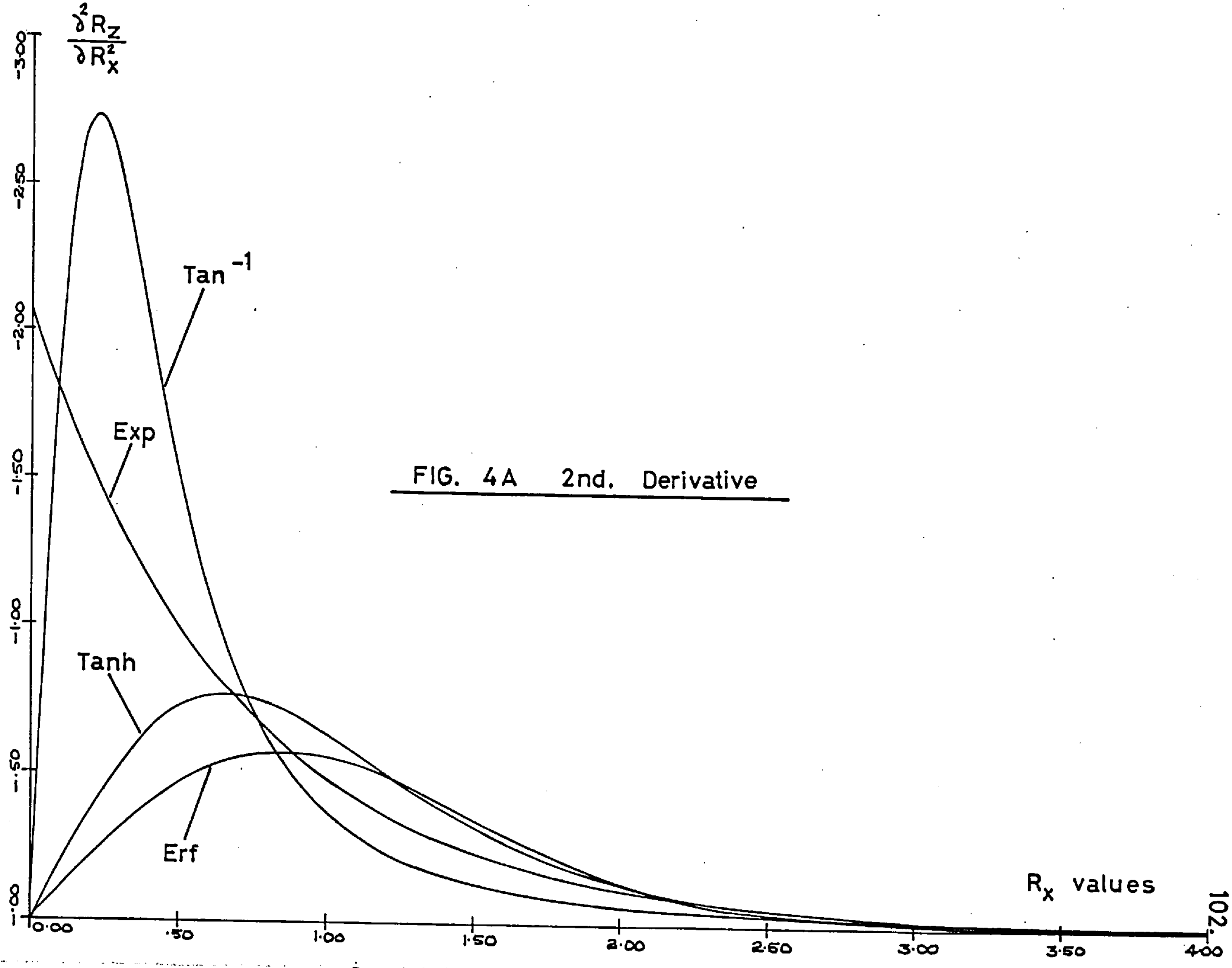


FIG. 4A 2nd. Derivative

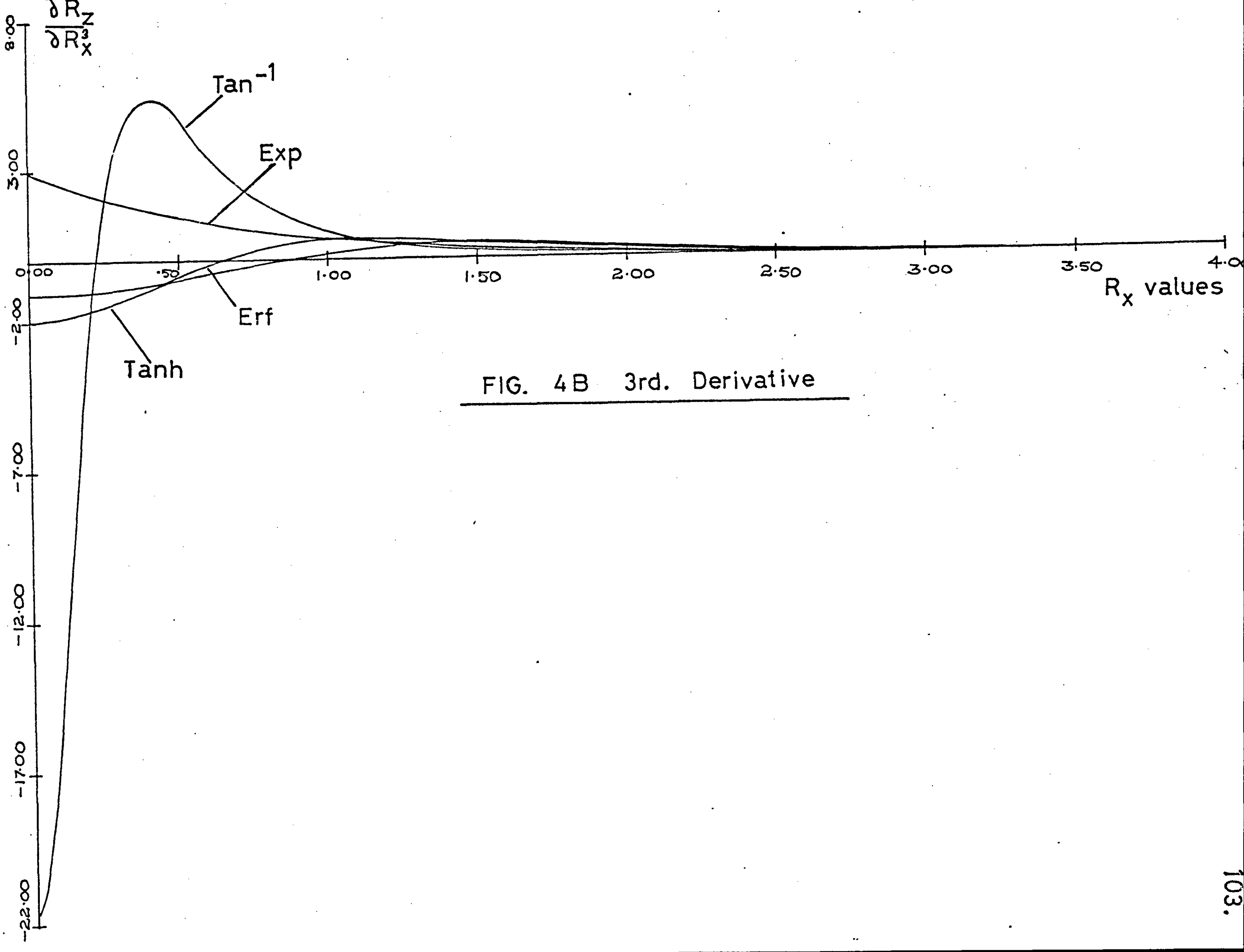


FIG. 4B 3rd. Derivative

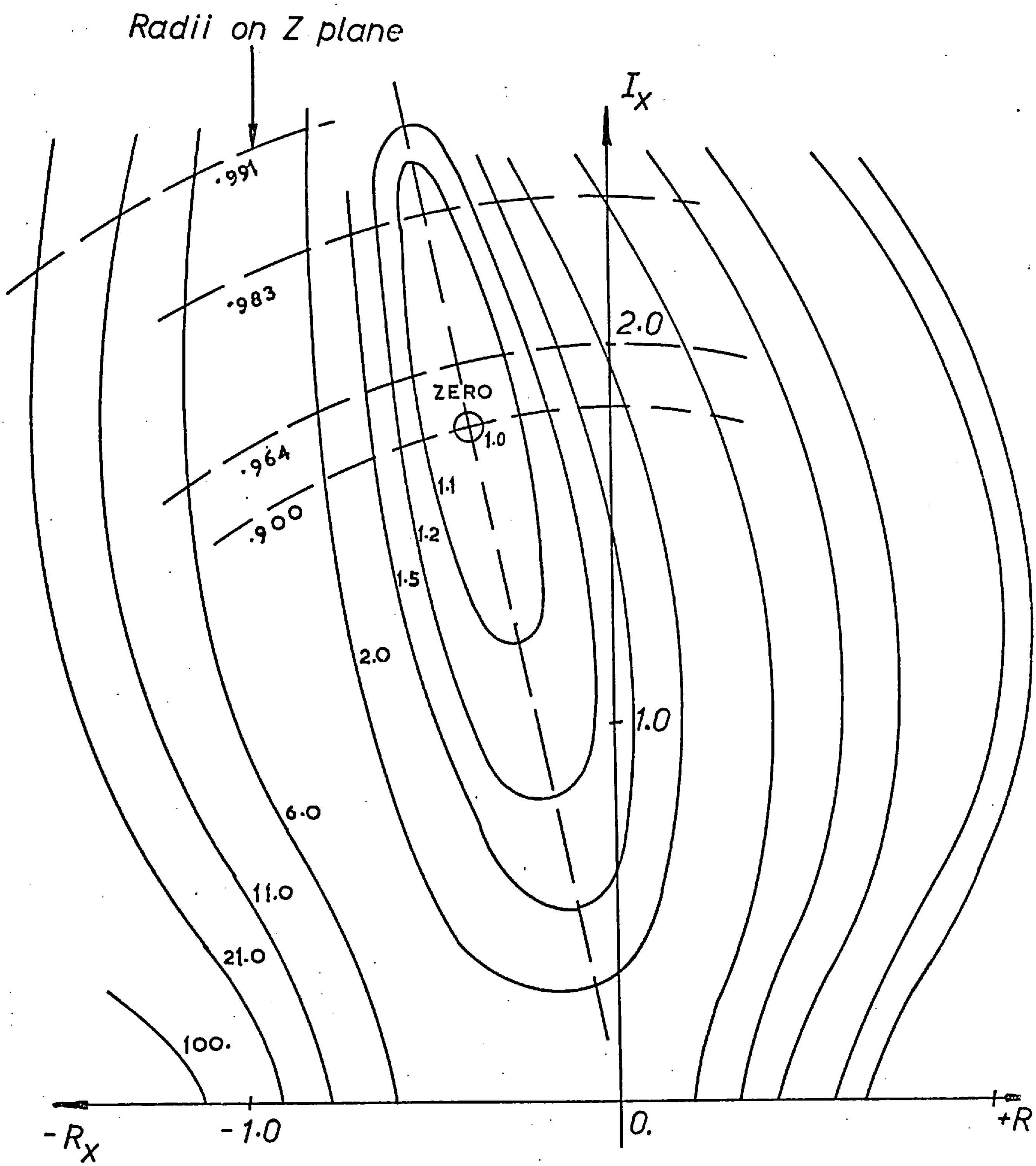


FIG. 5 Two fixed Zeros in X space

appear as elipses instead of the more tightly circular contours of figure 2. The gradient of the surface is now considerably less, and the whole hill appears smoother and more easily fitted locally by quadratic approximations. Figure 5 can be redrawn for various positions of the fixed zeros. The aspect ratios of the elipses surrounding the optimum increase as the zeros approach the unit circle. It will be noticed that the hill is almost symmetric about a radial as shown, and it would be natural to describe it in polar coordinates.

When the number of poles increases it is then more convenient to redefine X space in the sense that one dimension of X' space is reserved for each pole radius, with extra dimensions in X' being used to describe the angles of complex pole pairs. Thus X' space has the same dimensionality as the number of poles, while the Z plane is restricted to the usual two dimensions. The transformation now gives n points on the two dimensional Z plane corrosponding to one point in the n dimensional space of X', as defined by (3.61)

$$x \in X'^{(n)} \longrightarrow \delta_1, \delta_2 \;\ldots\ldots,\; \delta_n \in Z^{(2)} \tag{3.61}$$

$$\begin{aligned} \text{where}\quad /\delta_1/ &= \mathrm{Tanh}(x_1)\ ; \quad \mathrm{Ang}(\delta_{p+1}) = x_{r+1} \\ &\;\;\vdots \qquad\qquad\qquad\qquad \vdots \\ /\delta_r/ &= \mathrm{Tanh}(x_r) \qquad \mathrm{Ang}(\delta_r) = x_n \\ r &\triangleq \text{number } p \text{ of real poles + number of complex pole pairs} \\ \delta_{r+i} &= \text{complex conjugate of } \delta_{p+i},\ i=1,\ \ldots\ldots,\ n-r \\ n &= \text{total number of poles.} \end{aligned}$$

Multiple poles are covered by the definition (3.61), since the corrosponding components in X' space are quite distinct. An extra pole may easily be added to the arrangement as required. This merely increases the dimensionality of X', while leaving all other components and poles alone.

A hill climbing routine can now work in the n dimensional unconstrained space of X', while the system poles are constrained to move only within the unit circle on the two dimensional Z plane. Negative values of all the x_i components of X' are permissable since both the Tanh function and angular measure are odd functions. Values of angle greater than 2π radians simply cover again the same area of Z space as angles less than 2π. Hence there will be more than one solution in X' space corrosponding to a single Z plane configuration. This is in general of no concern as most hill climb procedures only climb to the local optimum, and all the optima in X' space will give identical solutions in the Z plane, of equal cost.

The only remaining consideration is that of uniqueness of the solution in the Z plane, and is part of the estimation theory alone. Naturally we may exchange pole positions in the Z plane with no effect on the estimation cost or model behaviour. This means that there will be again more than one point in X' space, for a single solution in the Z plane, but this is of no importance due once more to the local ability of the hill climbing routine. All of the optima in X' space will give identical solutions in the Z plane.

It will be seen later that on occasion we will want to freeze the radius of certain poles at less than 1.0 in the Z plane. To cover such a situation it is only necessary to fix one component of X' space, and continue to climb in the subspace remaining.

Because the transformation of (3.61) is defined in terms of analytic fuhctions, we can obtain the derivatives in X' space of the cost hill, if we can calculate the corrosponding gradients in the Z plane. This process will be described in more detail in Chapter 4. For the moment, it is enough that we have the information of the cost and its 1st derivative in the new X space, both being derived from pole-zero positions in the Z plane via the suggested transformation. We can therefore now employ fairly sophisticated minimisation routines[8,9] which work best in an unconstrained space, to solve the estimation problem.

3.11 More complex configurations.

Figures 6, 7 and 8 show the isovars for the case of three fixed poles and the system as (3.1); one complex pair is at (-0.20±j 0.876) and a third at (-0.980,j 0.0) in the Z plane. There is no local effect on the isovars of a complex pair of variable zeros $Z_{1,2}$ due to the pole P_3 or the pair $P_{1,2}$. The only effect is a global one on the hill system, which is a distinct contrast to the well known root locus plots also involving pole and zero positions. As the third zero Z_3 moves from the origin towards P_3 in successive figures, the hill system for the pair $Z_{1,2}$ changes from a single optimum on the real axis, to a double optimum near the poles $P_{1,2}$. (All these figures are symmetric about the real axis and a 'pair' is understood to be a complex conjugate pair with equal positive, and negative imaginary parts.)

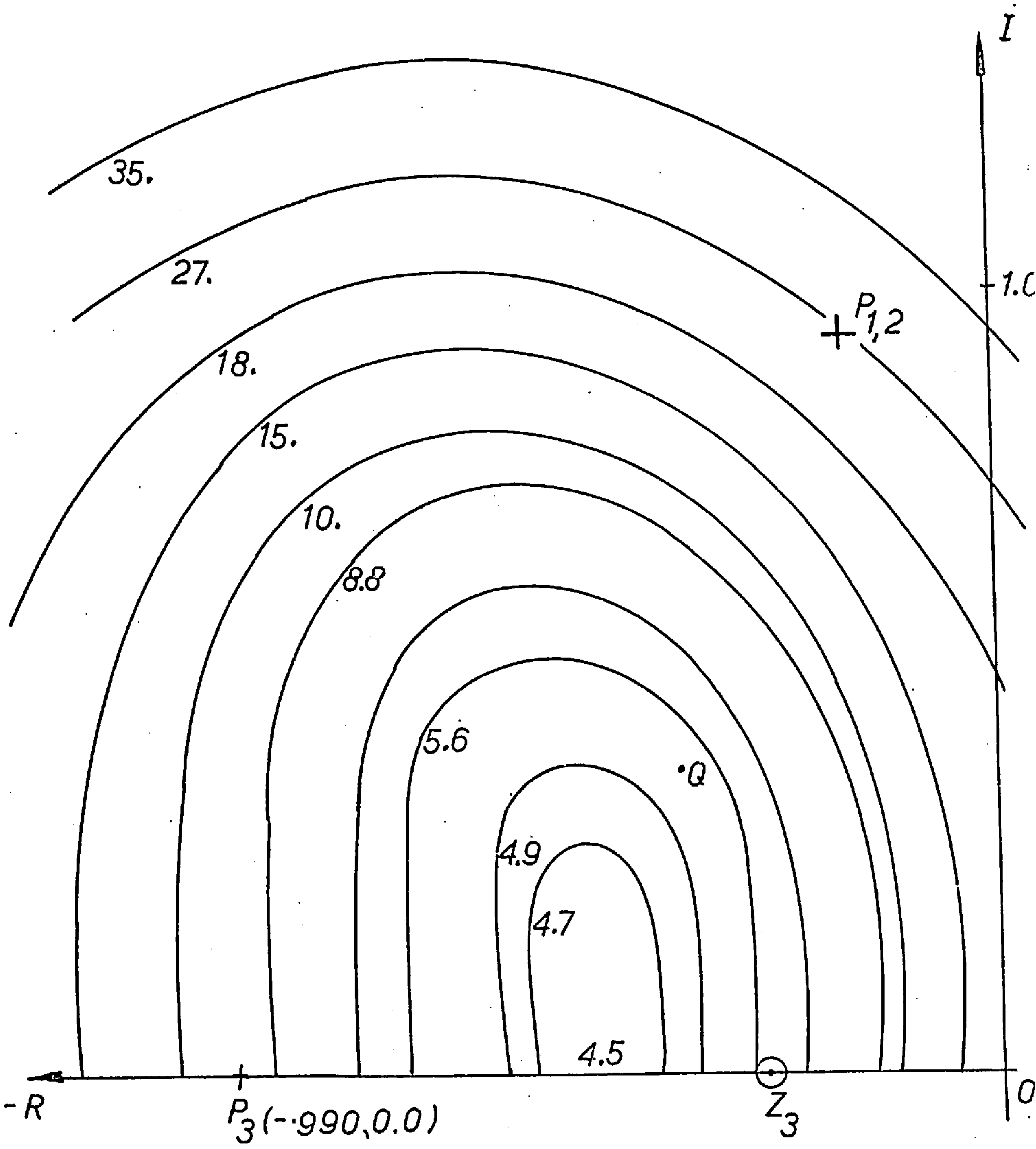


FIG. 6 Three fixed poles, Z_3 at -0.3

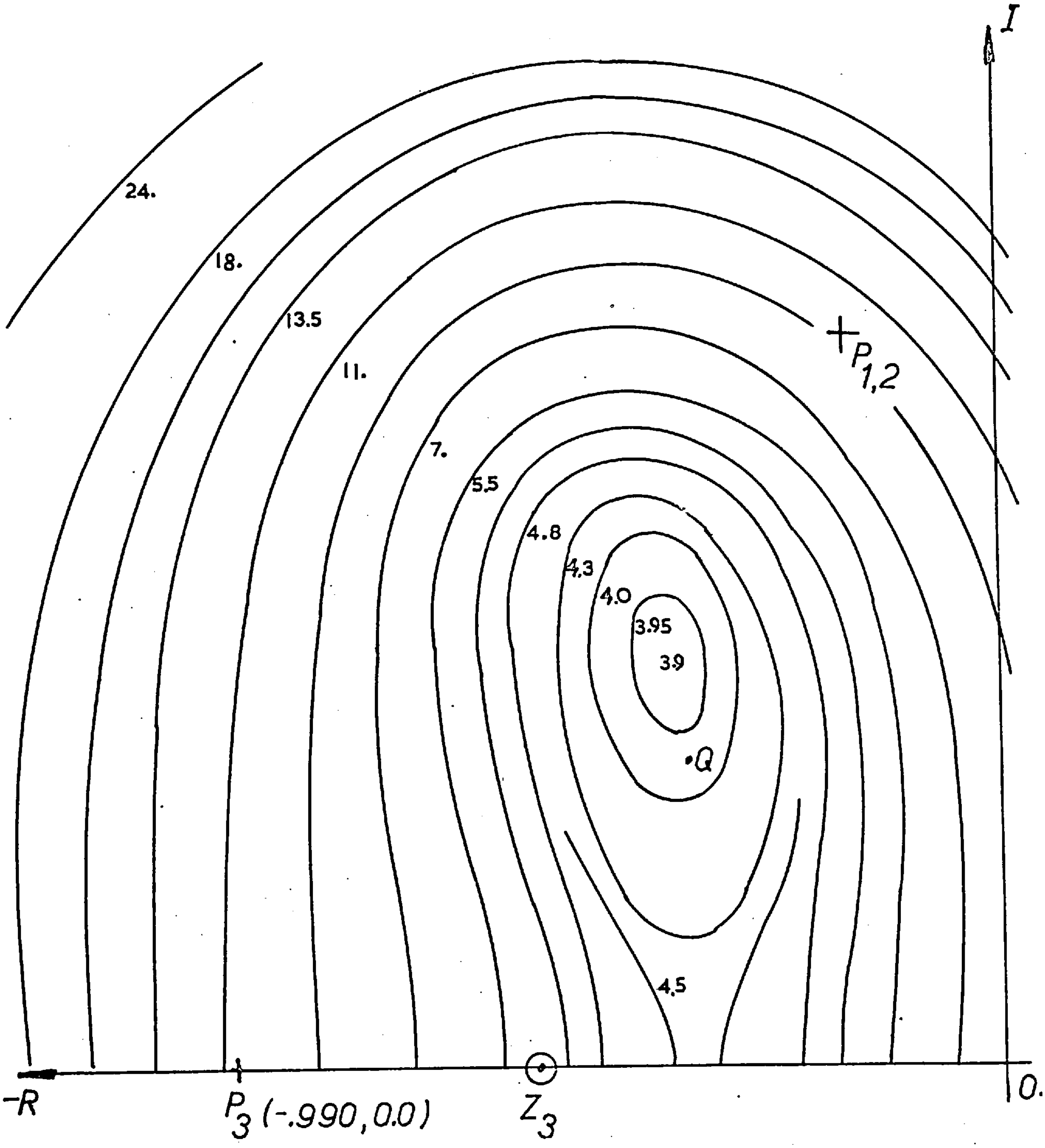


FIG. 7 Three fixed poles, Z_3 at -0.6

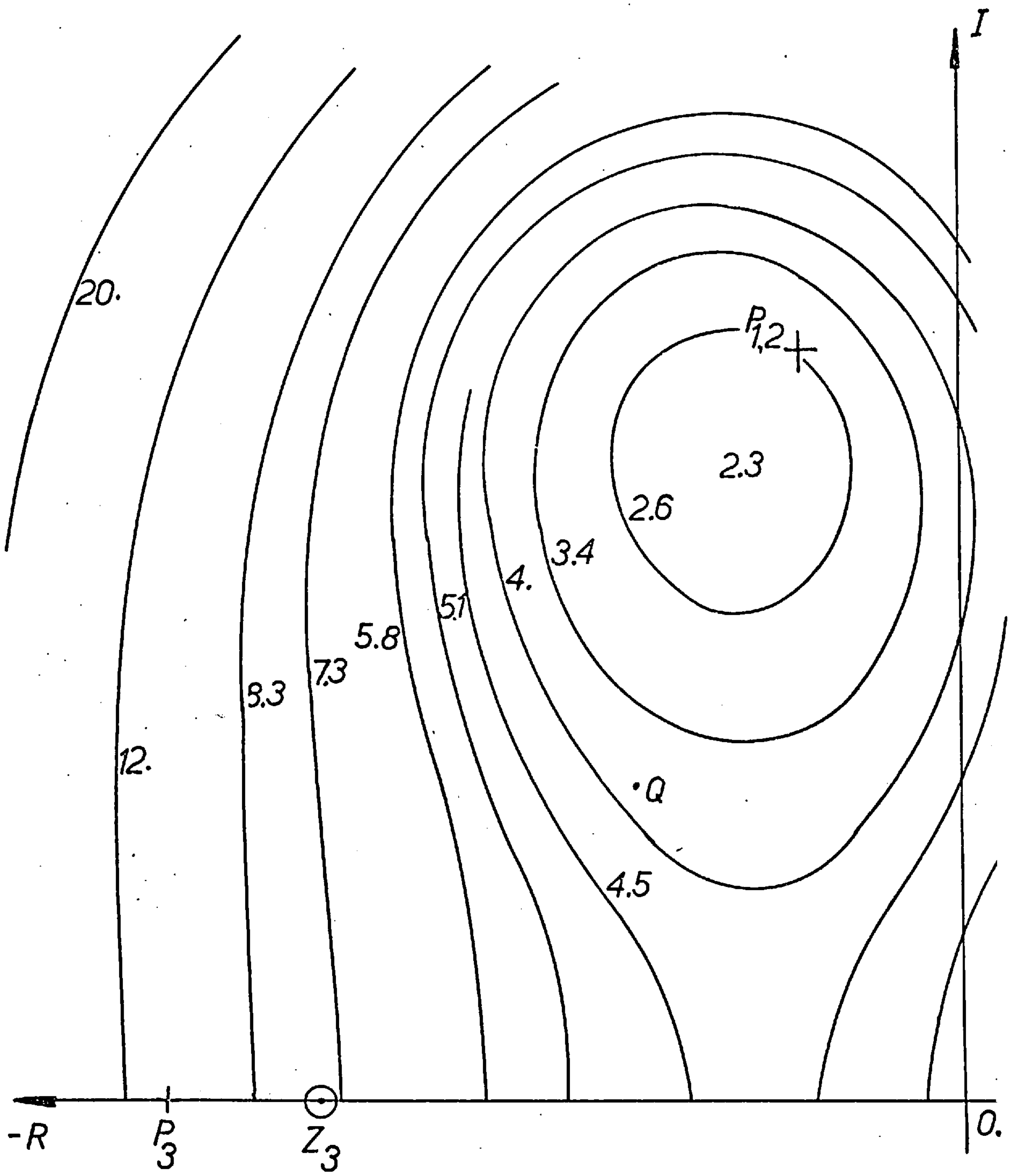


FIG. 8 Three fixed poles, Z_3 at -0.8

From a parameter estimation viewpoint $P_{1,2}$ cannot be correctly estimated unless P_3 has been reasonable well matched by Z_3. If we were estimating the poles $P_{1,2}$, and P_3, by matching them with the zeros, it implies that the positions of $Z_{1,2}$ and Z_3 are coupled. This estimation procedure can be interpreted as climbing a curved ridge hill in 3 space, one dimension being assigned to each of the degrees of freedom of $Z_{1,2}$ and Z_3. The altitude of the hill can be considered as the cost in a fourth dimension. Another requirement, that of following curved ridges has therefore been placed on the optimisation routine.

By detailed and repeated reference between figures 6,7 and 8, it is possible to see[45] that if the positions of $Z_{1,2}$ and Z_3 are separately and independantly optimised, then all three zeros will finish near point Q on the figures, The separate climbing of each can be iterated as an attempt to improve the estimation. As each zero will have its own respective local optima near Q, the estimation process will be trapped in a ridged hill situation. Only when the climb procedure couples $Z_{1,2}$ and Z_3, can a search be made along the ridge direction, and the absolute minimum be found. Advanced routines such as Fletcher and Powell[9] do have this ability to choose new search directions other than the co-ordinate axes of the space.

The situation for two pairs of poles is shown in figure 9. The isovars corrospond to the motion of two variable filter zeros $Z_{1,2}$ moving on the Z plane, while another pair $Z_{3,4}$ remain fixed as shown. An exactly similar isovar pattern applies if $Z_{1,2}$ were fixed and $Z_{3,4}$ were in motion. This gives rise to a similar climbing situation as the one in the previous case. Both pairs of zeros see local optima and

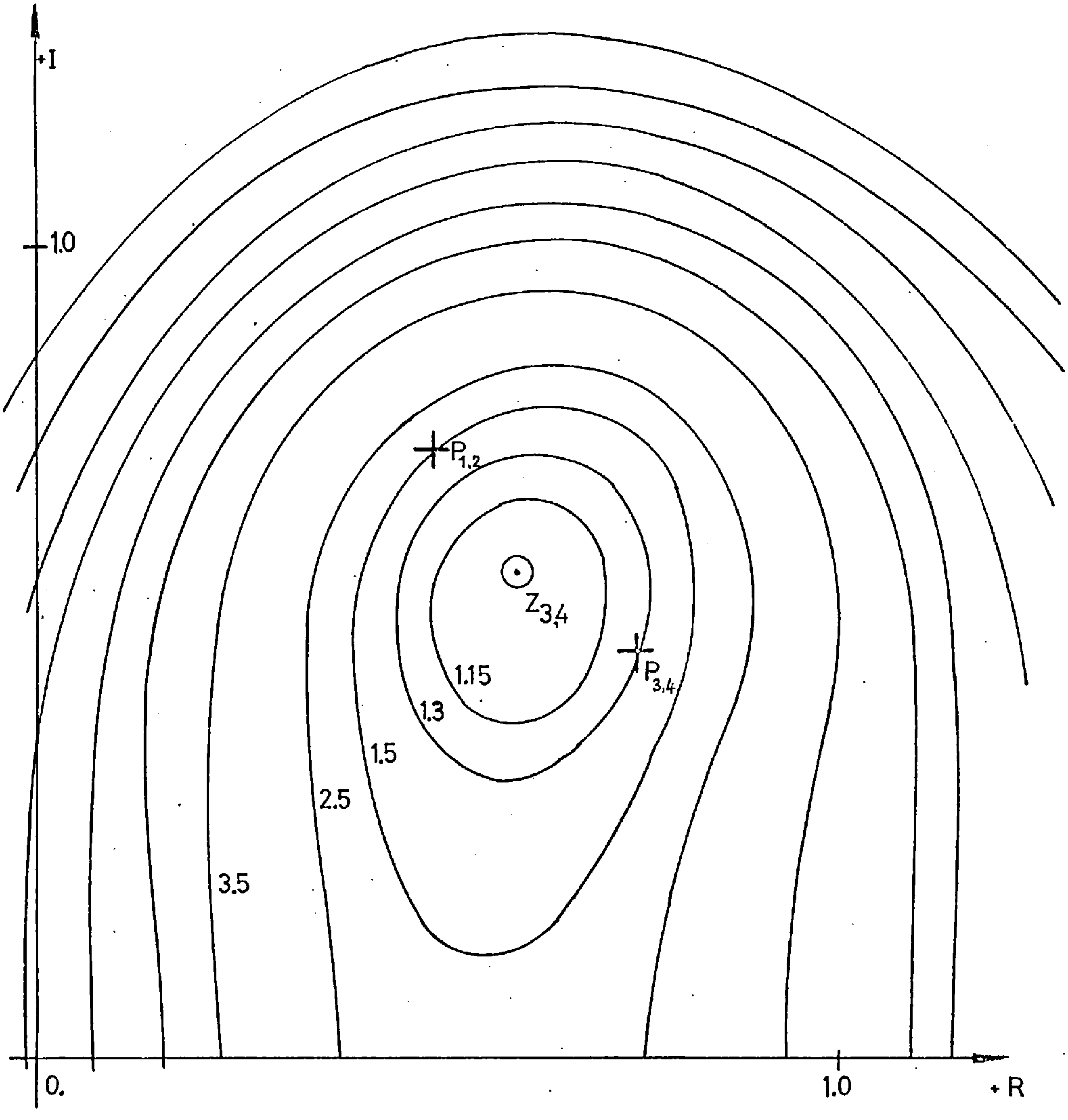


FIG. 9 Two pairs of poles

cannot escape by independant optimisation. Only if the coupling, i.e. the curved ridge hill, were recognised, could the correct estimate of the poles be achieved.

The inverse situation is shown in two dimensional X space in figure 10. Here we have two pairs of fixed complex zeros $Z_{1,2}$ and $Z_{3,4}$ with one complex pair of poles $P_{3,4}$ temporarily fixed, and one pair $P_{1,2}$ variable over the X plane, giving the isovars shown. Again a local optimum effect exists for independant climbing around the position of $P_{3,4}$ which can only be resolved by again recognising the curved ridge situation.

A third alternative is shown in figure 11 for completeness. One complex pair of zeros $Z_{3,4}$ and a complex pair of poles $P_{3,4}$ are fixed on the Z plane and hence in X space. The isovar plot is given for the variable poles $P_{1,2}$ with the zeros $Z_{1,2}$ temporarily fixed as shown. Besides the usual ridge hill situation, it is clear from this and the other figures that the variable poles and zeros do not anihilate each other to the extent of causing any discontinuity in the contours, i.e. such a condition does not give any local change in the smoothness of the cost surface.

3.12 Breakdown of the transformation.

Given a system model as in (3.1) and a data record v_k,k=1 ..... N, the parameters of the model might be estimated by adding a filter which was an approximate inverse of the system. The input to the filter would be v_k, and the parameters of the filter could be adjusted until the variance of the output signal e_k of (3.62) was minimised.

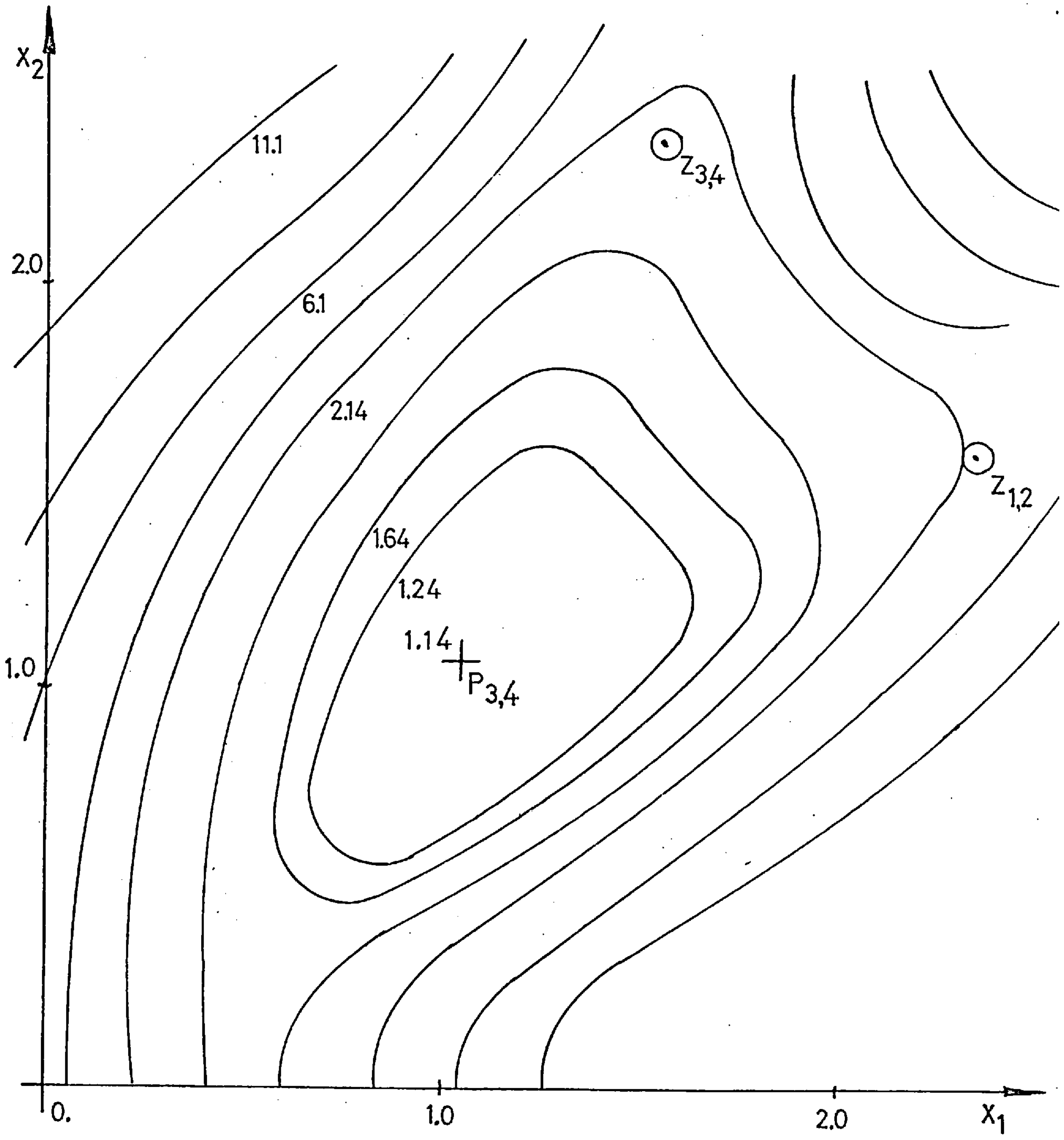


FIG.10 Two pairs of fixed Zeros in X space

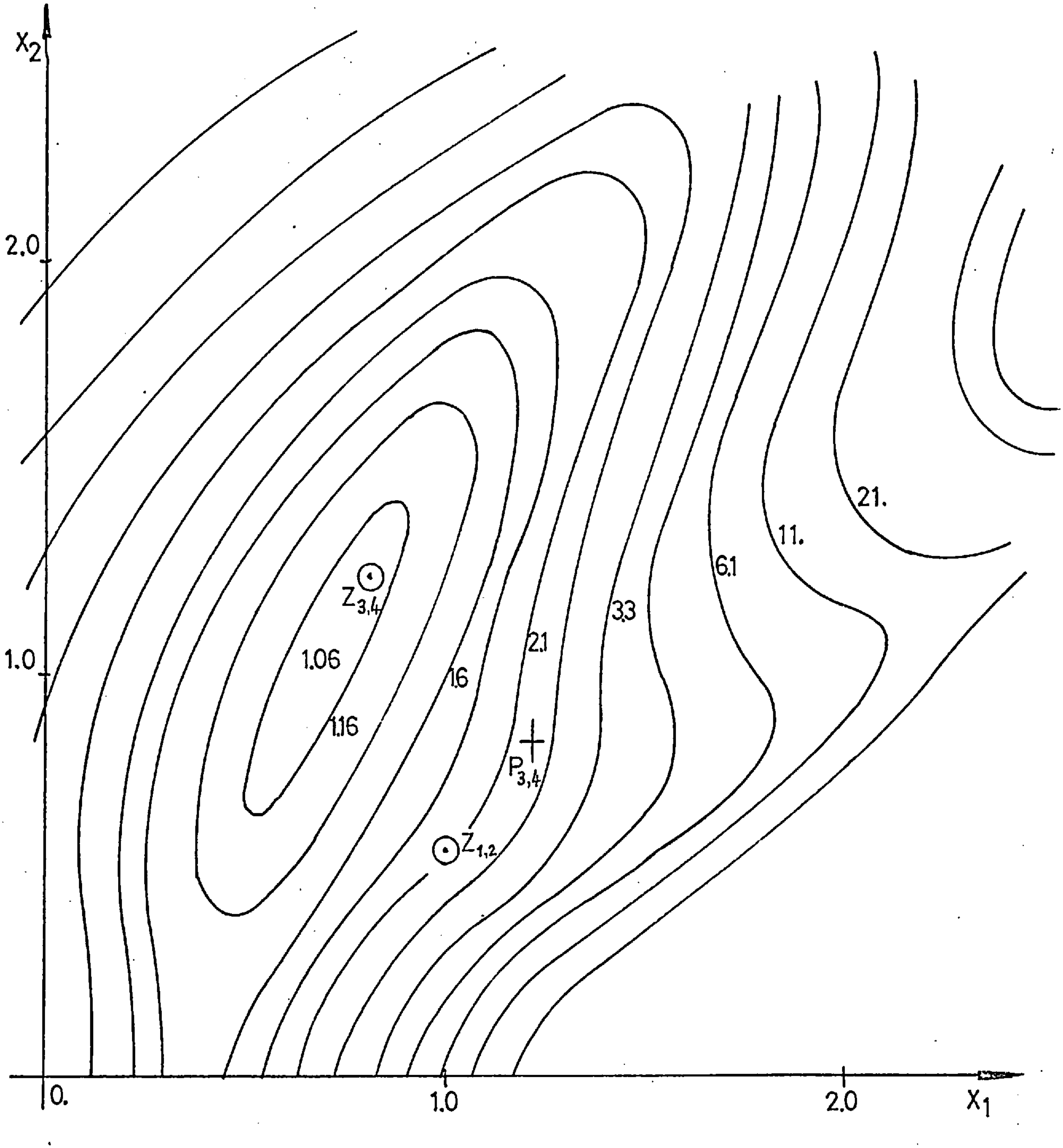


FIG. 11 Pair of fixed poles and a pair fixed zeros in X space

$$\hat{\epsilon}_k = \underbrace{\frac{\hat{D}(z)}{\hat{N}(z)}}_{\text{filter}} v_k = \frac{\hat{D}(z)}{\hat{N}(z)} \cdot \frac{N(z)}{D(z)} \epsilon_k \qquad (3.62)$$

Then $\hat{\delta}_i$ would be an estimate of δ_i, and $\hat{\eta}_i$ of η_i. The stability of the $\hat{\eta}_i$ roots of $\hat{N}(z)$ has to be ensured to make the sequence $\hat{\epsilon}_k$ bounded. The arrangement of (3.32) is very similar to the estimation procedure described in section 2.8. A data sequence v_k,k=1, ... N is filtered to give a signal $\hat{\epsilon}_k$ of residuals. For correct parameter estimation, the residuals should be white, i.e. independant, and minimum variance.

Since we have only a finite data set of length N, we ought to use the expression in (3.41) to calculate the expected sample variance of ϵ_k in order to investigate possible sensitive regions. The X' transformation described by (3.61) can be used to control the roots $\hat{\eta}_i$ within the unit circle on the Z plane. Let us apply this transformation to a simple 1st order process shown in (3.63).

$$\hat{\epsilon}_k = \frac{(z-\hat{\delta}_1)}{(z-\hat{\eta}_1)} \cdot v_k \qquad (3.63)$$

where v_k is for the moment taken as a white sequence;

$E(v_k v_{k-i}) = \sigma_v^2 \cdot \delta(i)$; $\delta(i) = 1.0$ for $i=0$; 0.0, $i\neq 0$

From (3.41), the expected sample variance of the $\hat{\epsilon}_k$ sequence is

$$\sigma_{\hat{e}}^2 = \left\{ 1.0 + \frac{(\hat{\eta}_1 - \hat{\delta}_1)^2}{1-\hat{\eta}_1^{\,2}} \left[1.0 - \frac{1-\hat{\eta}_1^{\,2N}}{N(1-\hat{\eta}_1^{\,2})} \right] \right\} \sigma_v^2 \qquad (3.64)$$

The transformation that we have described, was used to offset the effect of the term $\frac{1}{1-\hat{\eta}_1^{\,2}}$ in (3.64) e.g. figure 5. For a stable filter of the type of (3.63), $/\hat{\eta}_1/$ must be less than 1.0 and this must also be true for all poles of higher order filters. As $\hat{\eta}_1$ approaches the unit circle, with N finite, the magnitude of $\hat{\eta}_1^{2N}$ becomes significant compared to 1.0. Some idea of the size of this effect can be found from figure 12. Here the magnitude of $\hat{\eta}_1^{2N}$ has been plotted for various values of N, as $\hat{\eta}_1$ approaches the unit circle.

Equation (3.64) can be described in the transformation X space, as figure 13, which was drawn for η_1 equal to 0.998. The independant variable X, single dimension, gives the value of $\hat{\eta}_1$ along the real axis in the Z plane. The optimum value of $\hat{\eta}_1$ lies at X=3.45, which corrosponds through the transformation to $\hat{\eta}_1$ matching η_1 at a value of 0.998, and is the same solution which would be obtained for infinite data. It will be noticed that for N=∞, the curve appears convex, and the cost $\sigma_{\hat{e}}^2 = \frac{1}{N}\sum_{k=1}^{N} \hat{\epsilon}_k^2$ should prove to be easily minimised using standard optimisation routines. The corrosponding curve for N=200, however, shows non-convex behaviour away from the optimum. For a higher order filter, the space shown in figure 13 would be of higher dimension and non-convex, and the second derivative matrix would no longer appear positive definite. This fact would reduce the computational

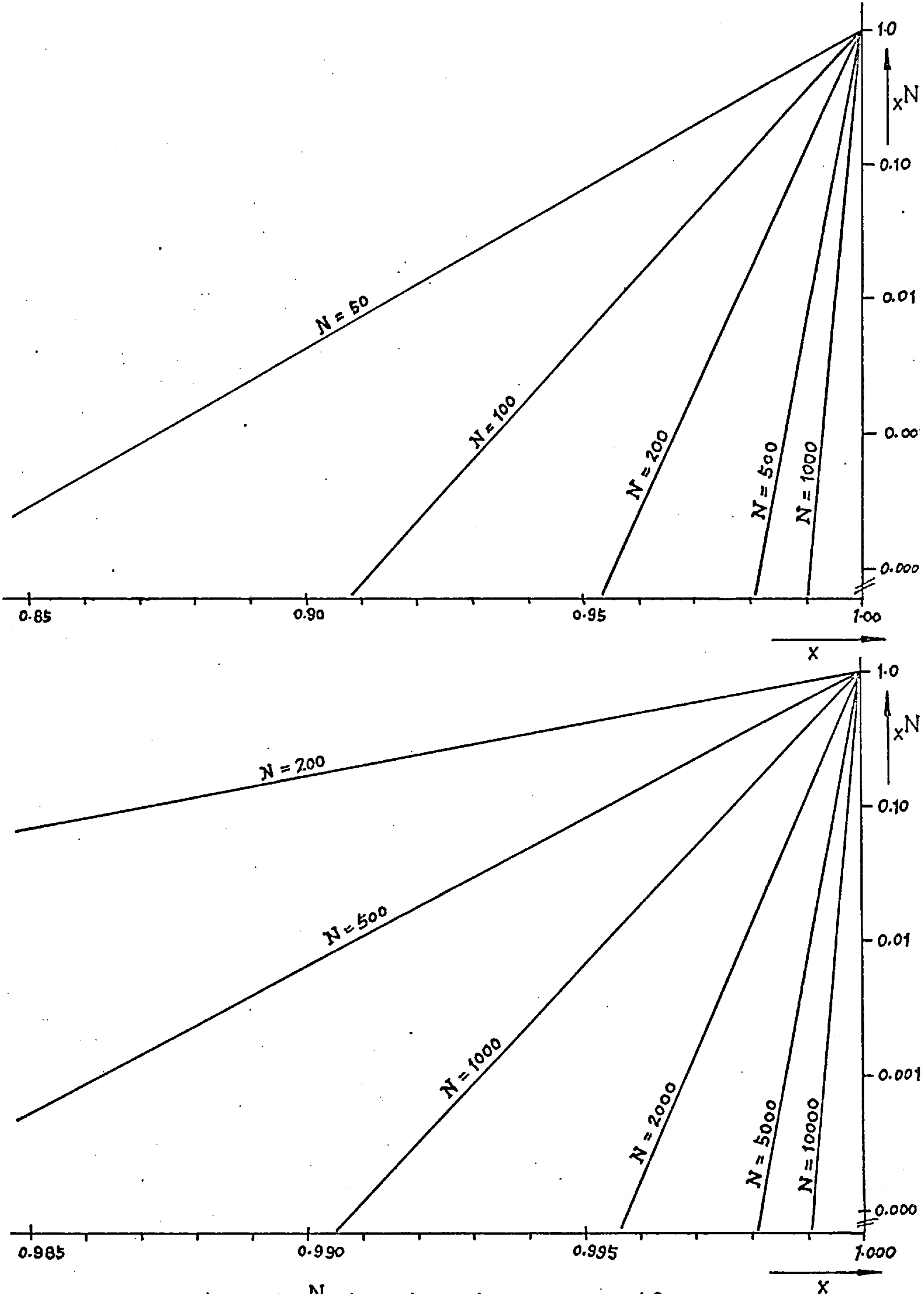


FIG. 12 Values of x^N plotted against x ; $x \leqslant 1.0$

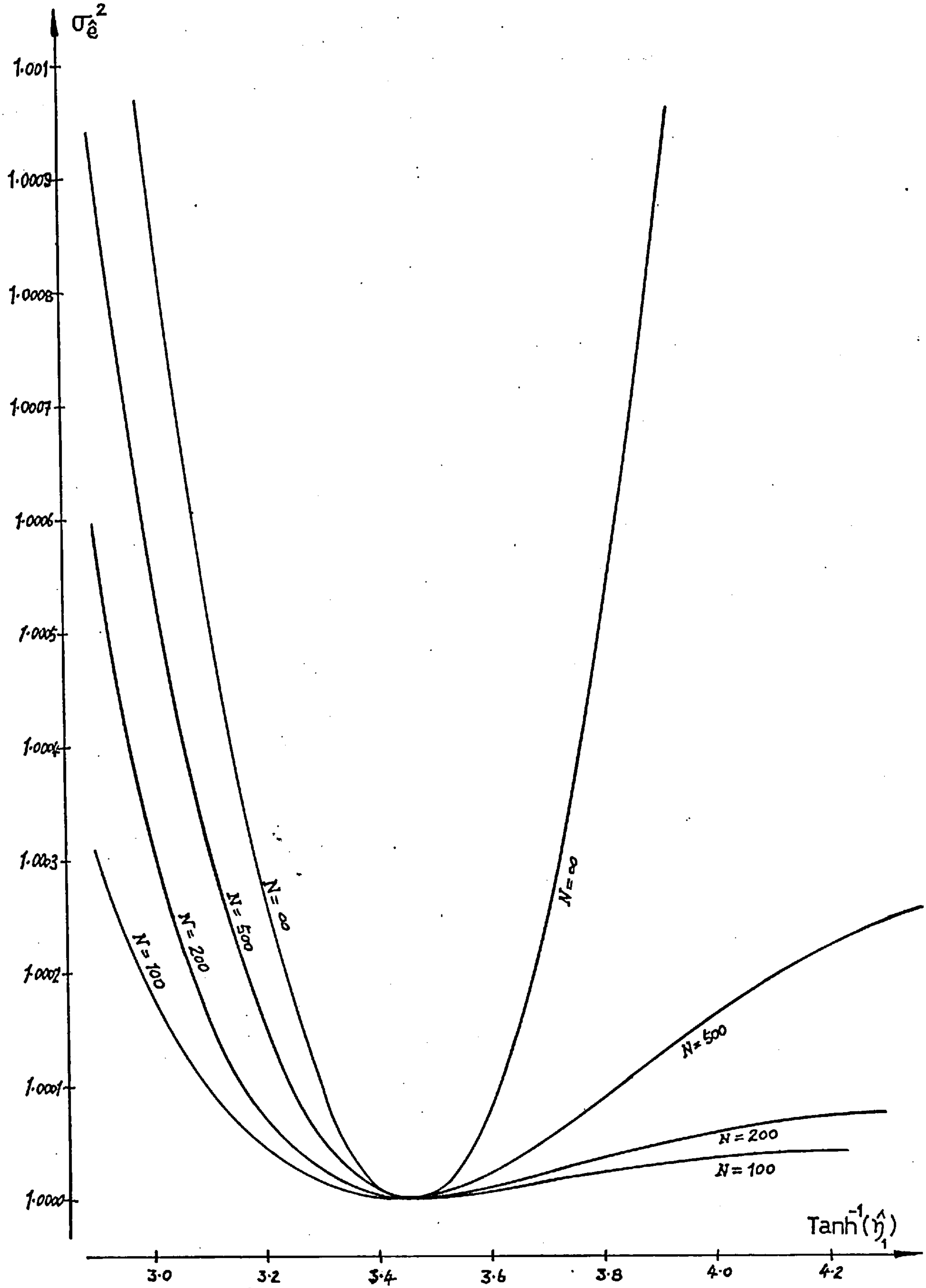


FIG. 13 Values of $\sigma_{\hat{e}}^2$ from (3.64) plotted in X Space

efficiency of hill climbers such as the Newton-Raphson, since a non-positive definite 2nd derivative matrix would make the algorithm step in the wrong direction.

From the above it should be obvious that there could be disadvantag in using the transformation blindly owing to this breakdown effect with finite data sequences. Comparing figures 12 and 13, we might decide that it was unwise to continue to search for an optimum if $/\hat{\eta}_1/>0.998$ and $N<500$ to ensure sufficient convexity in the neighbourhood of the optimum. This suggests a limit for $\hat{\eta}_1^{2N}$ of 0.10 in figure 12 as a criterion which strikes a balance between the data length N, and the nearness of $\hat{\eta}_1$ to the unit circle. To validly employ a stronger pole $\hat{\eta}_1$ to estimate some η_1, a longer data length should first be obtained to satisfy the above criterion. In chapter 4, we shall develop such ideas further and present other criteria with similar effects, but derived in other ways. These criteria will then automatically ensure that the breakdown effect described here, is fully controlled and not critical.

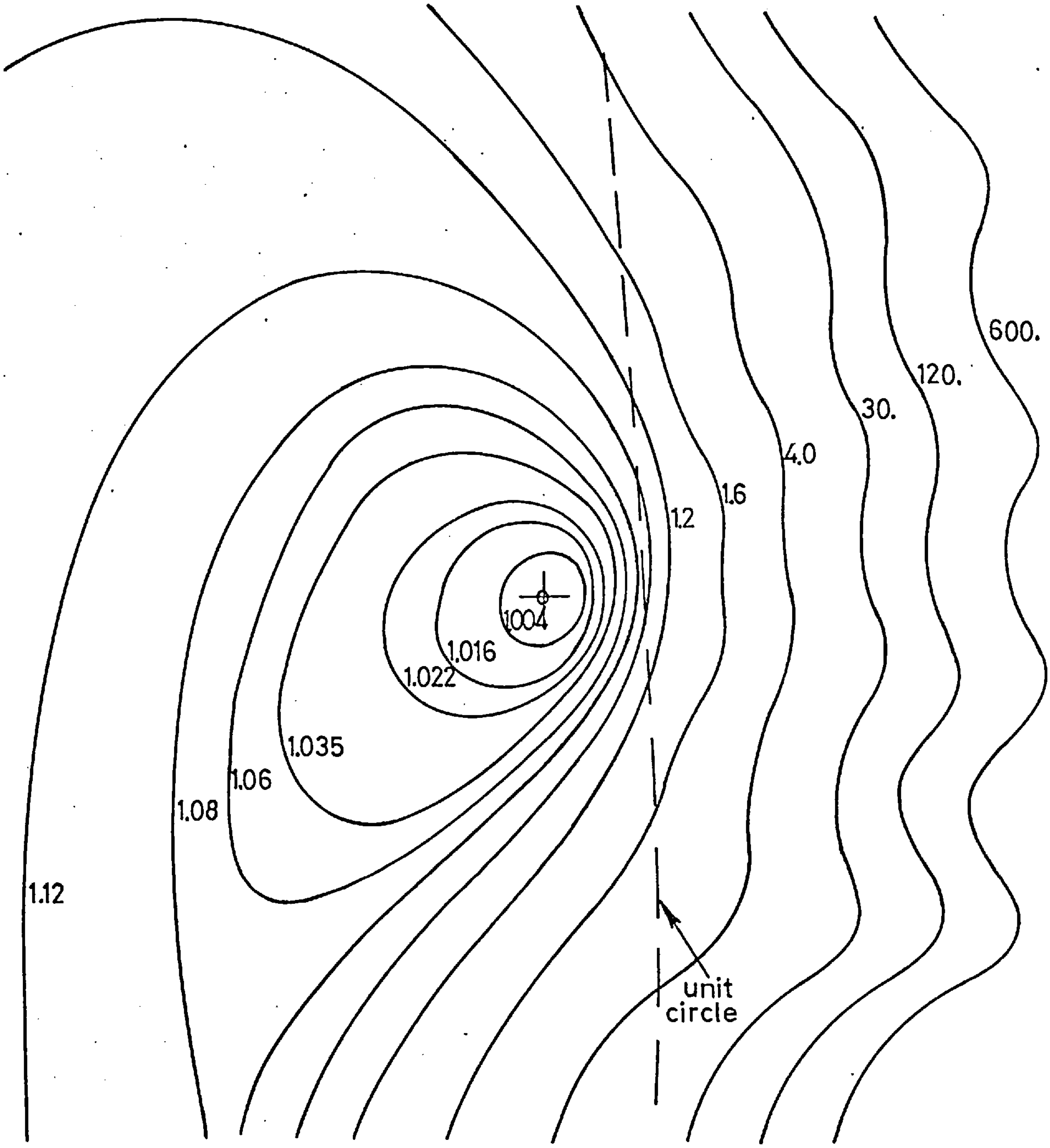
3.13 Finite data isovars.

If equations (3.64) are reworked, it is possible to demonstrate that we can have the poles of a system outside the unit circle for a finite data length, and yet produce a finite cost or sample variance. This is only to be expected, since the data length is not infinite and even an unstable system would have a bounded output sequence. 'Unstable' is defined here as producing an unbounded output within infinite time for an arbitrary, but bounded input sequence. Even if such a system can be validly estimated from a short data length, it would be useless to employ the estimate on a real plant which is effectively working for infinite time.

Despite this remark, we show in figure 14 the isovars for a system with a complex conjugate pair of fixed zeros inside the unit circle at a radius of 0.9777, and a variable complex pair of poles, which can also move outside the unit circle.

The relation (3.41) is used with a data length N of 50 to calculate the sample variance, and as a result the isovars exist outside the unit circle. The cost increases rapidly with radius in this region and there is some cyclic motion of the isovars as the angle of the poles changes in the Z plane. This is due to the aliasing effect of a finite data length, with the exponentially growing sinusoid produced by the pair of complex conjugate poles. The variance calculation can be viewed for this purpose in the manner of section (3.1) as the sum squared response to a unit pulse input at k=1. As the frequency changes with the angle in the plane, so the number of cycles within the data length changes. The final cycle will be the largest and produce a significant change in the sample variance depending on its phase at the end of the data length N. A similar cyclic repetition of maxima has been observed in the stochastic case when filtering a finite noise sequence with an external pole.

We could replot figure 14 in the transformation X space, but only for pole positions inside the unit circle on the Z plane. The isovars would be similar to those of figure 5, although in a different attitude. The most notable difference would be that some of the isovars would be



Pair of zeros at (·9669480 ±j ·14335699)

FIG. 14 Pair of fixed Zeros, finite data set

open ended at infinity in X space, since they cross the unit circle in figure 14. Thus, again we have another view of the breakdown of the hill in transformation X space. It would be possible for the hill climbing routine to fail with the open ended and infinite length contours, i.e. the hill appears to be singular in certain regions.

Once again the cure when estimating is to limit the pole radius in relation to the data length. We might limit the maximum pole radius, but continue to optimise the pole angle in the Z plane. At the final point we would expect to find the local gradient non-zero and aligned along a radial. A final check on the optimum cost could be made exactly on the unit circle at the intersection with the above radial. An example of this procedure is given in Chapter 6.

3.14 Zeros outside the unit circle.

An interesting situation develops when the zeros of a process (3.1) lie outside the unit circle. As demonstrated in figure 1, this is quite a valid condition for a system, and gives a finite variance of the output for all data lengths. Such processes are not uncommon in real plants which show non-minimum phase characteristics. A pole placed exactly over the top of the zero outside the circle would compensate for the latter's presence. Such an idea cannot be countenanced in practise, as exact matching could never be achieved. As a result the system would be unstable, and the method impossible to apply.

Consider the simple system described by (3.65), which is similar to (3.1) with the same definition of the e_k sequence.

$$v_k = \frac{1-\eta_1 z^{-1}}{1-\delta_1 z^{-1}} \cdot e_k \qquad (3.65)$$

The autocorrelations ϕ_r, defined in (3.15) can be repeated for this case from (3.26) and (3.43)

$$\phi_o = \left\{1.0 + \frac{(\delta_1-\eta_1)^2}{1-\delta_1^{\,2}}\right\}\sigma_e^2 \; ; \; \phi_1 = \left\{(\delta_1-\eta_1) + \frac{(\delta_1-\eta_1)^2}{1-\delta_1^{\,2}} \cdot \delta_1\right\}\sigma_e^2 \; ;$$

$$\phi_2 = \left\{(\delta_1-\eta_1)\delta_1 + \frac{(\delta_1-\eta_1)^2}{1-\delta_1^{\,2}} \cdot \delta_1^{\,2}\right\}\sigma_e^2 \quad \text{etc.} \qquad (3.66)$$

For the case $/\eta_1/<1.0$ i.e. Real zero inside the unit circle, all $\phi_r, r\geqslant 1$ can be made zero, by choosing $\delta_1=\eta_1$. This leaves ϕ_o, the central variance term, which is $1.0\,\sigma_e^2$. Thus the spectrum of v_k is white, an independant random signal with the same variance as the e_k sequence. If we closely examine the terms in (3.66) when $/\eta_1/>1.0$ i.e. Real zero outside the unit circle, then all $\phi_r, r\geqslant 1$, can be made zero by choosing $\delta_1=1.0/\eta_1$. The variance term ϕ_o is now however equal to $\eta_1^{\,2}.\sigma_e^2$. This result may also be derived from (3.26) for any number of matching poles and zeros. The system of (3.66) has been compensated by a choice of δ_1 in the sense that it has white output spectrum, but now has a gain of value $/\eta_1/$.

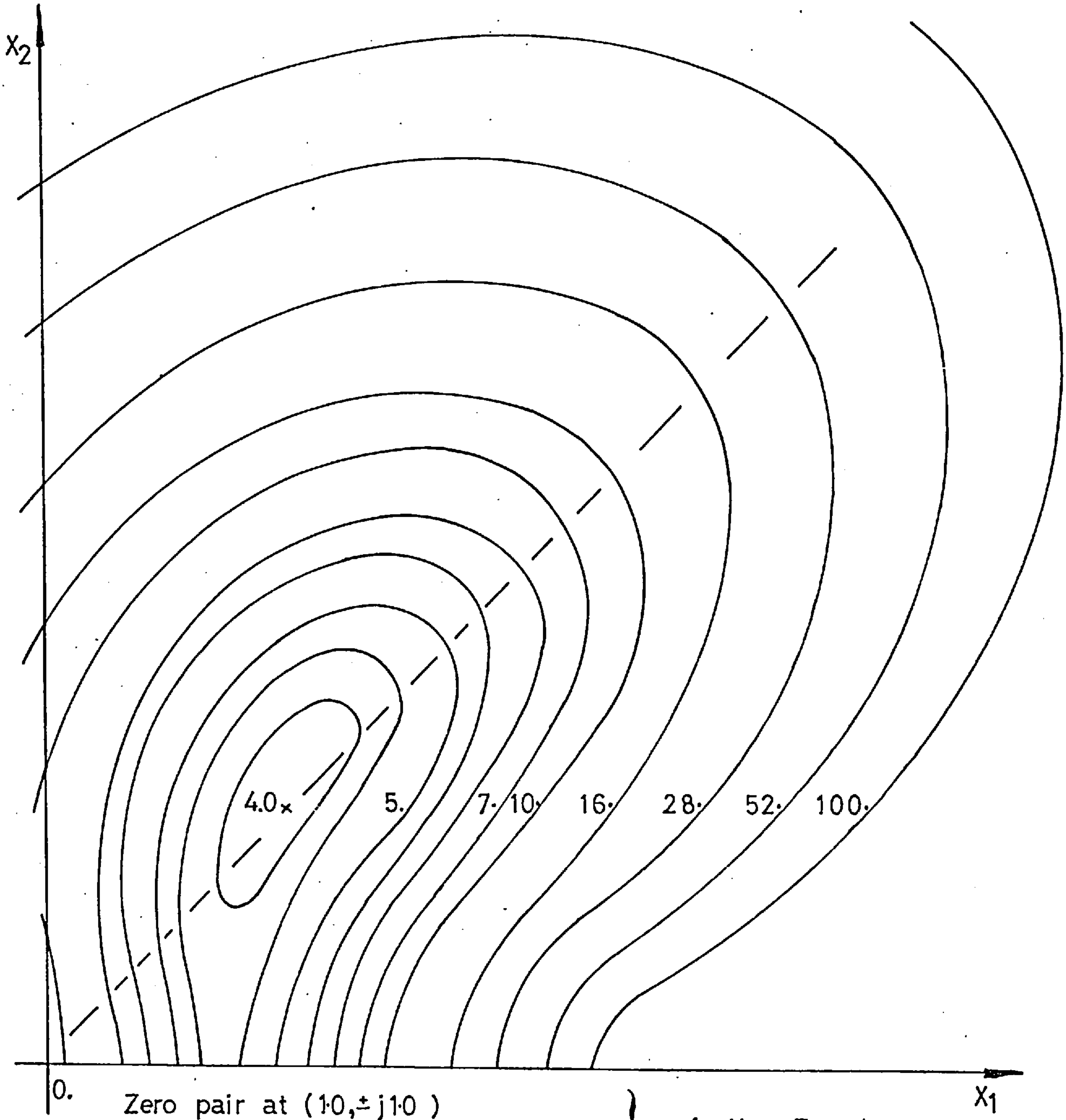
The philosophy of whitening the residual signal v_k as much as possible is in line with Wiener Theory, and it can be shown that the $\delta_1=1.0/\eta_1$ matching does in fact give the minimum mean-square value of v_k, for the condition that δ_1 is constrained to give a stable filter.

An example is shown in figure 15 for a complex conjugate pair of zeros at $(-1.0, \pm j1.0)$. The isovars of the variable pole pair are drawn in X space, and the figure shows an optimum at the internal radial inverse of $(1.0, \pm j1.0)$ in the Z plane i.e. at $(.50, \pm j.50)$ in the Z plane. The optimal cost is $4.0\sigma_e^2$ which corrosponds to the square of the external zero radius, and suggests a square law between the radial position of the system zeros and the minimum cost. This is in fact borne out by careful study of ϕ_o given by (3.26).

The complete system of a set of external zeros matched by a set of inverse internal poles is very similar to an 'all pass' system in continous time described in the Laplace transform s plane.[70] For a pole at s=-a, and a zero at s=+a in the complex s plane, the 'all pass' continuous time system has unit gain at all frequencies, but has a phase shift which changes from zero negatively as the frequency increases. If we were given a continuous time system as in (3.67), we could express this as a z transform as shown below. The term 1/s has been included to make the system physically realisable and gives a term $1/(1-z^{-1})$ in the discrete time description as expected.

$$\text{System Laplace transform } F(s) = \frac{s-b}{s+a} \cdot \frac{1}{s} \;;\; b \geqslant 0,\ a \geqslant 0 \qquad (3.67)$$

Sampled at every T seconds



Zero pair at (1·0, ± j1·0)
Above optimum is at (·50, ± j·50) } in the Z plane

FIG. 15 Pair of external zeros, in X space

$$z \text{ transform } F(z) = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \frac{(p-b)}{(p+a)p} \cdot \frac{1}{1-\exp(-T(s-p))} \cdot dp$$

$$= \sum \text{Residues at poles } p=0, \text{ and } p=-a$$

$$= \frac{p-b}{p+a} \cdot \frac{1}{1-z^{-1}\epsilon^{pT}}\Bigg|_{p=0} + \frac{p-b}{p} \cdot \frac{1}{1-z^{-1}\epsilon^{pT}}\Bigg|_{p=-a}$$ where $z=\exp(Ts)$ and T is the sampling period

$$= \frac{-b}{a} \cdot \frac{1}{1-z^{-1}} + \frac{a+b}{a} \cdot \frac{1}{1-z^{-1}\epsilon^{-aT}}$$

$$F(z) = \frac{1-z^{-1}(1+b/a\ (1-\epsilon^{aT}))}{(1-z^{-1}\epsilon^{-aT})\ (1-z^{-1})} \qquad (3.68)$$

When b is equal to a in (3.67), the all pass continuous time term $\frac{s-a}{s+a}$ can be seen from (3.68) to become (3.69) after sampling.

$$\frac{1-z^{-1}(2-\epsilon^{-aT})}{1-z^{-1}\epsilon^{-aT}} \qquad (3.69)$$

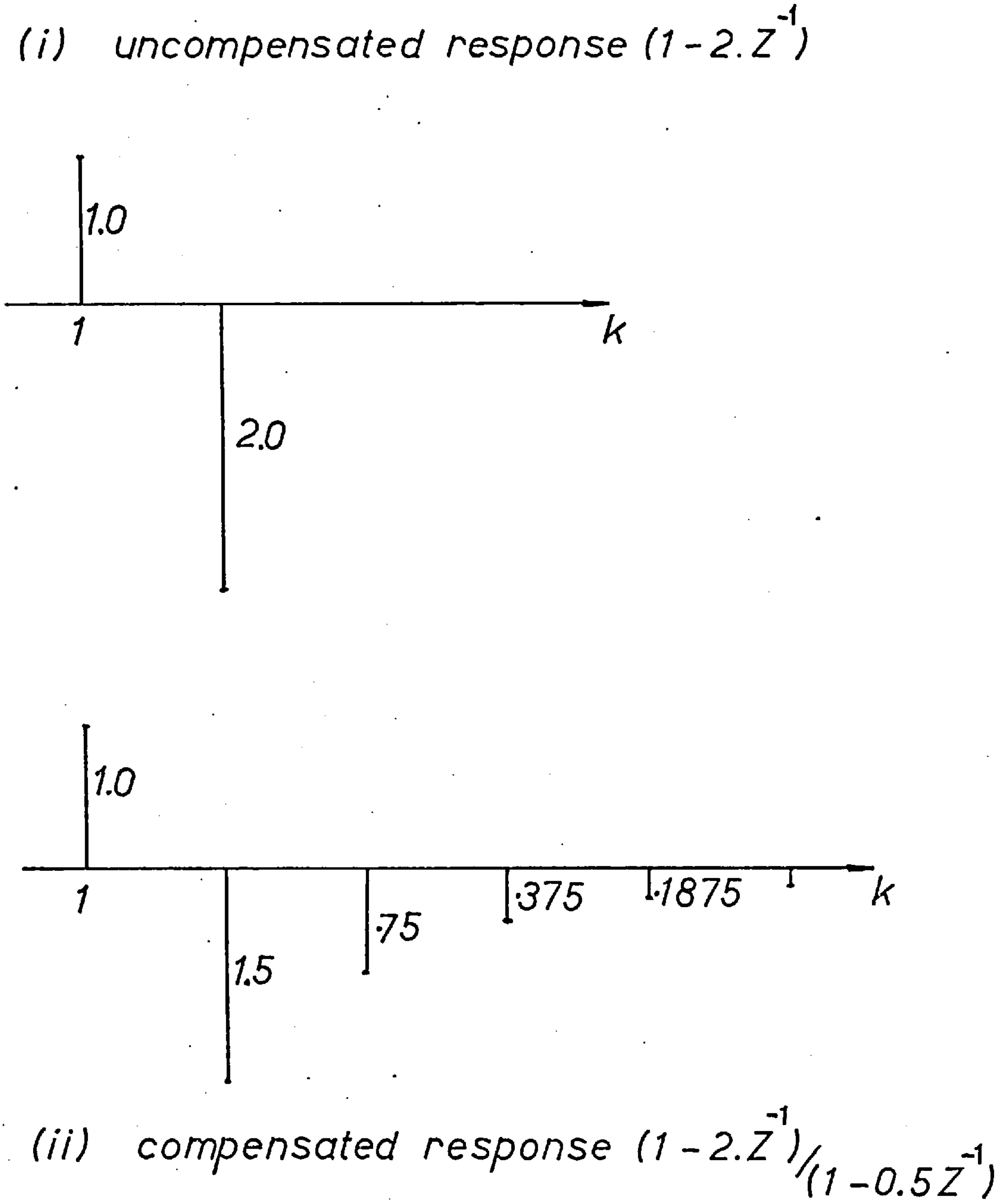
This can be compared to (3.65), but will not have the all pass characteristic $\delta=1.0/\eta$ as obtained before in discrete time. Sampling at discrete times T has destroyed the continuous time property of constant gain at all frequencies. i.e. the all pass property.

As an alternative, let the $\delta=1.0/\eta$ criterion be applied to (3.68). The term $1+b/a(1-\epsilon^{-aT})$ is then equal to ϵ^{+aT}, and this occurs when $b=a\epsilon^{aT}$. Such a system shows the all pass characteristics in discrete time, but obviously does not in continuous time, since we have lost the property a=b. As we might expect, if the sampling time T becomes very small, the two forms of discrete and continous time all pass filter tend to become identical. For the case $\delta=\eta;/\eta/<1.0$ for (3.68), the corrosponding value of b in (3.67) can be formed by equating $1+b/a(1-\epsilon^{-a}$ with ϵ^{-aT}. This gives the solution b/a=-1.0 which implies a zero exactly over the left half s plane pole in the continuous time description.

The discrete time system of (3.65) for the case $\delta=1.0/\eta;/\eta/>1.0$ cannot be identified as being any different from a system in which $\delta=\eta;/\eta/<1.0$ as far as output data on v_k alone is concerned. Both possibilities give a white spectrum output v_k due to their all pass nature. The only difference is that the variance of the output is either η^2 or 1.0 times σ_e^2 respectively. If we cannot measure e_k directly or know its variance, both possibilities are equally valid.

The results above,comparing the continuous and discrete time forms, indicate that if we changed the sampling time T, we should obtain a relative movement between the s plane pole and zero corrosponding to the discrete time form. This is naturally a reflection that what is really required is phase information between the e_k and v_k signals of (3.65). A non-minimum phase system, even when exactly 'matched' will show considerable phase shifts compared to the minimum phase alternative If e_k could be measured or was otherwise known, for example the signal u_k in (1.38), then a proper compensator/controller design could be employed.

Figure 16i shows the impulse response of a simple system similar to (3.65) for $\eta=2.0$ i.e. outside the unit circle, and a zero value for



(i) uncompensated response $(1-2.Z^{-1})$

(ii) compensated response $(1-2.Z^{-1})/(1-0.5Z^{-1})$

FIG.16 Response of simple system

δ. By the arguments of section 3.4, we would expect the output variance to be $(1.0^2+2.0^2)\sigma_e^2=5.0\sigma_e^2$. We cannot in practise apply the ideal compensation of a pole δ at z=2.0, and so we have to resort to a pole δ at z=0.5 which is the inverse condition. The result is the impulse response in figure 16 ii which can be recognised as a typical non-minimum phase impulse response. In fact the expected output variance comes to only $4.0\sigma_e^2$ compared to $5.0\sigma_e^2$ for response i. We can also form ϕ_r, $r=1,2,\ldots\ldots$ and show numerically as in section 3.5 that ϕ_r, $r\geqslant 1$ is zero for such a compensation. The system giving response ii could be used also for a special coding in time of an input signal without having any effect on the signal spectrum.

The results in parts of this section have also been given later by Rowe[16] who employed a spectral view point. The spectral density $\Phi_{vv}(z)$ of the output v_k of a system such as (3.1) or (3.65) is given by (3.70).

$$\Phi_{vv}(z) = \frac{N(z)\ N(z^{-1})}{D(z)\ D(z^{-1})}\ \Phi_{ee}(z) \tag{3.70}$$

where $\Phi_{ee}(z)$ is the power spectral density of the e_k sequence.

If there are some roots of N(z) which lie outside the unit circle, then the corrosponding roots of $N(z^{-1})$ will lie inside the unit circle. Thus we can choose D(z), all of whose poles lie inside the unit circle, to compensate for the roots of N(z) and the roots of $N(z^{-1})$ which lie within the unit circle. As pointed out by Åström[10] and by Doob[48], this choice can always be made for a process with a rational spectral density. With such a compensation scheme $\Phi(z)$ must be equal to $\Phi_{ee}(z)$, i.e. white, within a gain term given by (3.71). This is a slight advancement on Rowe's work, as the terms can be more simply expressed as here in root form.

$$\text{Gain} = \prod (\text{Radius of any zero of } N(z) \text{ outside the unit circle})^2 \qquad (3.71)$$

Such compensation schemes can ~~be~~ readily be applied when estimating non-minimum phase systems with the method of section 2.8 but will not return minimum variances. This must be of academic interest only, since the minimum variance estimator requires an unstable filter.

3.15 <u>Isovars for a Finite data set and external zeros.</u>

Figure 17 shows the isovars for a complex pair of zeros at (0.9935127, ± j 0.15075), radius 1.005 in the Z plane for a data length of 500. As before, the figure is symetric about the real axis, and only part of half of the complete figure is shown. Two minimum cost locations are found for the variable pole pair, and these corrospond to the exact matching point over the zeros and the inverse matching point at (0.9836527, ± j 0.15000). Since the data length is finite, we can have poles outside the unit circle and yet have a finite sample variance. The isovars have a steep slope around the external minimum and a saddle point also outside the unit circle. Clearly this part of the region could cause trouble with hill climb routines such as the Newton-Raphson. We would expect a large number of iterations without

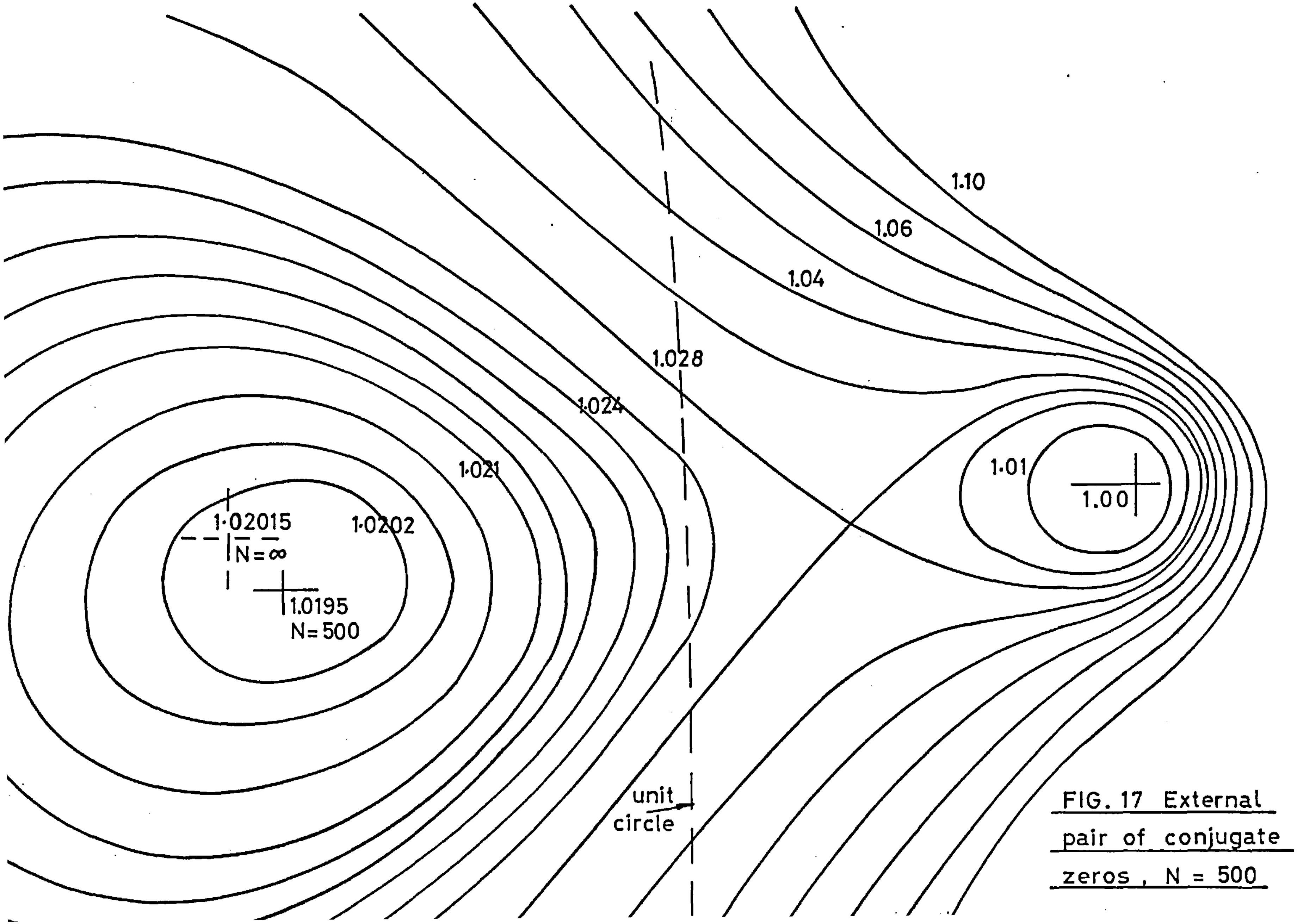


FIG. 17 External pair of conjugate zeros , N = 500

being sure of convergence to the optimum.

As the data length increases, the external contours shrink more tightly about the external zero and the saddle point moves radially outwards towards it. Finally for infinite data, the minimum must be needle sharp at the external zero. The isovars inside the circle would be less affected, as the data length increased, but would have to crowd into the radial space between the internal inverse matching point and the unit circle as in figure 2.

The transformation technique suggested in section 3.10 would allow us to find only the internal minimum at the inverse matching point. This would be valid since we require a filter which is stable both for the estimation process and for later use in the real plant controller. The absolute minimum cost solution would be difficult to estimate and useless in practise.

It will be noticed that in figure 17 the inverse optimum point for a finite data set is not exactly in the same place, nor has the same cost as the optimum point for infinite data. The variance of the output signal for a finite data set is given by (3.41), which includes terms such as $1.-(\delta_i \delta_\nu)^N$. Under the conditions of figure 17, the value of this term, from (3.6) and figure 12, is approximately 0.996 instead of 1.00 for infinite data. This difference is related to the strength of the poles δ_i and the length of the data set N. Some criterion can easily be developed, as for example in section 3.12, so that it is possible to predecide on a data length N for a given pole or vice versa, so that such position differences are kept small. This suggestion will be explored again in chapter 4.

During an actual estimation process the isovars and hills which have been shown here cannot be evaluated from the expected variance relations (3.26) or (3.41) in an analytic manner. They instead can be only evaluated from several runs over the data set of say y_k to give $\hat{\epsilon}_k$ for different values of $\hat{C}(z)$ in the manner of (3.63) in section 3.12. Figure 18 shows a section through such a practical hill for one data set. The system was similar to that of figure 17 but with an external zero pair and with a data set of only 50. The hill section shows a large amount of positive added white noise which is due to the digital round off noise in the computer being amplified by the unstable filter poles. The effect becomes worse as the pole radius increases as would be expected. When the calculation was repeated in double precision (16 decimal digits instead of 8), the digital noise was largely suppressed, at these pole radii.

One would suspect from figure 18 that estimating by a climbing procedure with poles outside the unit circle is very unsatisfactory. A movement of 1 part in 10^8 is sufficient to give a wild variation in the cost and destroy any logical decisions relying on surface smoothness Hill climbers of all types tend to get 'lost' and give very poor convergence. This yet again emphasises that it is only meaningful to climb with the poles constrained within the unit circle.

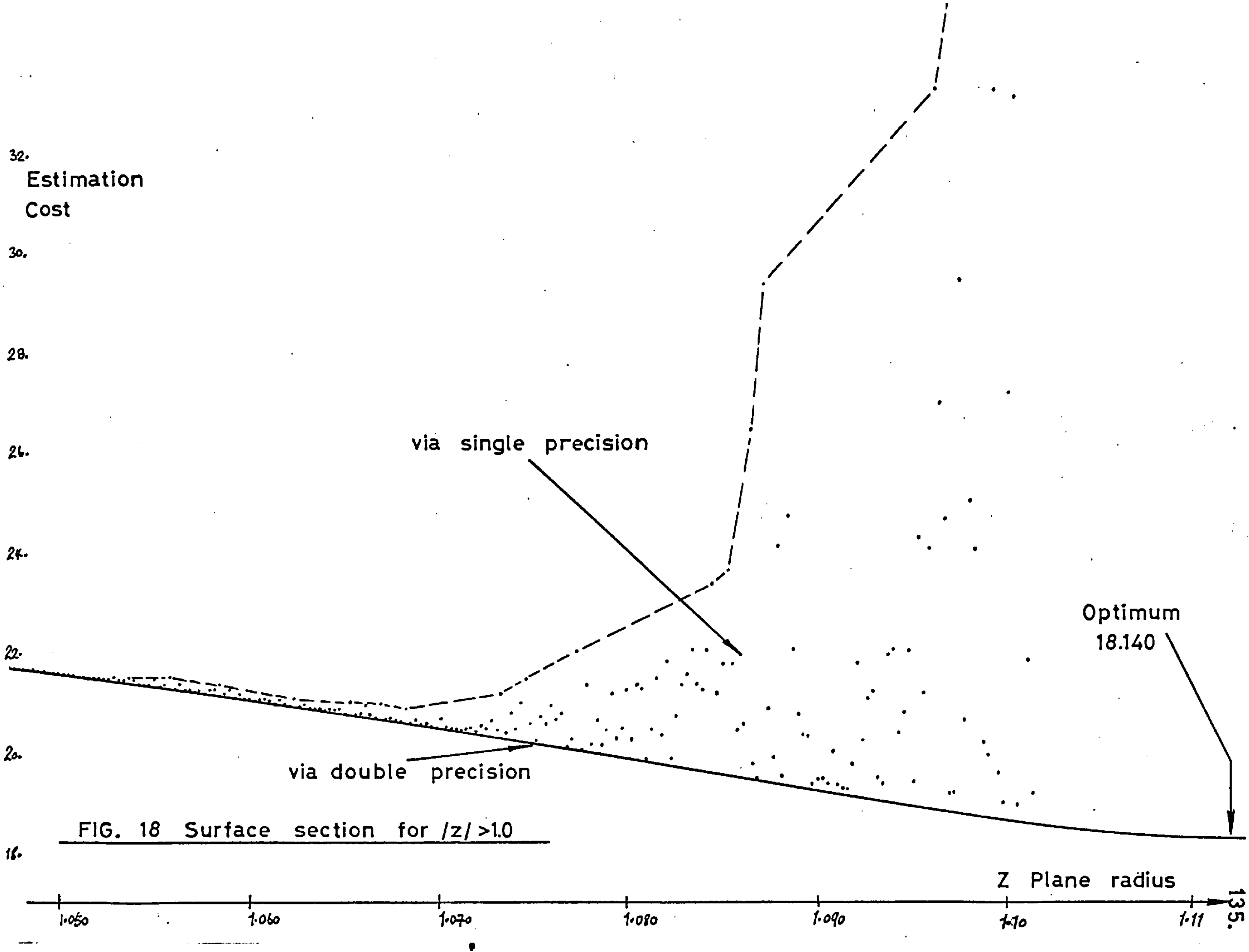


FIG. 18 Surface section for /z/ >1.0

CHAPTER 4

DERIVATIVES AND CRITERIA

4.1 The First Derivative of the estimation cost.

A number of simple closed form relations can easily be found by differentiating the variance expression (3.26) with respect to a pole δ_i or a zero η_i of the system of (3.1). Alternatively the derivatives of the estimation cost $V(\hat{\underline{\theta}}) \triangleq \sigma_{\hat{e}}^2$ can be obtained in open form by differentiating the $\hat{\epsilon}_k$ sequence directly as in (4.1) to (4.5), when applied to the estimation procedure suggested in section 3.12. For exact matching we require the number of poles equal to the number of zeros i.e. m=1, and hence the notation will be dropped. From (3.62) and (3.20)

$$\hat{\epsilon}_k = \frac{\prod_i (z-\hat{\delta}_i)}{\prod_i (z-\hat{\eta}_i)} \; v_k \qquad (4.1)$$

where $\hat{\delta}_i$ and $\hat{\eta}_i$ are estimates of the poles and zeros δ_i, η_i of a system such as (3.1)

$$\frac{\partial \hat{\epsilon}_k}{\partial \hat{\delta}_j} = \frac{-\prod_{i \neq j} (z-\hat{\delta}_i)}{\prod_i (z-\hat{\eta}_i)} v_k = \frac{-1}{(z-\delta_j)} \cdot \frac{\prod_i (z-\hat{\delta}_i)}{\prod_i (z-\hat{\eta}_i)} v_k = \frac{-1}{(z-\hat{\delta}_j)} \cdot \hat{\epsilon}_k \qquad (4.2)$$

$$\frac{\partial \hat{\epsilon}_k}{\partial \hat{\eta}_j} = \frac{+\prod_i (z-\hat{\delta}_i)}{(z-\hat{\eta}_j)^2 \prod_{i \neq j} (z-\hat{\eta}_i)} v_k = \frac{+1}{(z-\hat{\eta}_j)} \cdot \frac{\prod_i (z-\hat{\delta}_i)}{\prod_i (z-\hat{\eta}_i)} v_k = \frac{+1}{(z-\hat{\eta}_j)} \; \hat{\epsilon}_k \qquad (4.3)$$

$$\therefore \frac{\partial\sigma^2_{\hat{e}}}{\partial\hat{\delta}_j} = 2\sum_k\hat{\epsilon}_k \cdot \frac{\partial\hat{\epsilon}_k}{\partial\hat{\delta}_j} = 2\sum_k\hat{\epsilon}_k \cdot \frac{-z^{-1}}{1-\hat{\delta}_j z^{-1}}\hat{\epsilon}_k \qquad (4.4)$$

$$\frac{\partial\sigma^2_{\hat{e}}}{\partial\hat{\eta}_j} = 2\sum_k\hat{\epsilon}_k \cdot \frac{\partial\hat{\epsilon}_k}{\partial\hat{\eta}_j} = 2\sum_k\hat{\epsilon}_k \cdot \frac{+z^{-1}}{1-\hat{\eta}_j z^{-1}}\hat{\epsilon}_k \qquad (4.5)$$

where $\sigma^2_{\hat{e}}$ is defined as the sample variance of $\hat{\epsilon}_k$, taken as $V(\underline{\hat{\delta}})$ above.

At the optimum matching point, $\sigma_{\hat{e}}{}^2$ is a minimum, and both sets of derivatives in (4.5) must be zero. This is only true when the residues $\hat{\epsilon}_k$ are 'white' or independent i.e. $\hat{\epsilon}_k$ is uncorrelated with $\hat{\epsilon}_{k+i}$;/i/>0.

It will be seen from equations (4.4) and (4.5) that the filtering method of obtaining the 1st differential with respect to $\hat{\eta}_i$ or $\hat{\delta}_i$ is almost as simple as Åström's shifting method explained in section 2.8. The new procedure can be programmed very easily as follows as shown in the following example for the derivative with respect to $\hat{\eta}_j$.

```
  i)  Initially set a variable q=0.0, and k=1
 ii)  Sum $\hat{\epsilon}_k * q$ to give ½ differential (4.5)
iii)  Set new q=$\hat{\epsilon}_k + \hat{\eta}_j *$ previous value of q          (4.6)
 iv)  Recycle to ii) with new value of k, until the end of the data
```

The whole process requires two additions and two multiplications for each k step, whereas Åström's shifting method required one addition, one multiplication and one shift for each step. Since in general ~~the~~ the roots are complex, the arithmetic used in (4.6) appears at first sight to be in the complex mode, which is not required by Åström. However many modern computers can perform complex or double precision arithmetic in the same time as single precision.

Equations (4.4) and (4.5) can be further developed for complex conjugate roots. Suppose $\hat{D}(z)$ contains a complex conjugate pair $\hat{\delta},\hat{\delta}^*$ described by (a+jb),(a-jb).

$$\text{Then } \hat{\epsilon}_k = (z-a-jb)(z-a+jb)\,\frac{\hat{D}^{\Delta\Delta}(z)}{\hat{N}(z)}\,v_k \tag{4.7}$$

where $\hat{D}^{\Delta\Delta}$ equals $\hat{D}(z)$ less terms containing the pair of roots $\hat{\delta},\hat{\delta}^*$

$$\frac{\partial\hat{\epsilon}_k}{\partial a} = -(z-a+jb)\,\frac{\hat{D}^{\Delta\Delta}(z)}{\hat{N}(z)}\,v_k - (z-a-jb)\,\frac{\hat{D}^{\Delta\Delta}(z)}{\hat{N}(z)}\,v_k$$

$$\therefore\ \frac{\partial\hat{\epsilon}_k}{\partial a} = \frac{-2(z-a)}{(z-\hat{\delta})(z-\hat{\delta}^*)}\cdot\frac{\hat{D}(z)}{\hat{N}(z)}\,v_k = \frac{-2(z-a)}{(z-\hat{\delta})(z-\hat{\delta}^*)}\,\hat{\epsilon}_k \tag{4.8}$$

$$\frac{\partial\hat{\epsilon}_k}{\partial b} = -j(z-a+jb)\,\frac{\hat{D}^{\Delta\Delta}(z)}{\hat{N}(z)}\,v_k + j(z-a-jb)\,\frac{\hat{D}^{\Delta\Delta}(z)}{\hat{N}(z)}\,v_k$$

$$\frac{\partial\hat{\epsilon}_k}{\partial b} = \frac{+2b}{(z-\hat{\delta})(z-\hat{\delta}^*)}\,\hat{\epsilon}_k \tag{4.9}$$

Similarly if (a+jb) and (a-jb) represent a conjugate pair of roots $\hat{\eta},\hat{\eta}^*$ from $\hat{N}(z)$ we can show

$$\frac{\partial\hat{\epsilon}_k}{\partial a} = \frac{+2(z-a)}{(z-\hat{\eta})(z-\hat{\eta}^*)}\,\hat{\epsilon}_k \ ;\ \frac{\partial\hat{\epsilon}_k}{\partial b} = \frac{-2b}{(z-\hat{\eta})(z-\hat{\eta}^*)}\,\hat{\epsilon}_k \tag{4.10}$$

The derivative with respect to $\sigma_{\hat{\epsilon}}^2$ can be obtained in the same manner as (4.4) and (4.5). The simple filtering required in (4.8), (4.9) and (4.10) can ~~be~~ once more be performed by a method similar to (4.6), and requires only three additions and two multiplications at each step to provide both $\frac{\partial\hat{\epsilon}_k}{\partial a}$ and $\frac{\partial\hat{\epsilon}_k}{\partial b}$ above. Both the single root,and the conjugate pair of roots filter case can therefore be treated without using complex arithmetic.

The equations in (4.10) could be combined to give (4.11). This method does however require the filter, though simpler than before, to be run using complex arithmetic.

$$\frac{\partial\hat{\epsilon}_k}{\partial a} + j\,\frac{\partial\hat{\epsilon}_k}{\partial b} = \frac{2(z-a-jb)}{(z-\eta)(z-\eta^*)}\,\hat{\epsilon}_k = \frac{2}{(z-\eta^*)}\,\hat{\epsilon}_k \qquad (4.11)$$

where $\hat{\eta},\hat{\eta}^* = (a \pm jb)$

The individual gradients of $\sigma_{\hat{\epsilon}}^2$ with respect to a or b can now be taken from the real and imaginary parts of the sum of (4.5), using the simple filter of (4.11). An exactly similar form of (4.11) also holds for the roots $\hat{\delta},\hat{\delta}^*$, but with a negative sign.

4.2 Derivatives in the transform X space.

Whilst the hill climb routine is working in unconstrained X space as described in section 3.10, the cost $V(\underline{\theta})$ must be evaluated using poles and zeros described in the complex Z plane. Using equations (4.4) and (4.5), the gradient of the cost can be found with respect to the pole and zero positions in the Z plane. The hill climb routine requires the cost gradients with respect to points in X space, and these can be found by applying a transformation to the Z plane gradients. Consider a single pair of complex conjugate poles at ($a \pm jb$) in the Z plane. The transformation between the Z and X' description of their position is given by (4.12)

$$a = \mathrm{Rad}_z * \mathrm{Cos}(x_2) \quad ;$$

$$b = \mathrm{Rad}_z * \mathrm{Sin}(x_2) \quad \text{where } \mathrm{Rad}_z = \mathrm{Tanh}(x_1) = z \text{ plane pole radius.} \tag{4.12}$$

i.e. Point (x_1,x_2) in X' describes pole pair ($a \pm jb$) in Z plane.

As before, one dimension x_1 describes, via the Tanh function, the radius of the poles in the Z plane, while x_2 describes the angle in the Z plane, Ang_z, between the vectors (a+jb) and (1.0,0.0).

Given the cost gradients $\partial V/\partial b$ obtained as in section 4.1, we can express these in polar co-ordinates in the Z plane.

$$\frac{\partial V}{\partial \mathrm{Rad}_z} = \frac{\partial V}{\partial a} \cdot \mathrm{Cos}(\mathrm{Ang}_z) + \frac{\partial V}{\partial b} \cdot \mathrm{Sin}\,(\mathrm{Ang}_z) \tag{4.13}$$

$$\frac{\partial V}{\partial \mathrm{Ang}_z} = -\frac{\partial V}{\partial a} \cdot \mathrm{Sin}\,(\mathrm{Ang}_z) + \frac{\partial V}{\partial b} \cdot \mathrm{Cos}\,(\mathrm{Ang}_z) \tag{4.14}$$

Since we have related the angle in X space directly to the Z plane angle Ang_z, we can relate x_2 of X' space to Ang_z as in section 3.10. The component x_1 of X' space and the radius in the Z plane are related by the Tanh function, and therefore the cost differential with respect to x_1 is given by (4.16)

$$\frac{\partial V}{\partial x_1} = \frac{\partial \text{Radz}}{\partial x_1} \cdot \frac{\partial V}{\partial \text{Rad}_z} = \frac{\partial(\text{Tanh } x_1)}{\partial x_1} \cdot \frac{\partial V}{\partial \text{Rad}_z} = (1.0\text{-Rad}_z^2)\cdot \frac{\partial V}{\partial \text{Rad}_z} \qquad (4.16)$$

$$\text{Then } \frac{\partial V}{\partial x_1} = \left(\frac{\partial V}{\partial a}.\text{Cos}(\text{Ang}_z) + \frac{\partial V}{\partial b}.\text{Sin}(\text{Ang}_z)\right)*(1.0\text{-Rad}_z^2) \qquad (4.17)$$

$$\frac{\partial V}{\partial x_2} = -\frac{\partial V}{\partial a} \cdot \text{Sin}(\text{Ang}_z) + \frac{\partial V}{\partial b} \cdot \text{Cos}(\text{Ang}_z) \qquad (4.18)$$

Equations (4.17) and (4.18) will give the required gradients in X' space For a single real pole at z=a, described by x_3 in X' space, the angular terms in (4.17)do not exist, and the gradient in X' space reduces to (4.19)

$$\frac{\partial V}{\partial x_3} = +\frac{\partial V}{\partial a} \cdot (1.0\text{-Rad}_z^2) = \frac{\partial V}{\partial a} \cdot (1.0\text{-}a^2) \qquad (4.19)$$

As described in section 3.10, X' space is increased by one dimension for each extra pole in the Z plane. Clearly the above derivative transformation methods can be extended to any order of poles, including complex pairs, in order to provide a complete set of first derivatives in the X' space. Conceptually this can be extended to second derivatives for a Newton-Raphson procedure, but the suggested Fletcher-Powell hill climbing routine only requires the first derivatives in the climbing space.

A slight ambiguity arises between the X' and Z representations as regards the derivatives in (4.17). A point in the Z plane can be described by more than one possible point in X' space. This is because the angle Ang_z repeatedly sweeps over the Z plane as x_2 passes through increasing multiples of 2π radians. It turns out however that the relation in (4.15) still holds within any segment $0 \rightarrow 2\pi$ of the dimension x_2 and the hill climbing procedure is unaffected. Box describes similar transformations and shows that no extra local minima are produced in the constrained space because of this effect. It is possible however to have negative values of x_1 in X' space, or Rad_x in X space which describe, by convention, positive radii in the Z plane. This means that (4.17) should be modified by multiplying by the sign of x_1 or Rad_x to retain the correct relationship between $\frac{\partial V}{\partial \mathrm{Rad}_z}$ and $\frac{\partial V}{\partial x_1}$.

4.3 Second Derivatives of $\hat{\epsilon}_k$ in the Z plane.

As demonstrated by equation (2.62) the second derivative of the estimation cost $V(\underline{\hat{\theta}})$ can be calculated from the first and second derivatives of $\hat{\epsilon}_k$, with respect to the pole and zero positions. From (4.2), the first derivative with respect to $\hat{\delta}_i$ may be differentiated again as in (4.20), and (4.21).

$$\frac{\partial^2 \hat{\epsilon}_k}{\partial \hat{\delta}_i \partial \hat{\delta}_j} = \frac{\partial}{\partial \hat{\delta}_j}\left[\frac{-1}{(z-\hat{\delta}_i)}\hat{\epsilon}_k\right] = \frac{\partial}{\partial \hat{\delta}_j}\left[\frac{-(z-\hat{\delta}_j)}{(z-\hat{\delta}_i)} \cdot \frac{\hat{D}^{\Delta}(z)}{\hat{N}(z)} \quad v_k\right]$$

$$= + \frac{1}{(z-\hat{\delta}_i)} \frac{\hat{D}^{\Delta}(z)}{N(z)} v_k = \frac{1}{(z-\hat{\delta}_i)(z-\hat{\delta}_j)} \cdot \frac{\hat{D}(z)}{\hat{N}(z)} v_k = \frac{1}{(z-\hat{\delta}_i)(z-\hat{\delta}_j)} \cdot \hat{\epsilon}_k \tag{4.20}$$

where $\hat{D}^{\Delta}(z) = \hat{D}(z)$ less the term $(z-\hat{\delta}_j)$ i.e. $\dfrac{\hat{D}(z)}{(z-\hat{\delta}_j)}$

Similarly

$$\frac{\partial^2 \hat{\epsilon}_k}{\partial \hat{\eta}_i \partial \hat{\delta}_j} = \frac{\partial}{\partial \hat{\delta}_j}\left[\frac{1}{(z-\hat{\eta}_i)} \hat{\epsilon}_k\right] = \frac{\partial}{\partial \hat{\delta}_j}\left[\frac{(z-\hat{\delta}_j)}{(z-\hat{\eta}_i)} \cdot \frac{\hat{D}^{\Delta}(z)}{\hat{N}(z)} \quad v_k\right]$$

$$= \frac{-1}{(z-\hat{\eta}_i)} \cdot \frac{\hat{D}^{\Delta}(z)}{\hat{N}(z)} v_k = \frac{-1}{(z-\hat{\delta}_j)(z-\hat{\eta}_i)} \cdot \frac{\hat{D}(z)}{\hat{N}(z)} v_k = \frac{-1}{(z-\hat{\delta}_j)(z-\hat{\eta}_i)} \hat{\epsilon}_k \tag{4.21}$$

Again the 1st derivative in (4.3) can be differentiated ~~again~~ to give

$$\frac{\partial^2 \hat{\epsilon}_k}{\partial \hat{\eta}_i \partial \hat{\eta}_j} = \frac{\partial}{\partial \hat{\eta}_j}\left[\frac{1}{(z-\hat{\eta}_i)} \hat{\epsilon}_k\right] = \frac{\partial}{\partial \hat{\eta}_j}\left[\frac{1}{(z-\hat{\eta}_i)(z-\hat{\eta}_j)} \cdot \frac{\hat{D}(z)}{\hat{N}^{\Delta}(z)} \quad v_k\right]$$

$$= \frac{1}{(z-\hat{\eta}_i)(z-\hat{\eta}_j)^2} \cdot \frac{\hat{D}(z)}{\hat{N}^{\Delta}(z)} \cdot v_k = \frac{1}{(z-\hat{\eta}_i)(z-\hat{\eta}_j)} \hat{\epsilon}_k \tag{4.22}$$

for $i \neq j$; * factor 2.0 for $i=j$

where $\hat{N}^{\Delta}(z) \triangleq \hat{N}(z)$ less the term $(z-\hat{\eta}_j)$ i.e. $\dfrac{\hat{N}(z)}{(z-\hat{\eta}_j)}$

Clearly we are now in a position to assemble the second derivative of cost as in (2.62) by a simple filtering process similar to section 4.1. If the first derivative sequences (4.2) and (4.3) have already been stored, of if they are generated in parallel, then the second derivatives can be obtained by running the simple filters of (4.6) with the $\partial\hat{\epsilon}_k/\partial\hat{\delta}_i$ or $\partial\hat{\epsilon}_k/\partial\hat{\eta}_i$ signals as inputs. Indeed the same simple filtering program can be used for all these series.

As for the case of section 4.1, the operations appear to require complex arithmetic for other than real roots, but again the equations can be re-worked for conjugate complex pairs, so that only real arithmetic need be considered. Equation (4.8) can be differentiated to give (4.23) and (4.24)

$$\frac{\partial^2\hat{\epsilon}_k}{\partial a\partial a} = \frac{\partial}{\partial a}\left[\frac{-2(z-a)}{(z-\hat{\delta}_i)(z-\hat{\delta}_i^*)} \cdot \frac{\hat{D}(z)}{\hat{N}(z)}\, v_k\right] = \frac{+2}{(z-\hat{\delta}_i)(z-\hat{\delta}_i^*)}\hat{\epsilon}_k \qquad (4.23)$$

since $\dfrac{\hat{D}(z)}{(z-\hat{\delta}_i)(z-\hat{\delta}_i^*)}$ is independant of a.

where $\hat{\delta}_i = (a+jb)$; $\hat{\delta}_i^* = (a-jb)$

$$\frac{\partial^2\hat{\epsilon}_k}{\partial a\partial b} = \frac{\partial}{\partial b}\left[\frac{-2(z-a)}{(z-\hat{\delta}_i)(z-\hat{\delta}_i^*)} \cdot \frac{\hat{D}(z)}{\hat{N}(z)}\, v_k\right] = 0.0 \qquad (4.24)$$

From (4.9)

$$\frac{\partial^2\hat{\epsilon}_k}{\partial b\partial b} = \frac{\partial}{\partial b}\left[\frac{2b}{(z-\hat{\delta}_i)(z-\hat{\delta}_i^*)} \cdot \frac{\hat{D}(z)}{\hat{N}(z)}\, v_k\right] = \frac{+2}{(z-\hat{\delta}_i)(z-\hat{\delta}_i^*)}\hat{\epsilon}_k \qquad (4.25)$$

It will be noticed that in this case, two of the second derivatives (4.23) and (4.25) are the same, and (4.24) is zero, thus allowing simplification.

For the second derivatives of a complex pair $\hat{\eta}_i$ and $\hat{\eta}_i^*$, the derivation is more complicated than the above case. Differentiating (4.10) gives (4.26) and (4.27).

$$\frac{\partial^2 \epsilon_k}{\partial a \partial a} = \frac{\partial}{\partial a}\left[\frac{2(z-a)}{(z-\hat{\eta}_i)^2(z-\hat{\eta}_i^*)^2}\,\frac{\hat{D}(z)}{N^{\Delta\Delta}(z)}\quad v_k\right]$$

$$= \Big\{-2\ (z-\hat{\eta}_i)^2(z-\hat{\eta}_i^*)^2 - 2(z-a)\left[2(z-\hat{\eta}_i).-(z-\hat{\eta}_i^*)^2+2(z-\hat{\eta}_i^*).-(z-\hat{\eta}_i)^2\right]$$

$$* \frac{1}{(z-\hat{\eta}_i)^4(z-\hat{\eta}_i^*)^4} \cdot \frac{\hat{D}(z)}{N^{\Delta\Delta}(z)} \cdot v_k$$

$$= \frac{-2}{(z-\hat{\eta}_i)(z-\hat{\eta}_i^*)}\hat{\epsilon}_k + \frac{8(z-a)^2}{(z-\hat{\eta}_i)^2(z-\hat{\eta}_i^*)^2}\hat{\epsilon}_k \qquad (4.26)$$

where $\hat{\eta}_i = (a+jb)$; $\hat{\eta}_i^* = (a-jb)$; $\hat{N}^{\Delta\Delta}(z) \triangleq \dfrac{N(z)}{(z-\hat{\eta}_i)(z-\hat{\eta}_i^*)}$

The first term in (4.26) is very similar to the result in (4.23)

$$\frac{\partial^2 \hat{\epsilon}_k}{\partial a \partial b} = \frac{\partial}{\partial b}\left[\frac{2(z-a)}{(z-\hat{\eta}_i)^2(z-\hat{\eta}_i^*)^2}\,\frac{D(z)}{N^{\Delta\Delta}(z)}\quad v_k\right]$$

$$= -2(z-a)\left[2(z-\hat{\eta}_i).-j(z-\hat{\eta}_i^*)^2+2(z-\hat{\eta}_i^*).+j(z-\hat{\eta}_i)^2\right]$$

$$* \frac{1}{(z-\hat{\eta}_i)^4(z-\hat{\eta}_i^*)^4} \cdot \frac{D(z)}{N^{\Delta\Delta}(z)}\quad v_k$$

$$= \frac{-8b(z-a)}{(z-\hat{\eta}_i)^2(z-\hat{\eta}_i^*)^2}\hat{\epsilon}_k \qquad (4.27)$$

From the second half of (4.10)

$$\frac{\partial^2 \hat{\epsilon}_k}{\partial b \partial b} = \frac{\partial}{\partial b}\left[\frac{-2b}{(z-\hat{\eta}_i)^2(z-\hat{\eta}_i^*)^2} \cdot \frac{\hat{D}(z)}{N^{\Delta\Delta}(z)} \; v_k\right]$$

$$= \left\{-2(z-\hat{\eta}_i)^2(z-\hat{\eta}_i^*)^2 + 2b\left[2(z-\hat{\eta}_i)(z-\hat{\eta}_i^*)*2b\right]\right\}$$

$$* \frac{1}{(z-\hat{\eta}_i)^4(z-\hat{\eta}_i^*)^4} \cdot \frac{\hat{D}(z)}{N^{\Delta\Delta}(z)} \; v_k$$

$$= \frac{-2}{(z-\eta_i)(z-\eta_i^*)}\hat{\epsilon}_k + \frac{8b^2}{(z-\hat{\eta}_i)^2(z-\hat{\eta}_i^*)^2}\hat{\epsilon}_k \qquad (4.28)$$

As in the single pole case similarities and computational simplifications can be spotted for equations (4.26) to (4.28). Only simple real arithmetic filters are required and little storage need be involved.

4.4 Usage of the second derivatives.

This thesis presents an estimation scheme similar to Åström's maximum likelihood method described in section 2.8. The latter method required the calculation of the 1^{st} derivative vector and the 2^{nd} derivative matrix of the estimation cost $V(\underline{\hat{\theta}})$ with respect to the parameters estimates $\underline{\hat{\theta}}$. Two changes have been made here from that formulation. Chapter 3 has demonstrated the wisdom of estimating polynomial roots instead of coefficients, and the Fletcher-Powell algorithm (2.68) is used which avoids difficulties with a non-positive definite 2^{nd} derivative matrix. Although the 2^{nd} derivative matrix is not directly calculated during our hill climbing procedure, it will be computed at the final optimum to provide a statistical measure of the parameter estimation errors.

4.5 Computation of the 1^{st} derivatives of $V(\hat{\underline{\theta}})$

The estimation method we will be using requires the vector set of 1^{st} derivatives $\partial V/\partial\hat{\underline{\theta}}$. From equation (2.49), the prediction residual error is given by

$$\hat{\epsilon}_k = \frac{\hat{A}(z)}{\hat{C}(z)} y_k - \hat{G}_o \frac{\hat{B}(z)}{\hat{C}(z)} u_k \qquad (4.29)$$

The bias term χ should also be included in (4.29). This term may be modeled as convenient as a bias on y_k,u_k or ϵ_k signals, and will be discussed in Chapter 6. A scalar gain factor G_o has been included in order to redefine B(z) as a normalised polynomial $B(z) \triangleq \prod_{i=1}^{n}(z-\beta_i)$. For convenience we will consider $\hat{\epsilon}_k$ to be formed of two components w_k, v_k as in (4.30)

$$\begin{aligned} \hat{\epsilon}_k &= w_k - v_k \\ &\text{where } w_k \triangleq \frac{\hat{A}(z)}{\hat{C}(z)} y_k \; ; \; v_k \triangleq \hat{G}_o \frac{\hat{B}(z)}{\hat{C}(z)} u_k \end{aligned} \qquad (4.30)$$

The methods of sections 4.1 and 4.3 can be easily applied to derive the 1^{st} derivatives of the cost $V \triangleq \frac{1}{2}\sum_{k=1}^{N} \hat{\epsilon}_k^2$. The detailed workings will not be repeated, as they should clear from the above sections.

From (4.2) $$\frac{\partial\hat{\epsilon}_k}{\partial\alpha_i} = \frac{-1}{(z-\alpha_i)} w_k$$ where α_i is a real root of $\hat{A}(z)$

(4.31)

From (4.8) $\frac{\partial\hat{\epsilon}_k}{\partial a} = \frac{-2(z-a)}{(z-\alpha_i)(z-\alpha_i^*)} w_k$ where α_i, α_i^* are a pair of conjugate roots ($a \pm jb$) of $\hat{A}(z)$ (4.32)

From (4.9) $\frac{\partial\hat{\epsilon}_k}{\partial b} = \frac{+2b}{(z-\alpha_i)(z-\alpha_i^*)} w_k$ (4.33)

From (4.2) $\frac{\partial\hat{\epsilon}_k}{\partial\beta_i} = \frac{+1}{(z-\beta_i)} v_k$ where β_i is a real root of $\hat{B}(z)$ (4.34)

From (4.8) $\frac{\partial\hat{\epsilon}_k}{\partial a'} = \frac{2(z-a')}{(z-\beta_i)(z-\beta_i^*)} v_k$ where β_i, β_i^* (4.35)

are a pair of

From (4.9) $\frac{\partial\hat{\epsilon}_k}{\partial b'} = \frac{-2b'}{(z-\beta_i)(z-\beta_i^*)} v_k$ conjugate roots ($a' \pm jb'$) of $\hat{B}(z)$ (4.36)

From (4.3) $\frac{\partial\hat{\epsilon}_k}{\partial\gamma_i} = \frac{+1}{(z-\gamma_i)} \hat{\epsilon}_k$ where γ_i is real root of $\hat{C}(z)$ (4.37)

From (4.10) $\frac{\partial\hat{\epsilon}_k}{\partial a''} = \frac{+2(z-a'')}{(z-\gamma_i)(z-\gamma_i^*)} \hat{\epsilon}_k$ where γ_i, γ_i^* are a pair of conjugate (4.38)

From (4.10) $\frac{\partial\hat{\epsilon}_k}{\partial b''} = \frac{-2b}{(z-\gamma_i)(z-\gamma_i^*)} \hat{\epsilon}_k$ roots ($a'' \pm jb''$) of $\hat{C}(z)$ (4.39)

From (4.29) $\frac{\partial\hat{\epsilon}_k}{\partial\hat{G}_o} = \frac{-\hat{B}(z)}{\hat{C}(z)} u_k = -v_k/\hat{G}_o$ (4.40)

If the bias term χ were modelled as a bias χ_e on $\hat{\epsilon}_k$ then the derivative $\partial\hat{\epsilon}_k/\partial\chi_e$ would be 1.0 only. As from equation (2.58), the cost gradient can be obtained from $\frac{\partial\hat{\epsilon}_k}{\partial\underline{\theta}}$:

$$\frac{\partial V}{\partial\theta_i} = \sum_{k=1}^{N} \hat{\epsilon}_k * \frac{\partial\hat{\epsilon}_k}{\partial\theta_i} \qquad (4.41)$$

The gradient $\partial V/\partial \gamma_e$ is now equal to the mean of $\hat{\epsilon}_k$ which is easily calculated. It should be evident that equations (4.31) to (4.40) are very similar in structure, and employ the simple filter algorithm of (4.6) with different poles and input sequences. Thus it is quite natural for all these equations to use the same program to evaluate the components $\partial V/\partial \hat{\theta}_i$ of (4.41).

4.6 The estimation algorithm.

The full estimation procedure is similar to that of section 2.8, but uses the Fletcher-Powell[9] hill climbing method to avoid difficulties with non-positive definite second derivative matricies. The iterations progress as in (2.68), which is shown again in (4.42)

$$\underline{\hat{\theta}}_{j+1} = \underline{\hat{\theta}}_j - \alpha^o . H_j \frac{\partial V}{\partial \underline{\hat{\theta}}_j} \quad \text{at the jth iteration} \tag{4.42}$$

The estimate $\underline{\hat{\theta}}_j$ is updated from an initial $\underline{\hat{\theta}}_o$ by a term such that the optimum $\underline{\hat{\theta}}$ is approached. The positive definite matrix H_j, initially set to I, the unit matrix, is updated in turn by information obtained from the change in cost, and the change in gradient achieved during the last step. Finally at the optimum, H is an estimate of the inverse of the second derivative of cost matrix at the optimum. The scalar α^o is determined by minimisation along a line defined by $H_j . \frac{\partial V}{\partial \underline{\hat{\theta}}_j}$; the precise method used is open to selection. In our experience the simpler quadratic minimisation due to Powell[8] is far less ill-conditioned and is more effective in practise than the cubic minimisation due to Davidon and used in the published Fletcher-Powell[9] routine.

In our case the parameter set $\underline{\theta}$ for the model (4.29) describes the roots α_i, β_i, γ_i i=1, ... n of the polynomials $\hat{A}(z)$, $\hat{B}(z)$ and $\hat{C}(z)$, where n is the state vector dimensionality of (1.1). The normalising gain G_o is also included in $\underline{\theta}$. As in section 2.8, the evaluation of the cost requires a run over the data set y_k, u_k, k=1, ... N with the signals $\hat{A}(z)y_k$ and $\hat{G}_o.\hat{B}(z)u_k$ passed through a filter $1/\hat{C}(z)$. To ensure stability of this filter, and of the algorithm, the corrosponding $\hat{\theta}_i$ components must describe the roots γ_i of $\hat{C}(z)$ only through the X' space transformation. A suitable definition for $\hat{\underline{\theta}}$ is thus given by (4.42) to (4.48)

$$\hat{\theta}_i = \alpha_i \quad i=1, \ldots n \qquad \text{if } \alpha_i \text{ is a real root of } \hat{A}(z) \tag{4.42}$$

$$\text{or} \quad \hat{\theta}_i = a; \; \hat{\theta}_{i+1} = b \qquad \text{for a complex conjugate pair of roots } \alpha_i, \alpha^*_{i+1} = (a \pm jb) \text{ of } \hat{A}(z) \tag{4.43}$$

$$\hat{\theta}_i = \beta_j, \quad i=n+1, \ldots 2n; \; j=i-n \qquad \text{if } \beta_j \text{ is a real root of } \hat{B}(z) \tag{4.44}$$

$$\text{or} \quad \hat{\theta}_i = a' \; ; \; \hat{\theta}_{i+1} = b' \qquad \text{for a complex conjugate pair of roots } \beta_j, \beta^*_j = (a' \pm jb') \text{ of } \hat{B}(z) \tag{4.45}$$

$$\hat{\theta}_i = x_j \; ; \; \gamma_j = \text{Tanh}\,(x_j), i=2n+1, \ldots 3n \; ; \; j=i-2n \qquad \text{if } \gamma_i \text{ is a real root of } \hat{C}(z) \tag{4.46}$$

or $\hat{\theta}_i = x_j$; $\hat{\theta}_{i+1} = x_{j+1}$; $R_z = \mathrm{Tanh}(x_j)$;

$$\gamma_j = (R_z \cos x_{j+1}, +R_z \sin x_{j+1}) \;;\; \gamma^*_{j+1} = \mathrm{conjg}(\gamma_j)$$

for a complex conjugate pair of roots γ_j, γ^*_{j+1} of $\hat{C}(z)$, and $x \in X'$

(4.47)

$$\hat{\theta}_{3n+1} = \hat{G}_o \qquad (4.48)$$

The term χ' in (1.38) describing the constant bias level on y_k is not included at this point, and neither are the n initial system conditions which have been assumed to be zero so far. The estimate of λ as described in (2.55) can be left until the optimisation of $\underline{\hat{\theta}}$ has been concluded, and is then given by 2V/N.

The gradient terms $\frac{\partial V}{\partial \hat{\theta}_i}$ for each iteration of (4.42) can be easily calculated using (2.58) and the methods of sections 4.1 and 4.5. For the constrained roots, equations (4.17) to (4.19) have to be used to express the Z plane derivatives in terms of X' space variables. Thus the whole estimation algorithm can be expressed as follows in (4.49

i) Choose some initial value of $\underline{\hat{\theta}}_o$ and set $H_o = I$; $j=0$

ii) Set a temporary vector $\underline{\hat{\theta}} = \underline{\hat{\theta}}_j$

iii) Transform $\underline{\hat{\theta}}$ via equations (4.42) to (4.48) to give the roots $\alpha_i, \beta_i, \gamma_i$ of $\hat{A}(z)$, $\hat{B}(z)$, and $\hat{C}(z)$ and the scalar $\hat{G}_o$.

iv) Run the data set y_k, u_k though the model or filter as given by (4.29), generating the signals w_k and v_k, and evaluating the cost from (2.53). The filter system will be stable since the γ_i roots lie inside the unit circle, owing to the X' space transformation.

v) Run the simple filter described by (4.6) as required to generate $\partial\hat{\epsilon}_k/\partial\alpha_i$, $\partial\hat{\epsilon}_k/\partial\beta_i$, $\partial\hat{\epsilon}_k/\partial\gamma_i$, $\partial\hat{\epsilon}_k/\partial G_o$ as in equations (4.31)to(4.40

vi) Totalise the products in (4.41) to get the cost gradient $\partial V/\partial\theta_i$. For the roots γ_i of $\hat{C}(z)$, we have to first find $\partial V/\partial\gamma_i$ and then employ the gradient transform (4.17) to (4.19) and (4.46),(4.47) to obtain the correct values of $\partial V/\partial\hat{\theta}_i$, i=2n+1, ... 3n.

vii) Knowing the cost from iv) the value of $\hat{\underline{\theta}}$ can be modified by minimising along a direction $-H_j \dfrac{\partial V}{\partial \hat{\underline{\theta}}_j}$ as shown in (4.42) to find the optimum value of the scalar α^o. This involves repeated re-cycling back to stage iii) until $\hat{\alpha}^o$ is determined.

viii) $\hat{\underline{\theta}}_j$ and the matrix H_j can now be updated by the Fletcher and Powell[9] algorithm using $\hat{\alpha}^o$ and the change in gradient so that H tends towards an estimate of the second derivative matrix, but remains positive definite. This completion of an iteration j→j+1 requires recycling back to ii) following a test for convergence which is generally related to the rate of progress and the successive values of $\hat{\alpha}^o$. When the test is satisfied, the estimation procedure is ended

(4.49)

For the initial entry to the procedure stages v) and vi) have to be completed with $\hat{\underline{\theta}} = \hat{\underline{\theta}}_o$. Thereafter stages v) and vi) can be omitted for minimising along a line, if Powell's quadratic minimisation is employed, since this requires the functions $V(\hat{\underline{\theta}})$ only and no gradients. At the finish of each iteration, gradient information is again required to update H_j and $\hat{\underline{\theta}}_j$.

As a consequence of the simplicity of the filtering required in stages v) and vi) many of the operations can be made to run in parallel with stage iv). This avoids the storage of the intermediate signals such as $\partial\hat{\epsilon}_k/\partial\alpha_i$. In the extreme case the only storage required is that for the data set y_k, u_k, although the coding is by then rather complicated.

4.7 Computation of the second derivatives of V.

When the estimation procedure of the previous section has been completed, the matrix of second derivatives of V can be calculated using equations (4.20) to (4.28) and the first derivatives of V at the optimum point. The defining equation (2.62) is repeated here as (4.50)

$$\frac{\partial^2 V}{\partial\delta_i\partial\delta_j} = \sum_{k=1}^{N} \frac{\partial\hat{\epsilon}_k}{\partial\hat{\delta}_j}\cdot\frac{\partial\hat{\epsilon}_k}{\partial\hat{\delta}_i} + \sum_{k=1}^{N} \hat{\epsilon}_k\cdot\frac{\partial^2\hat{\epsilon}_k}{\partial\hat{\delta}_i\partial\hat{\delta}_j} \qquad (4.50)$$

The storage of the first derivatives signals can be avoided by multiplying and summing the signals during the final iteration of the estimation procedure, to give the first term of (4.50).

The second term of (4.50) can be computed by applying the equations of section 4.3 to the model equation (4.29) generating $\hat{\epsilon}_k$. Derivatives $\frac{\partial^2\hat{\epsilon}_k}{\partial\alpha_i\partial\alpha_j}$ and $\frac{\partial^2\hat{\epsilon}_k}{\partial\beta_i\partial\beta_j}$ can be easily obtained by substitutions in (4.20), but with the signals w_k and $-v_k$ replacing $\hat{\epsilon}_k$ respectively. Similarly $\frac{\partial^2\hat{\epsilon}_k}{\partial\gamma_i\partial\gamma_j}$ can be found by using $\hat{\epsilon}_k$ and equation (4.22). The cross product terms $\frac{\partial^2\hat{\epsilon}_k}{\partial\alpha_i\partial\beta_j}$ are zero by differentiation of (4.31) or (4.34).

If we differentiate (4.37) with respect to α_j or β_j, it is clear that $\frac{\partial^2\hat{\epsilon}_k}{\partial\gamma_i\partial\alpha_j}$ and $\frac{\partial^2\hat{\epsilon}_k}{\partial\gamma_i\partial\beta_j}$ again depend on the signals w_k and $-v_k$, with the poles of the simple filters being γ_i, α_j and γ_i, β_j respectively. The other case to be considered are the derivatives associated with $\hat{G}_o$.

From (4.40), $\frac{\partial^2\hat{\epsilon}_k}{\partial\hat{G}_o^{\,2}}$ and $\frac{\partial^2\hat{\epsilon}_k}{\partial\hat{G}_o\partial\alpha_i}$ are both zero, $\frac{\partial^2\hat{\epsilon}_k}{\partial\hat{G}_o\partial\beta_i}$ is given by $+\frac{v_k}{(z-\beta_i)}$ and $\frac{\partial^2\hat{\epsilon}_k}{\partial\hat{G}_o\partial\gamma_i}$ by $\frac{-v_k}{(z-\gamma_i)}$.

All these terms can be found for the case of complex conjugate roots in an equally simple manner, either by re-running the above items using complex arithmetic or by employing equations (4.23) to (4.28) in place of (4.20) to (4.22).

At the final iteration, the residuals $\hat{\epsilon}_k$ of (4.29) should be independant. If this is not true either the optimum $\hat{\underline{\theta}}_j$ has not yet been reached or it must be concluded that the model does not fit in the sense that it is lower order[37] than the plant from which the data y_k, u_k was obtained.

By studying the second derivative eqations for $\partial^2\hat{\epsilon}_k/\partial\hat{\theta}_i\partial\hat{\theta}_j$ in (4.20) to (4.28) it can be seen that the second derivative signals contain no undelayed terms in $\hat{\epsilon}_k$ as there is at least one delay term z^{-1} in each equation. Since the signal $\hat{\epsilon}_k$ is independant, we must expect the second term of (4.50) to have a zero mean and a variance proportional to N for large data sets of length N. This is similar to Åström's case in section 2.8. Again if these second terms in (4.50) were ignored, the matrix would at least be positive semi-definite, and indicate that the Newton-Raphson algorithm (2.67) could be applied successfully.

4.8 Calculation of the expected second derivative matrix.

The theory due to Cramer and Rao can be used to assess the achievable accuracy of an estimation procedure. This will be described in the next section. The information required is the expected matrix of second order partial derivatives of the estimation cost V. We can then place a lower bound on the covariance matrix of the estimation errors. Due to the expectation operation the second term of (4.50) will be taken as zero i.e. assuming $\hat{\epsilon}_k$ to be independant when the estimation has been completed.

Consider the system as in equation (1.38), with the output y_k generated by input signals u_k and ϵ_k. A notation is now introduced which will be useful in writing down the first terms of (4.50)

$\phi_{\tau} \; \frac{eD}{fC} \cdot \frac{gI}{hJ} \cdot \sigma_e^2$ is defined as the sample cross correlation for delay τ between two systems $\frac{(z-e)D(z)}{(z-f)C(z)} \cdot e_k$ and $\frac{(z-g)I(z)}{(z-h)J(z)} \; e_k$

(4.51)

Large polynomials are written as capital subscripts such as A,B, or C, while individual roots such as α, β, or γ, are in lower case letters.

Consider a root α of $\hat{A}(z)$, then from (4.2),(4.29) and (4.50), the first term of the second derivative is given by (4.52)

$$E \frac{\partial^2 V}{\partial\alpha\,\partial\alpha} = E\sum_{k=1}^{N}\left[\frac{-1}{(z-\alpha)} \; \frac{A(z)}{C(z)} \; y_k\right]^2 \text{ where N is the data length.} \qquad (4.52)$$

$$= E \sum_{k=1}^{N} \left[\frac{-1}{(z-\alpha)} \cdot \frac{A(z)}{C(z)} \left[\frac{G_o.B(z)}{A(z)} u_k + \frac{C(z)}{A(z)} \epsilon_k \right] \right]^2$$

$$= E \sum_{k=1}^{N} \left[\frac{-1}{(z-\alpha)} \cdot \frac{G.B(z)}{C(z)} u_k \right]^2 + E \sum_{k=1}^{N} \left[\frac{-1}{(z-\alpha)} \epsilon_k \right]^2$$

$$+ E \sum_{k=1}^{N} 2 \left[\frac{-1}{(z-\alpha)} \cdot \frac{G.B(z)}{C(z)} \cdot u_k * \frac{-1}{(z-\alpha)} \epsilon_k \right] \qquad (4.53)$$

The last term of (4.53) is taken as zero if u_k and ϵ_k are independant of each other. The signal u_k is assumed to arise from a stationary uncorrelated source of variance σ_u^2. Then (4.53) can be written as (4.54) using the notation defined in (4.51)

$$E \frac{\partial^2 V}{\partial\alpha\,\partial\alpha} = N_* G_o^{\,2} \phi_o \frac{B}{\alpha C} \cdot \frac{B}{\alpha C} \cdot \sigma_u^{\,2} + N_* \phi_o \frac{1}{\alpha} \cdot \frac{1}{\alpha} \cdot \sigma_e^{\,2} \qquad (4.54)$$

For a root β of B(z) we similarly obtain (4.55)

$$E \frac{\partial^2 V}{\partial\beta\,\partial\beta} = E \sum_{k=1}^{N} \left[\frac{1}{(z-\beta)} \cdot \frac{G.B(z)}{C(z)} u_k \right]^2$$

$$= N_* G_o^{\,2} \phi_o \frac{B}{\beta C} \cdot \frac{B}{\beta C} \cdot \sigma_u^{\,2} \qquad (4.55)$$

Also the cross products:

$$E \frac{\partial^2 V}{\partial\alpha\,\partial\beta} = E \sum_{k=1}^{N} \left[\frac{-1}{z-\alpha} \cdot \frac{A(z)}{C(z)} y_k * \frac{1}{(z-\beta)} \; \frac{G.B(z)}{C(z)} u_k \right]$$

$$= N* -G_o \phi_o \frac{B}{\alpha C} \cdot \frac{B}{\beta C} \cdot \sigma_u^{\,2} \qquad (4.56)$$

For the C(z) polynomial, consider one root γ:

$$E\,\frac{\partial^2 V}{\partial\gamma\,\partial\gamma} = E\sum_{k=1}^{N}\left[\frac{1}{(z-\gamma)}\,\epsilon_k\right]^2$$

$$= N_*\ \phi_{o\,\frac{1}{\gamma}\cdot\frac{1}{\gamma}}\cdot\sigma_e^{2} \qquad (4.57)$$

Again the cross product terms can be found:

$$E\,\frac{\partial^2 V}{\partial\alpha\,\partial\gamma} = E\sum_{k=1}^{N}\left[\frac{-1}{z-\alpha}\left[\frac{G.B(z)}{C(z)}\,u_k + \epsilon_k\right]\right]*\ \frac{1}{(z-\gamma)}\,\epsilon_k$$

$$= N_*\ -\ \phi_{o\,\frac{1}{\alpha}\cdot\frac{1}{\gamma}}\cdot\sigma_e^{2} \qquad (4.58)$$

$E\,\dfrac{\partial^2 V}{\partial\beta\,\partial\gamma}$ = zero, since u_k and ϵ_k are assumed to be uncorrelated (4.59)

The cross products with G_o:

$$E\,\frac{\partial^2 V}{\partial\alpha\,\partial G_o} = E\sum_{k=1}^{N}\left[\frac{-1}{z-\alpha}\left[\frac{G_oB(z)}{C(z)}\,u_k + \epsilon_k\right]\right]*\quad \frac{-B(z)}{C(z)}\,u_k$$

$$= N_*\ G_o\ \phi_{1\,\frac{B}{C}\cdot\frac{B}{\alpha C}}\cdot\sigma_u^{2} \qquad (4.60)$$

$$E\,\frac{\partial^2 V}{\partial\beta\,\partial G_o} = E\sum_{k=1}^{N}\left[\frac{1}{(z-\beta)}\cdot G_o\frac{B(z)}{C(z)}\,u_k\right]*\ -\ \frac{B(z)}{C(z)}\,u_k$$

$$= N_*\ G_o\ \phi_{1\,\frac{B}{C}\cdot\frac{B}{\beta C}}\cdot\sigma_u^{2} \qquad (4.61)$$

$$E\,\frac{\partial^2 V}{\partial\gamma\partial G_o} \;=\; \text{zero, as } u_k \text{ and } {}_k \text{ are uncorrelated} \qquad (4.62)$$

$$E\,\frac{\partial^2 V}{\partial G_o \partial G_o} = N_* \,\phi_o \;\frac{B}{C}\cdot\frac{B}{C}\;\cdot\sigma_u^{\,2} \qquad (4.63)$$

The above equations (4.52) to (4.63) have been given essentially for real roots only unless the calculations are done in complex arithmetic, and then ϕ will have real and imaginary components. For our purpose it is wiser to choose parameters θ_i, θ_{i+1} as in (4.43) to describe,for example,the components a and b of a complex conjugate pair of roots α, α^* in A(z). This could also be repeated for the B(z) and C(z) polynomials. In the case of $\hat{C}(z)$, this involves a re-definition of θ_i, θ_{i+1} as in (4.47) from measuring in X' space to measuring in the Z plane. Once the estimation process has been completed and $\hat{C}(z)$ is a stable polynomial, we may well choose to describe it in the Z plane for our own convenience.

Consider a complex conjugate pair of roots α, α^* of A(z), then from (4.8),(4.29) and (4.32), the first term of (4.50) is given by

$$E\,\frac{\partial^2 V}{\partial a \partial a} = E\sum_{k=1}^{N}\left[\frac{-2(z-a)}{(z-\alpha)(z-\alpha^*)}\cdot\frac{A(z)}{C(z)}\;y_k\right]^2 \qquad (4.64)$$

$$= E\sum_{k=1}^{N}\left[\frac{-2(z-a)}{(z-\alpha)(z-\alpha^*)}\cdot\frac{A(z)}{C(z)}\left[\frac{G_o B(z)}{A(z)}\cdot u_k + \frac{C(z)}{A(z)}\;{}_k\right]\right]^2$$

where $\alpha, \alpha^* = (a+jb),\ (a-jb)$

$$= E \sum_{k=1}^{N} \left[\frac{-2(z-a)}{(z-\alpha)(z-\alpha^*)} \cdot \frac{G_o B(z)}{C(z)} u_k \right]^2 + \left[\frac{-2(z-a)}{(z-\alpha)(z-\alpha^*)} \epsilon_k \right]^2$$

$$+ E \sum_{k=1}^{N} 2 \left[\frac{-2(z-a)}{(z-\alpha)(z-\alpha^*)} \cdot \frac{G_o B(z)}{C(z)} u_{k^*} \frac{-2(z-a)}{(z-\alpha)(z-\alpha^*)} \epsilon_k \right] \qquad (4.65)$$

The last term of (4.65) is taken as zero for the u_k and ϵ_k signals being independant of each other. The other terms can be written as (4.66) using the notation defined in (4.51)

$$E \frac{\partial^2 V}{\partial a \partial a} = N_* 4 G_o^2 \not{\phi}_o \frac{aB}{\alpha\alpha^* C} \cdot \frac{aB}{\alpha\alpha^* C} \cdot \sigma_u^2 + N_* 4 \not{\phi}_o \frac{a}{\alpha\alpha^*} \cdot \frac{a}{\alpha\alpha^*} \cdot \sigma_e^2 \qquad (4.66)$$

For the imaginary components b of α, α^* we can use (4.9) and (4.50)

$$E \frac{\partial^2 V}{\partial b \partial b} = E \sum_{k=1}^{N} \left[\frac{2b}{(z-\alpha)(z-\alpha^*)} \cdot \frac{A(z)}{C(z)} y_k \right]^2 \qquad (4.67)$$

$$= E \sum_{k=1}^{N} \left[\frac{2b}{(z-\alpha)(z-\alpha^*)} \frac{A(z)}{C(z)} \left[\frac{G_o \cdot B(z)}{A(z)} u_k + \frac{C(z)}{A(z)} \epsilon_k \right] \right]^2$$

where $\alpha, \alpha^* = (a+jb), (a-jb)$

$$\therefore E \frac{\partial^2 V}{\partial b \partial b} = N_* 4 G_o^2 b^2 \not{\phi}_o \frac{B}{\alpha\alpha^* C} \cdot \frac{B}{\alpha\alpha^* C} \cdot \sigma_u^2 + N_* 4 b^2 \not{\phi}_o \frac{1}{\alpha\alpha^*} \cdot \frac{1}{\alpha\alpha^*} \cdot \sigma_e^2$$

$$(4.68)$$

The cross products are similarly obtained:

$$E\,\frac{\partial^2 V}{\partial a \partial b} = E \sum_{k=1}^{N} \left[\frac{-2(z-a)}{(z-\alpha)(z-\alpha^*)} \cdot \frac{A(z)}{C(z)}\, y_k * \frac{2b}{(z-\alpha)(z-\alpha^*)} \cdot \frac{A(z)}{C(z)}\, y_k \right]$$

(4.69)

$$= E \sum_{k=1}^{N} \left[\frac{-2(z-a)}{(z-\alpha)(z-\alpha^*)} \left[\frac{G_o \cdot B(z)}{C(z)}\, u_k + \epsilon_k \right] * \frac{2b}{(z-\alpha)(z-\alpha^*)} \left[\frac{G_o \cdot B(z)}{C(z)} u_k + \epsilon_k \right] \right]$$

$$E\,\frac{\partial^2 V}{\partial a \partial b} = N_* 4b\, \phi_1 \; \frac{b}{\alpha\alpha^*} \cdot \frac{1}{\alpha\alpha^*} \cdot \sigma_e^2 - N_* 4b G_o^2\, \phi_1 \; \frac{bB}{\alpha\alpha^* C} \cdot \frac{B}{\alpha\alpha^* C} \cdot \sigma_u^{\;2}$$

(4.70)

Similar equations to (4.66) to (4.70) can easily be written for root pairs in B(z),C(z), and any cross product terms. These are quite large in number and will not be shown here as they do not shed any more light on the situation.

The various autocorrelations ϕ_τ can be analytically calculated either for finite or infinite data sets using the methods of chapter three. The final expected second derivative matrix and the matrix obtained at the end of the estimation process for the same parameter set can be compared and conclusions drawn in the sense of the next section.

4.9 The Cramér-Rao theorem.

The inequality due to Cramér and Rao[69] can be used to describe a lower bound to the accuracy of estimating a parameter. This result can be shown in different ways[12,21,35,69], but its derivation will not be shown here.

Consider a probability density function $f(\psi,\underline{\theta})$ and an independant set of samples $\psi_1 \ldots\ldots \psi_N$ drawn from the population of density f. Assume f is continuous in and differentiable with respect to $\underline{\theta}$. Based on those samples we wish to find $\hat{\underline{\theta}}$, an unbiassed estimate of $\underline{\theta}$. The covariance matrix Ψ of the estimation errors of $\hat{\underline{\theta}}$ is defined by (4.71).

$$\begin{aligned}\Psi &\triangleq E\ (\hat{\underline{\theta}}-\underline{\theta})(\hat{\underline{\theta}}-\underline{\theta})^t \\ &= \int\ldots\ldots\int \ (\hat{\underline{\theta}}-\underline{\theta})(\hat{\underline{\theta}}-\underline{\theta})^t f.\ d\psi_1,\ \ldots\ldots\ d\psi_N \qquad (4.71)\end{aligned}$$

The Cramér-Rao inequality is now given in (4.72) and (4.73). Kendall[35] shows the two expressions are equivalent if $f(\psi,\underline{\theta})$ can be differentiated twice with respect to $\underline{\theta}$.

$$\Psi \geqslant \left[E\left[\frac{\text{Log}_e f(\psi,\underline{\theta})}{\partial\underline{\theta}}\right] * \left[\frac{\text{Log}_e f(\psi,\underline{\theta})}{\partial\underline{\theta}}\right]^t\right]^{-1} \qquad (4.72)$$

$$\geqslant \left[-E\left[\frac{\partial 2\text{Log}_e f(\psi,\underline{\theta})}{\partial\underline{\theta}\partial\underline{\theta}}\right]\right]^{-1} \qquad (4.73)$$

Equation (4.72) thus describes a lower bound for Ψ, and it would be desirable to have an estimator which achieved this lower bound. It has to be proved in each particular case that the maximum likelihood estimation method used has this property at least asymptotically.

The probability density function that we are concerned with is the likelihood function $L(\underline{\theta})$ defined in (2.40) in chapter 2. For our own convenience we have been considering $L'=\text{Log}_e(L)$ instead of L for maximisation, and this is permissable since L' is monotonic in L. Thus the vector set of (2.42) is the same as the vector in (4.72). For the purpose of equation (4.73) we therefore require a value of $E\left[\frac{\partial^2 \text{Log}_e(L)}{\partial\underline{\theta}\partial\underline{\theta}}\right]$ and this can be obtained by differentiating (2.54) to give (4.74), and hence Fisher's information matrix $I^N(\underline{\theta})$.

$$E\frac{\partial^2 L'}{\partial\underline{\theta}\partial\underline{\theta}} = \frac{-1}{\lambda^2} E\left[\frac{\partial^2 V}{\partial\underline{\theta}\partial\underline{\theta}}\right] \qquad (4.74)$$

where $V(\underline{\theta}) = \frac{1}{2}\sum_{k=1}^{N}\hat{\epsilon}_k^2$ from (2.53)

Now expressions have already been found in section 4.8 for $E\frac{\partial^2 V}{\partial\underline{\theta}\partial\underline{\theta}}$ and now these can be substituted into (4.74) and hence into (4.73) to give a lower bound for Ψ, the covariance matrix of the parameter estimation errors (4.71). This enables us to examine an estimation scheme to see how the lower bound on Ψ and hence the expected accuracy of the estimated were influenced by the structure of the model and different input signals. At this point the matrix term corrosponding to $\hat{\lambda}$ can be inserted. By differentiating (2.55) we get $\frac{\partial^2 L'}{\partial\hat{\lambda}^2} = \frac{3}{\hat{\lambda}^4}\sum_{k=1}^{N}\hat{\epsilon}_k^2 - \frac{N}{\hat{\lambda}^2}$. But $\hat{\lambda}$ is given by $\frac{1}{N}\sum_{k=1}^{N}\hat{\epsilon}_k^2$ and hence $\frac{\partial^2 L'}{\partial\lambda^2}$ reduces to $\frac{2N}{\lambda^2}$.

As Åström has pointed out, if the matrix $E\frac{\partial^2 L'}{\partial\underline{\theta}\partial\underline{\theta}}$ is singular then there are probably too many parameters in the problem and only linear combinations can be estimated. It now remains to investigate whether the estimation process which we have constructed allows the lower bound to be achieved. Chapter 5 attempts to prove that the process we have used has at least asymptotically the minimum variance property, and chapter 6 shows some comparison results to demonstrate this in practise.

For a practical situation we cannot know the true parameter values $\underline{\theta}$ and have only a finite data set. Thus we cannot obtain the expectation results required in (4.73), and have to resort to an estimate of the information matrix of (4.74) by using (4.50). This can then provide us with some confidence region in which we believe the true parameters will lie. Such an application is mentioned again in section 4.13 and demonstrated in Chapter 6.

4.10 Breakdown due to a finite data set.

Åström has noted in reference 37 at least one example where his method has failed to converge owing to a short data length. The optimum roots of $\hat{C}(z)$ in his case were very close to, or lay outside the unit circle. With more data the estimation procedure was successful. Examples of this behaviour have been often seen for pole positions near the unit circle. In particular the second derivative matrix of the estimation cost, derived from the finite data set in the manner of section 4.7, the 'practical' matrix, does not favourably compare for these cases with the 'theoretical' expected matrix for the same data length, derived as in section 4.8. As the poles of $\hat{C}(z)$ approach the unit circle the 'practical' matrix elements appear to contain a large correlated random factor, and thus the matrix tends to be singular. The climbing routines can frequently be seen to lose their convergence properties and produce an estimate which differs significantly from the theorectical value. The actual estimate will still be the one with the maximum likelihood for that data set, but probably indicates an error distribution outside the range suggested by the Cramér-Rao theory using the theoretical 2nd. derivative cost matrix.

The effect is definitely due to a relation between the length of the data set and the position of the poles, and can provide a 'rule of thumb' to guide the whole procedure.

We shall examine an element of the 'theore~~c~~tical' matrix (4.74) for a data length N as described in section 4.8. In particular it will have a mean value affected by the bias demonstrated in (3.41) which will arise since $\hat{\epsilon}_k$ is strictly a non-stationary sequence for non-infinite N.

Consider a typical element of the matrix from section 4.8 for the β terms as given in equation (4.55). We will calculate the variance of the practical summation repeated here as (4.75). Since u_k is a finite data set, k=1, ... N, the signals used in the summation are strictly non-stationary as described in section 3.4. The mean of (4.75) must therefore be given by a form similar to (3.41)

$$\frac{\partial^2 V}{\partial\beta\,\partial\beta} = \sum_{k=1}^{N}\left[\frac{1}{z-\beta}\cdot\frac{G_o B(z)}{C(z)}\quad u_k\right]^2 \qquad (4.75)$$

$$= G_{\sigma}^2 \sum_{k=1}^{N}\left[u_k v'_o + u_{k-1}v'_1 + u_{k-2}v'_2 \cdots\cdots u_1 v'_{k-1}\right]^2$$

where the expansion of (3.52) is implied, with v'_i being the unit impulse response at delay i of the system

$$\frac{1}{z-\beta}\cdot\frac{B(z)}{C(z)}$$

The variance of this chosen matrix element is defined in (4.77) into which we substitute (4.76)

$$\mathrm{var}\left[\frac{\partial^2 V}{\partial\beta\,\partial\beta}\right] \triangleq E\left[\frac{\partial^2 V}{\partial\beta\,\partial\beta} - E\,\frac{\partial^2 V}{\partial\beta\,\partial\beta}\right]^2 \qquad (4.77)$$

$$= G_{\sigma}^2\; E\Big[\sum_{k=1}^{N}\left(u_k v'_o + u_{k-1}v'_1 \cdots\cdots u_1 v'_{k-1}\right)^2$$
$$- E\sum_{k=1}^{N}\left(u_k v'_o + u_{k-1}v'_1 \cdots\cdots u_1 v'_{k-1}\right)^2\Big]^2$$

where v'_i is the unit impulse response of the system of (4.75) at delay i

$$= G_o^2 E\left\{\begin{bmatrix} u_1^2\, v_o'^2 \\ + u_2^2 v_o'^2 + 2\,u_1 u_2 v'_1 v'_o + u_1^2 v_1'^2 \\ + \cdots\cdots \\ + u_N^2\, v_o'^2 \cdots\cdots + u_1^2\, v_{N-1}'^2 \end{bmatrix} - E\begin{bmatrix} \\ \text{do.} \\ \\ \end{bmatrix}\right\}^2$$

$$= G_o^2\, E\left\{\left(\sum_{m=1}^{N} \sum_{i=1}^{m} \sum_{j=1}^{m} u_i u_j v'_{m-i}\; v'_{m-j} \right) - E\left(\qquad \text{do.} \qquad \right)\right\}^2$$

$$= G_o^2\, E\left\{\left(\sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{n} \sum_{l=1}^{n} u_i u_j u_k u_l v'_{m-i} v'_{m-j} v'_{n-k} v'_{n-l} \right)\right.$$

$$-2\left(\sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{n} \sum_{l=1}^{n} u_i u_j * E(u_k u_l) v'_{m-i} v'_{m-j} v'_{n-k} v'_{n-l}\right.$$

$$+\left(\sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{n} \sum_{l=1}^{n} E(u_i u_j) * E(u_k u_l) v'_{m-i} v'_{m-j} v'_{n-k} v'_{n-l} \right)$$

(4.78)

Now u_k has been assumed to be an independant sequence with zero mean and a Normal distribution with a variance σ_u^2. This means that the only terms to survive from (4.78) will be those described by (4.79)

$E(u_i u_j) = \sigma_u^2$ for i = j, and zero otherwise.

$E(u_i u_j u_k u_l) = E(u_i u_j) * E(u_k u_l) = (\sigma_u^2)^2 = \sigma_u^4$

for $i = j$; $k=l$; $i \neq k$, and zero otherwise.

$E(u_i u_j u_k u_l) = 3\,\sigma_u^4$ for $i=j=k=l$, being the fourth moment of the u sequence which was assumed to have a Normal distribution

(4.79)

The First term of (4.78) reduces to

$$\sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{i=1}^{m} \sum_{\substack{k=1 \\ k \neq i}}^{n} E(u_i^2 u_k^2) v'^{2}_{m-i} v'^{2}_{n-k} \qquad \text{for } j=i \;;\; k=l \;;\; i \neq k$$

$$+ \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{j=1}^{\min(m,n)} E(u_j^4)\; v'^{2}_{m-j}\; v'^{2}_{n-j} \qquad \text{for } j=i=k=l$$

$$= \sigma_u^4 \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{i=1}^{m} \sum_{k=1}^{n} v'^{2}_{m-i}\; v'^{2}_{n-k} + 2\sigma_u^4 \sum_{m=1}^{N} \sum_{n=1}^{N} \sum_{j}^{\min(m,n)} v'^{2}_{m-j}\; v'^{2}_{n-j}$$

(4.80)

The second term of (4.78) reduces to

$$-2\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{j=1}^{m}\sum_{k=1}^{n} E\,(u_j^2)*E(u_k^2)v'^{2}_{m-j}v'^{2}_{n-k} \quad \text{for } i=j \;;\; k=1$$

$$= -2\,\sigma_u^4\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{k=1}^{m}\sum_{k=1}^{n} v'^{2}_{m-j}\;v'^{2}_{n-k} \qquad (4.81)$$

The third term similarly reduces to

$$\sigma_u^4\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{j=1}^{m}\sum_{k=1}^{n} v'^{2}_{m-j}\;v'^{2}_{n-k} \qquad (4.82)$$

Most of these terms therefore cancel to leave (4.83)

$$\text{var}\left[\frac{\partial^2 V}{\partial\beta\partial\beta}\right] = G_o^2 \;*\cdot 2\sigma_u^4\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{j=1}^{\min(m,n)} v'^{2}_{m-j}\;v'^{2}_{n-j} \qquad (4.83)$$

A similar result to (4.80) can be obtained for any element of the cost 2nd. differential matrix of section 4.8. Thus we can expect each element to have a mean value given by an equation similar to (4.55) and a variance given by an equation similar to (4.80). The above derivation can be repeated for the variance of ϕ_1^N or ϕ_2^N which occurs in some elements, and also for the covariance $\phi_o^N * \phi_1^N$

4.11 Calculation of the variance of a matrix element.

The relation given in (4.83) can be pursued further to provide a closed form expression. If the summation limits of (4.83) are examined it will be seen that they can be modified to give (4.84)

$$G_o^2 * \left[4\sigma_u^4 \sum_{m=1}^{N} \sum_{n=1}^{m-1} \sum_{j=1}^{n} v_{m-j}^{\prime 2} v_{n-j}^{\prime 2} + 2\sigma_u^4 \sum_{m=1}^{N} \sum_{j=1}^{m} v_{m-j}^{\prime 4} \right] \qquad (4.84)$$

Expression (3.52) can be used to replace the $v_k^{\prime}$ elements with terms $\sum_i R_i \delta_i^{k-1}$ etc. The second term of (4.84) can then be expressed as follows

$$G_o^2 * 2\sigma_u^4 \sum_{m=1}^{N} \sum_{i=1}^{l} \sum_{\nu=1}^{l} \sum_{j=1}^{l} \sum_{k=1}^{l} R_i R_\nu R_j R_k \left[(\delta_i \delta_\nu \delta_j \delta_k)^{m-2} + \ldots\ldots + (\delta_i \delta_\nu \delta_j \delta_k) \right]$$

where R_i, δ_i ; $i=1, \ldots\ldots l$ are the residuals and roots of $\frac{1}{z-\beta} \cdot \frac{B(z)}{C(z)}$

$$= G_o^2 * 2\sigma_u^4 \sum_{m=1}^{N} \frac{\sum^4 R^4}{\delta^4} \cdot \frac{1-(\delta^4)^m}{1-\delta^4}$$

where the obvious notation $\sum^4 R^4$ and δ^4 has been introduced as an economy.

$$= G_o^2 * 2\sigma_u^4 \frac{\sum^4 R^4}{\delta^4} \cdot \frac{1}{(1-\delta^4)} \left[N - \left\{ (\delta^4)^1 + (\delta^4)^2 \ldots\ldots (\delta^4)^N \right\} \right]$$

$$= G_o^2 * 2\sigma_u^4 \frac{\sum^4 R^4 \cdot N}{\delta^4 (1-\delta^4)} \left[1 - \frac{\delta^4}{N} \cdot \frac{1-(\delta^4)^N}{1-\delta^4} \right] \qquad (4.85)$$

We might have expected var$\left[\frac{\partial^2 V}{\cdots} \right]$ to be linear in N as the part shown in (4.85). A bias term can be seen in (4.85) which arises as in (3.41) from the statistically non-stationary filtering done on the u_k sequence. This term is of more importance when N is "short".

The above process can be repeated for the first term of (4.84):

$$G_o^2 {*} 4\sigma_u^4 \sum_{m=1}^{N} \sum_{n=1}^{m-1} \sum_{i=1}^{l} \sum_{y=1}^{l} \sum_{j=1}^{l} \sum_{k=1}^{l} R_i R_y R_j R_k (\delta_i \delta_y)^{m-n} \Big[(\delta_i \delta_y)^{n-2} {*} (\delta_j \delta_k)^{n-2} +$$

$$\ldots\ldots + (\delta_i \delta_y \delta_j \delta_k)^{o} + (\delta_i \delta_y \delta_j \delta_k)^{-1} \Big]$$

$$= G_o^2 {*} 4\sigma_u^4 \sum_{m=1}^{N} \sum_{n=1}^{m-1} \frac{\Sigma^4 R^4 (\delta_i \delta_y)^{m-n}}{\delta^4} \cdot \frac{1-(\delta^4)^n}{1-\delta^4}$$

where $\Sigma^4 R^4$ and δ^4 are defined as before.

$$= G_o^2 {*} 4\sigma_u^4 \sum_{m=1}^{N} \frac{\Sigma^4 R^4}{\delta^4 (1-\delta^4)} \Big[(\delta_i \delta_y)^{m-1} + \ldots\ldots + (\delta_i \delta_y)^1$$

$$- \Big\{ (\delta_i \delta_y)^{m-1} (\delta_i \delta_y)^1 (\delta_j \delta_k)^1 + \ldots\ldots (\delta_i \delta_y)^1 (\delta_i \delta_y)^{m-1} (\delta_j \delta_k)^{m-1} \Big\} \Big]$$

$$= G_o^2 {*} 4\sigma_u^4 \sum_{m=1}^{N} \frac{\Sigma^4 R^4}{\delta^4 (1-\delta^4)} \left[\frac{1-(\delta_i \delta_y)^m}{1-\delta_i \delta_y} - (\delta_i \delta_y)^m \cdot \frac{1-(\delta_j \delta_k)^m}{1-\delta_j \delta_k} \right]$$

$$= G_o^2 {*} 4\, \sigma_u^4 \frac{\Sigma^4 R^4}{\delta^4 (1-\delta^4)} \Big[\frac{1}{1-\delta_i \delta_y} \Big\{ N - (\delta_i \delta_y)^1 - \ldots\ldots - (\delta_i \delta_y)^N \Big\}$$

$$\frac{-1}{1-\delta_j \delta_k} \Big\{ (\delta_i \delta_y)^1 + \ldots\ldots + (\delta_i \delta_y)^N - (\delta^4)^1 - \ldots\ldots - (\delta^4)^N \Big\} \Big]$$

$$= G_o^2 {*} 4\sigma_u^2 \frac{\Sigma^4 R^4}{\delta^4 (1-\delta^4)} \Big[\frac{N}{1-\delta_i \delta_y} - \delta_i \delta_y \cdot \frac{1-(\delta i \delta y)^N}{(1-\delta_i \delta_y)^2} - \delta_i \delta_y \frac{1-(\delta i \delta y)^N}{(1-\delta_j \delta_k)(1-\delta_i \delta_y)}$$

$$+ \delta^4 \cdot \frac{1-(\delta^4)^N}{(1-\delta^4)(1-\delta_j \delta_k)} \Big]$$

$$= G_o^2 * 4\sigma_u^2 \frac{\Sigma^4 R^4 \qquad N}{\delta^4(1-\delta^4)(1-\delta_i\delta_\gamma)} \left[1.0 - \frac{\delta_i\delta_\gamma}{N} \cdot \frac{1-(\delta_i\delta_\gamma)^N}{1-\delta_i\delta_\gamma} - \frac{\delta_i\delta_\gamma}{N} \frac{1-(\delta_i\delta_\gamma)^N}{1-\delta_j\delta_k} + \frac{\delta^4}{N} \cdot \frac{1-(\delta^4)^N}{1-\delta^4} \right] \qquad (4.86)$$

where $\Sigma^4 R^4 = \Sigma_{i=1}^{1} \Sigma_{\gamma=1}^{1} \Sigma_{j=1}^{1} \Sigma_{k=1}^{1} R_i R_\gamma R_j R_k$; $\delta^4 = \delta_i \delta_\gamma \delta_j \delta_k$

It can also be shown that the terms in (4.85) can be considered to be included in (4.86) with small error provided that $(3.-\delta_i\delta_\gamma) \rightarrow 2.0$ and $(1. + \delta_i\delta_\gamma) \rightarrow 2.0$ which is true for 'strong' poles i.e. close to the unit circle.

Equation (4.86) is now the closed form expression for the variance of $\frac{\partial^2 V}{\partial\beta\partial\beta}$, and is a linear function in N. The bias term apparent in (4.86) again arises from the strictly non-stationary filtering required in the estimation problem. The magnitude of the bias term can be seen plotted in figures 19 and 20. If for a given set of poles δ_i the data length N is short, then the bias term is significant in (4.86). As a result the standard deviation of $\underline{\tilde{\theta}}$ will not be a function of $1/\sqrt{N}$, but will be larger.

The above results clearly are valid for any of the elements of the matrix $\frac{\partial^2 V}{\partial\underline{\theta}\partial\underline{\theta}}$, but will be more important when considering roots of $\hat{A}(z)$ or $\hat{C}(z)$ close to the unit circle. The results enable a criterion to be established from figures 19 and 20, which relates the magnitude of a pole to the minimum data length N. Unless such criteria are followed, the data length N may well prove too short for a satisfactory estimate to be obtained. A maximum likelihood estimate can be obtained for short N, but the actual estimate $\underline{\theta}$ may have a large error

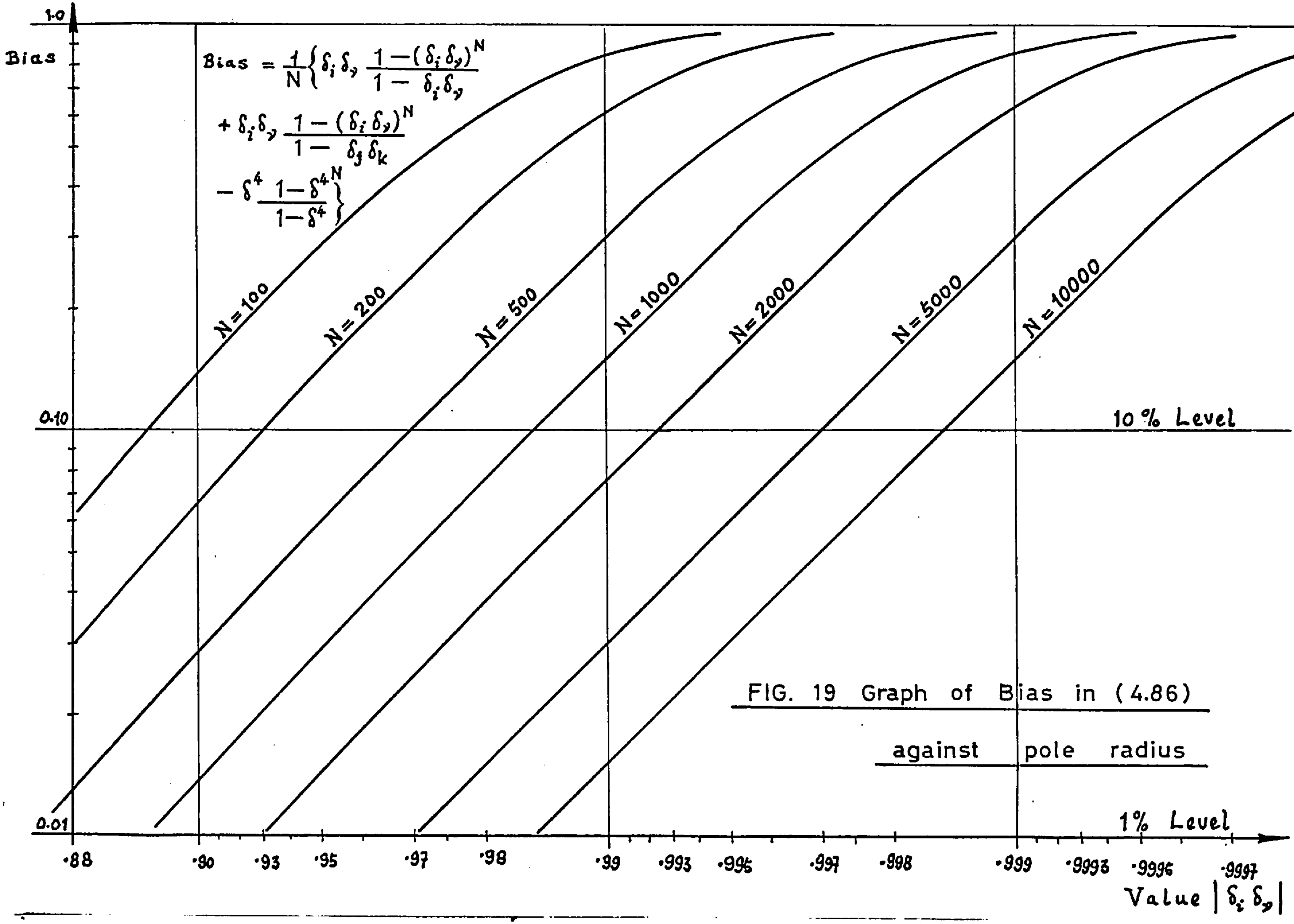


FIG. 19 Graph of Bias in (4.86) against pole radius

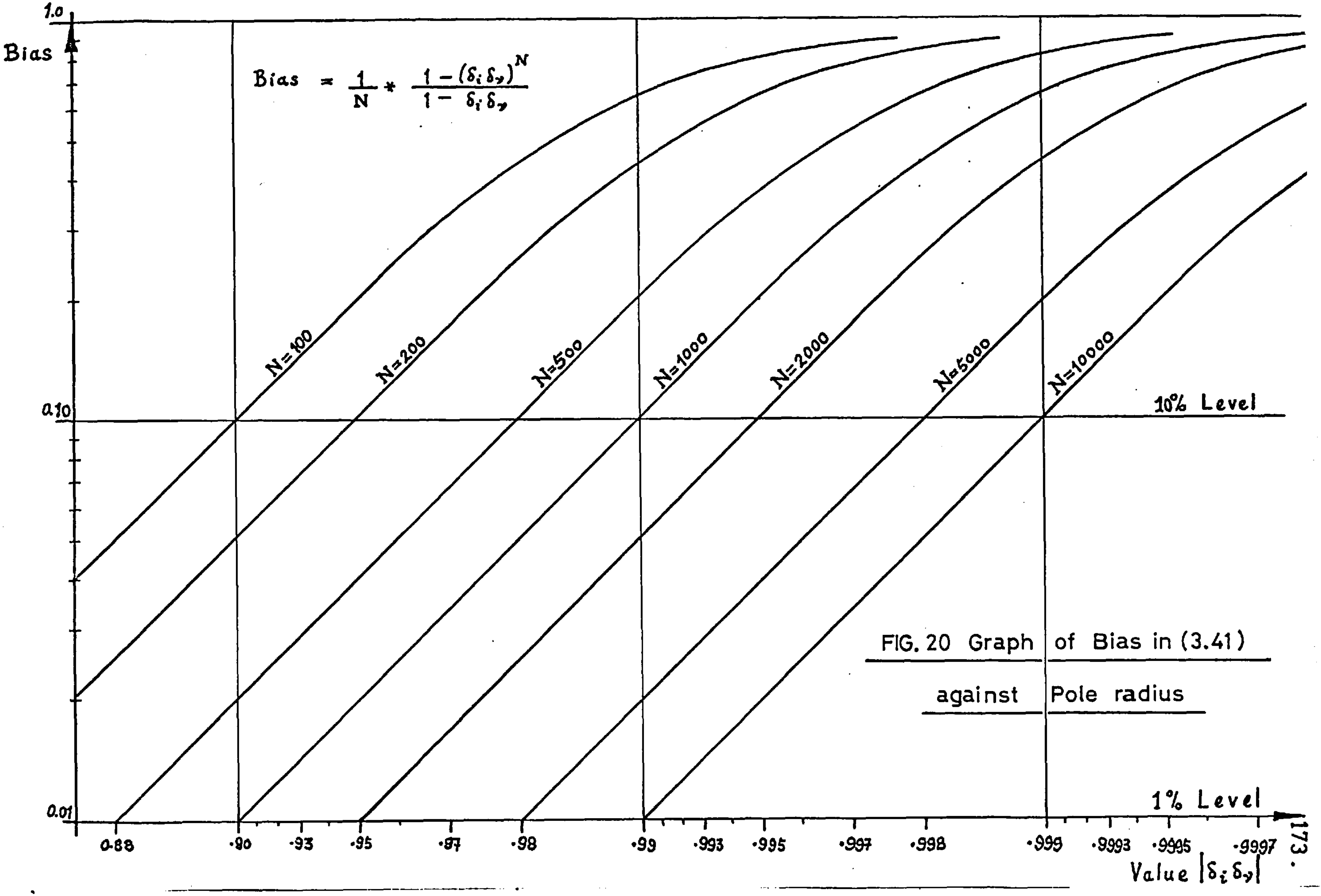


FIG. 20 Graph of Bias in (3.41) against Pole radius

variance and therefore not of practical use.

From figures 19 and 20 we can set a criteria relating N and any $\delta_i \delta_y$; i,y =1, ... l so that the bias terms are less than say 10% of the full value in (3.41) or (4.86). This criteria can be stated as :

$$N \geqslant \frac{10}{1-\delta_i\delta_y} \tag{4.87}$$

4.12 Equivalence of several criteria.

Criteria connecting the length of the data set and the estimated system parameters could be used by plant engineers to assess experiments during plant identification in the field. The criterion suggested in (4.87) is in fact similar to others derived from other considerations. Intuition might suggest that the data length N ought to be significantly longer than the decay time of $1/\hat{A}(z)$ or $1/\hat{C}(z)$. From (3.52) the decay envelope of the impulse response is proportional to δ^{i-1} at delay i from the initial impulse. If we specify that δ^{N-1}/δ be less than some small value ε then

$$\delta \leqslant \frac{1}{\varepsilon^{N-2}} \tag{4.88}$$

This is very similar to requiring δ^{2N} to be small compared with 1.0 for a given value of δ in the expressions shown in figures 19 and 20.

A significant error may arise when using the suggested estimation method due to ignoring the initial conditions on the system at time k=1. This gives a further viewpoint, and the criteria for this is again very similar to the above impulse decay idea. We would require any such initial condition effects to decay appreciable within the data length N compared to the continual disturbances.

The breakdown of the constraining transformation mentioned in section 3.12 yet again suggests that δ^{2N} should be small compared to unity. All these viewpoints can be seen to reduce to a genuine 'rule of thumb' which can be used to make judgements, even during the estimation procedure itself, on the data lengths required for particular problems. The suggested criterionwhich satisfies these demands has been given in (4.87).

We require the 'strongest' roots, i.e. those nearest the unit circle, of $\hat{A}(z)$ or $\hat{C}(z)$ to satisfy the above criteria. The $\hat{B}(z)$ polynomial does not appear from equations (4.54) to (4.63) to be directly involved. This is probably due to the role of B(z) in both the system equation (1.38) and in the estimation filter equation (2.49) being that of providing the zeros of the process. As shown in figures 1 to 11, zeros unlike poles may lie at any point in the z plane without causing any irregularity in the shape of the isovars.

The cross product term of equation (4.58) between a root of $\hat{A}(z)$ and a root of $\hat{C}(z)$ is also of interest. If during estimation $\hat{A}(z)$ was discovered to be very similar to $\hat{C}(z)$ i.e. pole-zero cancellation between α and γ occuring in (2.49), then the matrix element in (4.58) when calculated practically as in section 4.7 would be very similar to

(4.57) and the second term in (4.54). From the preceeding analysis we would expect in practise the matrix to be ill-conditioned if the above criteria were not satisfied.

4.13 A Conflict of Philosophies.

The Bayesian approach as explained in section 2.7 assumes that a parameter θ varies randomly and has a known probability density function $\phi(\theta)$. The problem considered is of estimating the value of θ, from a data set $X_1 \ldots\ldots X_N$ drawn from the parent distribution $f(X,\theta)$. The estimate of θ should be based on the conditional probability density function of θ given the data $X_1 \ldots\ldots X_N$ since this contains all the statistical information. The basic assumption which is made in Bayesian theory is that the probability density function $g(\theta)$ is known in all its detail. It is then possible, provided the algebraic manipulations are not intractable, to obtain an analytic expression for the mean of θ and the variance of θ. Thus before any experimental data X_1 to X_N is collected, we can define fixed limits which would have say a 95% probability of containing the value of the random θ. If in fact the estimation process does not give estimates of θ which lie within these limits in about 95% of a large number of cases, we might suspect the validity of either the estimation procedure or of the initial assumption about $g(\theta)$.

The classical statistical approach assumes θ is fixed and deterministic, but that any estimate $\hat{\theta}$ of θ will be random and have a probability density function. No assumptions need be made about the parent distribution f(X). After the data for an experiment is collected an estimation procedure is used to derive a value of the estimate $\hat{\theta}$ and its variance var.($\hat{\theta}$) about $\hat{\theta}$. The Central Limit theorem might well now be invoked to show that the sample distribution of the estimate could be taken as Normal. Indeed it is well known[21,23] that maximum likelihood estimates asymptotically approach a normal distribution as the data length $N \rightarrow \infty$. The desire now would be to define a fixed interval which with say a 95% probability contain the true value of θ. However since θ is fixed it can only lie either inside, or outside the interval with no intermediate possibility. The statements which should be made is that the confidence interval suggested above is random and covers the true θ with a probability of 95%. The limits can be found from the assumed Normal distribution of $\hat{\theta}$ whose parameters are the estimated mean and variance of $\hat{\theta}$.

The experiment may be repeated and a new data set collected, but the assumed distribution for $\hat{\theta}$ will have a different mean and variance for each experiment and hence the confidence limit will also be different for each experiment. It cannot be necessarily expected that the limits obtained will be similar to the Bayesian limit based on an assumed g(θ). If the estimators are unbiassed and consistent, then for large data sets $N \rightarrow \infty$, we could expect the statistical confidence interval to be small and centred on the true value of θ, the mean of the Bayesian interval.

The matrix elements which we have discussed in the previous sections represent for a vector $\underline{\hat{\theta}}$, the inverse of variance of the estimates of $\underline{\theta}$ and are therefore the parameters of the probability density function of $\hat{\underline{\theta}}$ in the above statistical approach. We have also been calculating the variance in turn of those matrix elements themselves It seems clear that if these variances were kept small compared to the mean value of the elements, we might expect that the classical and Bayesian confidence intervals would more nearly corrospond.

This demonstrates the utility of a criterion in showing the relation between the parameters $\underline{\theta}$ and the data length required to make the estimation philosophies agree in some sense. This enables a practical data length N to be decided for a maximum likelihood estimation method, whose properties in general can only be proven asymptotically as $N \to \infty$.

Given only one data set, it is commonly accepted that the estimation procedures should be allowed to iterate or 'climb' until the gradient of the likelihood cost function such as (2.53) is zero. This gives a maximum likelihood estimate in the classical statistical approach as above. The confidence limits however are unlikely, unless N is large enough, to appear like the Bayesian limits. Further experiments will only confirm an apparent wide spread in the estimates $\hat{\underline{\theta}}$. A wiser procedure would be to run one experiment with a chosen data length, to get some idea of the $\underline{\theta}$ parameters. Later experiments would be run with the data length N chosen via some criterion as above so that the estimates $\hat{\underline{\theta}}$ for different experiments could show some practical agreement.

This approach could be built into a single experiment estimation method. As successive estimates of $\underline{\theta}$ were obtained at each iteration, equation (4.87) could be used to verify that the iterations should be continued, or that the method should be halted. The recomendation to the operator then would indicate that the estimate $\hat{\underline{\theta}}$ was losing validity in the above sense, because the data length was not long enough.

CHAPTER 5

CONSISTENCY AND CONVEXITY.

5.1 Åström's work.

We are concerned in this chapter with defining the conditions which are required for the maximum likelihood estimates to have the desirable properties mentioned in section 2.1.

Åström has done a lot of work in this area as given in references 11 and 37. He has described the systems in terms of the coefficients of Z polynomials in forms similar to (1.38). We have shown in previous chapters the utility of estimating systems in terms of the roots of the Z polynomials.

For a satisfactory estimation procedure we require the probability that an estimate lies close to the true value, to approach unity as the number of available data points N approaches infinity; i.e. a consistent estimate in the sense of section 2.1. This property is often demonstrated by using the law of large numbers, but Åström has used the method of Wald modified for samples which are not independant. Since the theorem proofs of Åström are very fully given in references 11 and 37, we shall not repeat them here in detail. In certain areas the proofs are affected by our system description in terms of roots instead of coefficients, and therefore require more detailed explanation

5.2 Notation.

The vector of parameter estimates is defined as $\underline{\hat{\theta}}$. Åström takes $\underline{\hat{\theta}}$ as (4n+3) in length and defined as in equation (2.50), but including the D.C. level χ', the value of λ and the n initial conditions of the system (1.1) or (1.38). For the work of this thesis $\hat{\underline{\theta}}$ is again (4n+3) long but describes the n roots of the A,B, and C polynomials, the gain G_o, together with χ',λ, and n initial conditions. Changes in definition can be used as before to avoid complex values by describing conjugate pairs in terms of their real and imaginary components as in chapter 4.

The vector Y denotes the N vector of output observations as in equation (2.9). Similarly the u_k sequence is denoted by U. The operator E_o is the mathematical expectation with respect to the distribution of Y when the parameter estimates $\hat{\underline{\theta}}$ have their true values $\underline{\theta}$. The logarithm of the likelihood function of section 2.5 is defined as $L'(Y,\hat{\underline{\theta}})$. This and the estimate $\hat{\underline{\theta}}$ depend on the number of data points N and therefore have the notation $L'^N(Y,\hat{\underline{\theta}})$ and $\hat{\underline{\theta}}^N$.

5.3 Assumptions about the input.

In order for the following proofs to hold, the input signal must be assumed bounded and Ce~~s~~saro summable, that is the limits in (5.1) exist.

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} u_k \;;\; \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} u_k u_{k+i} \qquad \text{for } i=0,1,2\ \ldots.. \tag{5.1}$$

5.4 Lemma 1.

The following Lemmas and theorems are numbered and derived as in references 11 and 37 but are reproduced here using the notation of this thesis. The theorems deal with the asymptotic properties of functions of a single data sample. This data is assumed to have ergodic properties so that results for a single realisation are equivalent to those of an ensemble.

Lemma 1 states that

$$\mathcal{L}im_{N\to\infty}\frac{1}{N}\, L'^{N}(Y,\underline{\hat{\theta}}) = \mathcal{L}im_{N\to\infty}\frac{1}{N}\, E_o L'^{N}(Y,\underline{\hat{\theta}}) = L'(\underline{\hat{\theta}},\underline{\theta}_o) \qquad (5.2)$$

with probability 1.0 provided that the input satisfies the condition in (5.1) and that $\underline{\hat{\theta}}$ and $\underline{\theta}$ both belong to a region $\mathcal{R}$ in an r dimensional Euclidian space, where r is the dimensionality of $\underline{\theta}$, which is strictly (4n+3). This region is defined as in (5.3)

$\mathcal{R}$=all values of $\underline{\hat{\theta}}$ for which the roots of the polynomials $\hat{A}(z)$ and $\hat{C}(z)$ have magnitudes less than unity, and for which λ is greater than zero equation (1.38) ..... (5.3)

This Lemma gives the asymptotic properties of the likelihood function for the problem studied, and implies that a single realisation of the data can be used instead of an ensemble. The proof Åström gives expresses $L'^{N}(Y,\underline{\hat{\theta}})$ as in (5.4)

$$L'^{N}(Y,\underline{\hat{\theta}}) = \text{constant} - N\,\text{Log}_e\lambda\ \frac{-1}{2\lambda^2}\sum_{k=1}^{N}\hat{e}_k^{\,2} \qquad (5.4)$$

$$= -\,N\,\text{Log}_e\lambda\ \frac{-1}{\lambda^2}\,V(\underline{\hat{\theta}}) \text{ from equation (2.52)}$$

where $V(\underline{\hat{\theta}})$ is a cost defined in(2.53)

The convergence of $L'^{N}(Y,\underline{\hat{\theta}})$ is now equivalent to the convergence of $V(\underline{\hat{\theta}})$ since $\lambda \neq 0$. The value of $\hat{e}_k$ can be substituted in terms of the known signals y_k, u_k from (2.49) and then the various terms in $V(\underline{\hat{\theta}})$ can be examined. Kolmogoroff's criterion[71] of the strong law of large numbers can be applied to each term and implies almost certain convergence if the partial sums are bounded. These partial sums are shown[37] to be generated by difference equations, which produce bounded results provided that the roots of the polynomials $\hat{A}(z)$ and $\hat{C}(z)$ have magnitudes less than one. Thus Lemma 1 holds provided that $\underline{\hat{\theta}}$ belongs to region $\mathcal{R}$. This means that both the system and the model are asymptotically stable.. In this thesis we have so chosen the roots of $\hat{C}(z)$ with the aid of the X' transformation of section 3.10 , that the above criterion is satisfied. The roots of $\hat{A}(z)$ should also have been so chosen to ensure the validity of (5.2), however in the practical examples chosen the roots of $\hat{A}(z)$ did not in fact exceed the unit circle. The original test systems were also stable i.e. the roots of $A(z)$ had magnitudes less than one.

Lemma 2

Let the input u_k satisfy assumption (5.1) and let $\mathcal{R}'$ be a closed set contained in $\mathcal{R}$, then $L'(\underline{\hat{\theta}},\underline{\theta}_o)$ is ana~~l~~nalytic function of $\underline{\hat{\theta}}$ within the set $\mathcal{R}'$. Lemma 2 states that

$$\mathcal{L}im_{N\to\infty} \frac{1}{N} \cdot \frac{\partial}{\partial\theta i}\left[L'^{N}(Y,\underline{\hat{\theta}})\right] = \frac{\partial}{\partial\theta i} \cdot \mathcal{L}im_{N\to\infty} \frac{1}{N} \cdot E_o L'^{N}(Y,\underline{\hat{\theta}}) = \frac{\partial}{\partial\theta i} \cdot L'(\underline{\hat{\theta}},\underline{\theta}_o)$$

with probability 1.0 ..... (5.6)

This relation also holds for higher derivatives. These results can be demonstrated[37] because $L'^N(Y,\hat{\underline{\theta}})$ is infinitely differentiable in R' and by analytic continuation we can define an analytic function of a complex variable $\hat{\underline{\theta}}$. The function $L'^N(Y,\hat{\underline{\theta}})$ increases monotonically with N, but not faster than N. Therefore $\frac{1}{N} L'^N(Y,\hat{\underline{\theta}})$ is bounded and converges uniformly for $\hat{\underline{\theta}}$ belonging to R', and is thus an analytic function. These Lemmas establish that the average over the sequence length N of $\hat{e}_k^2$ converges to its ensemble average as $N\rightarrow\infty$, which is differentiable in $\underline{\theta}$.

5.5 Theorem 1

This theorem is concerned with the uniqueness of the maximum of the likelihood function and requires the previous Lemmas to hold true. Let S_o be a set in r dimension Euclidean space defined by (5.3) such that

$$S_o \triangleq \hat{\underline{\theta}} \text{ for which } L'(\hat{\underline{\theta}},\underline{\theta}_o) \doteq L'(\underline{\theta}_o,\underline{\theta}_o) \qquad (5.7)$$

Assuming that the signal u_k satisfies assumption (5.1) and that for all N sufficiently large $\hat{\underline{\theta}}^N \in R'$, where R' is a closed set contained in R, then

$$\left\| \hat{\underline{\theta}}^N - \pi\, \underline{\theta}^N \right\| \rightarrow 0. \qquad (5.8)$$

with probability 1.0, where $\pi\underline{\theta}$ is the projection of $\underline{\theta}$ on both S_o and R' i.e. the nearest point in both S_o and R'

Åstrom[37] proves this theorem by following Kendall's work in reference 35 but uses Lemma 1 in place of the strong law of large numbers since the observations are not independant samples. The proof depends critically on the fact that $\hat{\underline{\theta}}^N$ is chosen so that the likelihood function has an absolute maximum, which cannot be guaranteed in pr~~ø~~actise. The climbing algorithms used will find the local optima of a function but not necessarily the global one unless the function is convex. This difficulty is well known for maximum likelihood estimates and can only be covered in practise by assuming that $\hat{\underline{\theta}}^N$ is the globally optimum value when N is sufficiently large.

Theorem 1 implies that the estimate $\hat{\underline{\theta}}^N$ converges into the set S_o as $N \rightarrow \infty$. If the set S_o contains only one point $\underline{\theta}_o$ then the estimate is strongly consistent. The maximum is then unique if N is large enough for S_o to be a point, even if the function has several equal magnitude maxima. Any model with $\hat{\underline{\theta}}$ belonging to S_o generates realisations with the same statistical properties as the given data set of system output Y. We can no longer tell which of the points in S_o generated the observed output. For our representation in root form, S_o would contain all the permutations of the roots which gave the same coefficient values. This point will be covered again later.

Huzurbazar has discussed[39] consistency at length and shows that logically consistency is first proved and then a statement should run "The consistent solution of the likelihood equation is a maximum of the likelihood function with a probability approaching 1.0 as $N \rightarrow \infty$." He also proves that a consistent solution of the likelihood equation is unique.

By employing Lemma 2 and theorem 1, Åström[37] shows (5.9) to hold.

$$\frac{1}{N}\left\| L'^{N}_{\theta\theta}\ (Y,\underline{\hat{\theta}}^{N}) - L'_{\theta\theta}\ (\underline{\hat{\theta}}^{N},\underline{\theta}_{o})\right\| \rightarrow 0. \qquad \ldots\ldots \quad (5.9)$$

with probability 1.0 as $N \rightarrow \infty$, where $L'_{\theta\theta}$ is the notation for the 2nd derivative matrix of L' with respect to θ.

This means that the quantity $L'^{N}_{\theta\theta}\ (Y,\underline{\hat{\theta}}^{N})$ which is computed in the manner of section 4.7 is an almost sure estimate of the information matrix $I^{N}(\underline{\theta}) = N * L'_{\theta\theta}\ (\underline{\theta},\underline{\theta}_{o})$ for large values of N. This is the matrix used in the Cramer-Rao theory described in section 4.9, and Åström's result shows that it is not necessary to compute it separately. The practical difficulties mentioned in section 4.10 cast doubt on this approach for some cases. This only arises when the value of N is not large enough for (5.9) to be true, and it was for this reason the criteria of section 4.12 were introduced.

5.6 Theorem 2

So far all the results of the previous Lemmas and theorem 1 can be shown to be valid for $\underline{\theta}$ being any complete set of parameters as in section 5.2 provided that the $\hat{A}(z)$ and $\hat{C}(z)$ polynomials have roots which lie within the unit circle. Theorem 2 shows that some parameters cannot be consistently estimated.

Let $\underline{\theta}$ be defined as in equation (2.50) and also include the n initial conditions on the system (1.38) as well as the D.C. component χ' and the value λ. Let $\Lambda^{N}(Y,\underline{\hat{\theta}}^{N})$ be the diagonal matrix of eigenvalues of $\frac{1}{N} L'^{N}_{\theta\theta}(Y,\underline{\hat{\theta}}^N)$, and $T^N(Y,\underline{\hat{\theta}}^N)$ be a corrosponding matrix of orthogonal eigenvectors. Then

$$\mathcal{L}im_{N\to\infty}\left\|\Lambda^N(Y,\underline{\hat{\theta}}^N)\ T^{N^t}(Y,\underline{\hat{\theta}}^N)\ \underline{\hat{\theta}}^N-\Lambda^N(Y,\underline{\hat{\theta}}^N)\ T^{N^t}(Y,\underline{\hat{\theta}}^N)\ \underline{\theta}_o\right\| = 0. \quad \text{with probability 1.0} \qquad (5.10)$$

The proof by Åström requires the previous results of Lemmas 1 and 2 and theorem 1, but will not be repeated here.

The theorem indicates the linear combinations of $\underline{\hat{\theta}}^N$ that are consistent, even if some or all of the components of $\underline{\hat{\theta}}^N$ are inconsistent. As a corrolary Åström shows that if $\frac{1}{N}L'^{N}_{\theta\theta}(Y,\underline{\hat{\theta}}^N)$ converges to a value $L'_{\theta\theta}$ then

$$\frac{1}{N}L'^{N}_{\theta\theta}(Y,\underline{\hat{\theta}}^N)\ \underline{\hat{\theta}}^N \to L'_{\theta\theta}\cdot\underline{\theta}_o \qquad \text{with probability 1.0} \qquad (5.11)$$

This implies that the estimate is strongly consistent if $L'_{\theta\theta}$ is non-singular.

Consider now the set S_o in Theorem 1. One of Åström's results is that a set S'_o in S_o is linear in the 3n + 1 coefficient parameter in equation (2.50) and in χ' and λ. Hence $\underline{\hat{\theta}}^N$ will at least converge into a hyperplane S_o for these parameters i.e. components orthogonal to this hyperplane will be consistently estimated. However S_o will always contain the n dimensional sub-space spanned by the parameters associated with the n initial conditions of (1.38) or (1.1) since the corrosponding sub-matrix of the information matrix $I^N(\underline{\theta})$ reaches a finite lower bound as $N \rightarrow \infty$. Thus these initial condition parameters cannot be consistently estimated. This situation is frequently overcome in practise by choosing the data length N long enough. The initial conditions then decay in magnitude compared with the stochastic disturbances and can be justifiably ignored as not contributing greatly to the total cost. If the set of parameters in S_o' or their transformed equivalents have consistent maximum likelihood estimates i.e. S_o contains only one point, then the system could be described as "completely identifiable."

In this thesis we have tried to show a case for representing the $\hat{A}(z)$,$\hat{B}(z)$ and $\hat{C}(z)$ polynomials in terms of their roots α_i,β_i, and γ_i respectively, where i=1, ... n, in place of the coefficient description that Åström uses in $\underline{\hat{\theta}}$. For example, the a_j coefficients of $\hat{A}(z)$, j=1, ... n which are a subset of the components of $\underline{\hat{\theta}}$, can be described in terms of sums and products of the roots α_i. This relation is defined in (5.12).

$$a_j = {}^n\mathcal{C}_j(\alpha_i) \quad ; \quad j=1, \;..... \; n \qquad (5.12)$$

where ${}^n\mathcal{C}_j$ denotes the sum of combinations of n roots taken as a product of order j

Naturally the order of the roots α_i may be permutated amongst themselves and yet give the same value of the coefficient a_j as shown in (5.12).

The conclusion therefore is that a subset of $\underline{\hat{\theta}}$ can be defined in terms of the roots α_i and these estimates will be consistent within a permutation of the order of the set of α_i. Similar reasoning applies to the sets of β_i and γ_i roots belonging to the $\hat{B}$ and $\hat{C}$ polynomials. Thus $\underline{\hat{\theta}}$ can be defined in terms of the roots and Åström's consistency result will still apply with the permutation proviso added. In practice this is no difficulty and the effect will only show itself for example by the occasional exchange of the imaginary components in complex conjugate root pairs for similar estimation runs.

5.9 Excitation and Identifiability.

Åström[37] defines an input signal u_k to be "persistently exciting" of order m if the limits of (5.1) exist and if the matrix in (5.13) is positive definite.

$$\begin{bmatrix} \phi_o & \phi_1 & \cdots\cdots & \phi_m \\ \phi_1 & \phi_o & \cdots\cdots & \phi_{m-1} \\ & & & \\ \phi_m & \phi_{m-1} & \cdots\cdots & \phi_o \end{bmatrix} \quad \text{where} \quad \phi_i = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} u_k u_{k-i} \tag{5.13}$$

Theorem 3

This theorem states that the system (1.38) is completely identifiable if u_k is persistently exciting of order 2n. Åström's proof involves the detailed definition of the set S'_o of the previous section. For S'_o to be a linear set a quadratic form in u_k must equal zero as $N\to\infty$. This in turn requires the matrix in (5.13) to be positive definite. If this holds then S'_o only contains the set $\underline{\theta}_o$, and therefore the estimate $\underline{\hat{\theta}}$ as in section 5.6 is consistent.

The requirements of S_o' can eventually be reduced to having every state of the system controllable either from the input u_k or from the disturbance e_k. In the initial problem statement in chapter 1, this controllability condition was assumed and would also be true for all the other system transformations. In practise this might be violated since in general the properties of the physical system will not be known before data is collected. As a result the estimate of the information matrix $I^N(\underline{\theta})$ can become singular for large values of N. This is a further effect than that described in section 4.10. The two sources of the same effect can be separated only if the criteria of section 4.12 are applied during estimation to decide on a reasonable length of the data set N for the current estimate.

5.8 Asymptotic Normality.

Theorem 4

Define $\underline{\theta}_\Delta$ to be the (3n+3) vector of parameters in $\underline{\theta}$ when those corrosponding to the initial states have been discarded. Thus $\underline{\theta}_\Delta$ is the same as the vector defined in (2.50) but including χ' and λ terms. If the set S_o contains only the point $\underline{\theta}_{o\Delta}$ then $\hat{\underline{\theta}}_\Delta^N$ is consistent. Theorem 4 then states that the stochastic variable $L'_{\theta\theta}(\underline{\theta}_{o\Delta},\underline{\theta}_{o\Delta})\sqrt{N}(\hat{\underline{\theta}}_\Delta^N-\underline{\theta}_o)$ is asymptotically normal with zero mean and variance $-L'_{\theta\theta}$, with probability 1.0 as $N\rightarrow\infty$. If $L'_{\theta\theta}(\underline{\theta}_{o\Delta},\underline{\theta}_{o\Delta})$ is non-singular as well, then $\hat{\underline{\theta}}_\Delta^N$ is asymptotically normal with a mean of $\underline{\theta}_{o\Delta}$ and variance $\frac{1}{N}L'^{-1}_{\theta\theta}$. Since this converges as $N\rightarrow\infty$, the estimates are also asymptotically efficient and we cannot expect to find an estimator with a greater accuracy for long samples.

The proof of this theorem is similar to the standard methods, but in this case the samples are dependent. Åström therefore invokes the statements of the previous Lemma's and theorems. In particular he uses the result of (5.9) and the fact that $L'^{N}_{\theta}(Y,\hat{\underline{\theta}}^N) = 0.$ at the end of the estimation procedure. The proof also depends of the boundedness of u_k expressed in (5.1) and the fact that stable difference equations are obtained if the poles of $\hat{A}^{-1}(z)$ and $\hat{C}^{-1}(z)$ lie within the unit circle. This latter property is satisfied if the system (1.38) is stable, which is a basic assumption. It is possible for the roots of $C(z)$ in (1.38) to lie outside the unit circle for non minimum phase systems and $C^{-1}(z)$ to be unstable. However section 3.14 demonstrated that such a system can be estimated as if the roots γ_i lay within the unit circle. The estimates would not however be minimum variance as described in section 3.14. Without any phase information the non-minimum phase physical system can only be identified as a minimum phase system.

The asymptotic normality implies that the distribution of $\hat{\underline{\theta}}^N$ is known and confidence regions can be determined. This requires an estimate of the covariance matrix so that approximate significance tests can be made on the results.

In general we have demonstrated that all of Åström's Lemmas and theorems hold for the parameter set of roots chosen for $\underline{\theta}$ in this thesis. It is obvious from the preceding work that it is desirable for $L'_{\theta\theta}$ to be non-singular as was suggested in section 4.10. Theorem 1 would give a stronger statement if the likelihood function were convex and $L'_{\theta\theta}$ were positive definite. A local unimodal hill climbing routine could then be employed in confidence to obtain the single optimum $\hat{\underline{\theta}}^N$.

5.9 The Information Matrix.

The information matrix, so named by Fisher, was mentioned in section 4.9 and is defined here in equation (5.14)

$$I^N(\hat{\underline{\theta}}) = - E_o \, L'^N_{\theta\theta}(Y,\underline{\theta}_o) = E_o L'^N_{\theta}(Y,\underline{\theta}_o). \; L'^N_{\theta}(Y,\underline{\theta}_o)^t \qquad (5.14)$$

This matrix has been implicitly mentioned in theorems 2 to 4. It was shown in theorem 2 that an analysis of its rank revealed which components could be consistently estimated. Theorem 3 implies that if the (3n+3) square sub-matrix corrosponding to parameters describing the polynomials A,B,C, and the scalars G,χ' and λ, is positive definite then those parameters are consistently estimated. Theorem 4 shows the estimates to be asymptotically normal with a covariance matrix obtainable from the information matrix. The asymptotic value of the information matrix at the true parameters $\underline{\theta}_o$ is given as (5.15), and this can be calculated for a given $\underline{\theta}_o$ and N as demonstrated in section 4.8.

$$I^N(\underline{\theta}_o) = -N_* L'_{\theta\theta}(\underline{\theta}_o,\underline{\theta}_o) \qquad (5.15)$$

In practise $\underline{\theta}_o$ is unknown and only the form $L'_{\theta\theta}{}^N(Y,\hat{\underline{\theta}}^N)$ can be evaluated as shown in section 4.7. According to one of Åström's results expressed here as equation (5.9), this estimate converges with probability 1.0 to the information matrix as $N\to\infty$. As explained in section 4.10 it frequently appears that N is not sufficiently large for the problem in hand for this ideal convergence to occur. Typically the 'practical' matrix contains correlated random terms and the matrix tends to be singular. The elements in this case are not good estimates of the corrosponding elements of the expected matrix in (5.15).

Sections 4.10 to 4.12 introduced the idea of a relation between the roots α_i, δ_i of the system and the length N of the data set. Now if a theorem similar to 4 allowed us to take $L'^{N}_{\theta\theta}\,(Y,\underline{\hat{\theta}}^{N})$ as normally distributed for finite N and certain related pole positions, then the matrix elements would be defined statistically by their means and variances. Probabilistic statements could be then made about the condition of the matrix for a given value of N and decisions could then be made about further estimation work. For example a test could be employed at each iteration of the estimation procedure as described in sections 4.10 to 4.12 and a decision taken to continue the iterations or stop for want of a longer data length.

Unfortunately a normal distribution cannot be lightly assumed since the results are not derived from independant samples and are valid only asymptotically as $N\to\infty$. In practise however such a working assumption might well be made and schemes similar to those of chapter 4 developed.

5.10 Positive Definiteness.

It is of interest to examine the second derivative matrix $V_{\theta\theta}$ of the estimation cost V. This matrix describes the surface of the estimation hill being climbed, and also appears in the Cramer-Rao theory of section 4.9 and in Fisher's Information matrix in equation (5.14). If it can be shown that $V_{\theta\theta}$ is positive definite, then the hill surface would be convex and have a single unique maximum. We would therefore expect a simple hill-climbing routine with only a capability of finding a nearby unimodal solution would be sufficient for the estimation problem.

For simplicity consider initially a simplified system similar to (3.62) which has been used before and given here as (5.16).

$$\hat{e}_k = \frac{\hat{D}(z)}{\hat{N}(z)} \; v_k \tag{5.16}$$

This system is simply a junior version of the full system used in sections 2.5 and 4.6 for maximum likelihood estimation.

As shown in section 2.5 the likelihood L and its logarithm L' are maximised by minimising the estimation cost $V \triangleq \sum_{k=1}^{N} \hat{e}_k^2$, with respect to the parameter set $\underline{\hat{\theta}}$. The first and second derivatives of V have been covered already in sections 4.5 and 4.7 for $\underline{\hat{\theta}}$ as the root description of the Z polynomials. The results of sections 4.1 and 4.3 allow us to express the matrix (4.50) of the 2nd derivatives of V as (5.17) here when dealing with the junior system above.

$$\frac{\partial^2 V}{\partial\hat{\delta}_i \partial\hat{\delta}_j} = \sum_{k=1}^{N} \frac{\partial\hat{e}_k}{\partial\hat{\delta}_i} \cdot \frac{\partial\hat{e}_k}{\partial\delta_j} + \sum_{k=1}^{N} \hat{e}_k * \frac{\partial^2\hat{e}_k}{\partial\hat{\delta}_i \partial\hat{\delta}_j}$$

$$= \sum_{k=1}^{N} \frac{-1}{z-\hat{\delta}_i} \hat{e}_k * \frac{-1}{z-\hat{\delta}_j} \hat{e}_k + \sum_{k=1}^{N} \hat{e}_k * \frac{1}{(z-\hat{\delta}_i)(z-\hat{\delta}_j)} \hat{e}_k \qquad (5.17)$$

These expressions hold for differentiating with respect to the roots $\hat{\delta}_i$ of $\hat{D}(z)$. Very similar results apply for the roots $\hat{\eta}_i$ of $\hat{N}(z)$ and also for the cross product terms.

Consider the first term of (5.17) alone. This is equivalent to the sum of a product of two signals from different filters, $f_k = \frac{1}{z-\hat{\delta}_i} \hat{e}_k$ and $h_k = \frac{1}{z-\hat{\delta}_j} \hat{e}_k$ over a finite data set N. It should be noted that no shift terms arise in equation (5.17) unlike the case for the coefficient description covered in section 2.8.

Define the N vectors $\underline{f}$ and $\underline{h}$ to corrospond to the scalar sequences f_k and h_k, k=1, ... N. Then $\underline{f}$ and $\underline{h}$ are algebraically independant if either $\hat{\delta}_i \neq \hat{\delta}_j$ or the filters have different inputs. Define a sum vector $\underline{s}$ as the inner product shown in (5.18)

$$\underline{s} = \begin{bmatrix} \underline{f} \underline{h} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \quad \text{where } \alpha_1, \alpha_2 \text{ are arbitary constants.} \qquad (5.18)$$

The vector $\underline{s}$ cannot be null for any α_1, α_2 not zero. Thus the inner products of (5.19) must be positive only.

$$S = \underline{s}^t \underline{s} = \begin{bmatrix} \alpha_1 \alpha_2 \end{bmatrix} \begin{bmatrix} \underline{f}^t \\ \underline{h}^t \end{bmatrix} \begin{bmatrix} \underline{f} \; \underline{h} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \qquad (5.19)$$

$$= \begin{bmatrix}\alpha_1\alpha_2\end{bmatrix}\begin{bmatrix}\underline{f}^t\underline{f} & \underline{f}^t\underline{h} \\ \underline{h}^t\underline{f} & \underline{h}^t\underline{h}\end{bmatrix}\begin{bmatrix}\alpha_1 \\ \alpha_2\end{bmatrix} \qquad (5.20)$$

This implies that the matrix shown in (5.20) is positive definite, and this is true even for finite N. The expectation of this matrix is in fact the zero delay sample cross correlation matrix between the signals f_k and h_k and must be positive definite. This result holds for any number of outputs derived from filters as shown. The restriction is only that the filters must be non-identical if the corrosponding inputs are the same, otherwise a positive semi-definite matrix results.

These arguments can be extended to show that the first term of the 2nd derivative matrix with respect to $\hat{\underline{\Theta}}$ for the complete system in equation (4.50) is positive definite. This arises from the very similar form of the derivatives of $\hat{e}_k$ which have been given in detail in section 4.5 for the complete system.

Non-distinct or multiple roots will give a submatrix which is singular in those roots. This is natural since we cannot then distinguish one root from the other. A singular matrix also arises if there are common root terms between the $\hat{A}$, $\hat{B}$, and $\hat{C}$ polynomials. This is equivalent to a zero cancelling a pole in (2.49) and violates the controllability requirement of section 5.7. A similar non-positive definite situation might be thought to arise for a submatrix corrosponding to a complex conjugate pair of poles. As shown in figures 1 and 2 twin optima appear on the Z plane and this strictly implies an inconsistent estimate. However, by expressing such root pairs in their (a+jb) form, it can be seen from equations (4.8 - 10) and (4.23 - 28) that the hill is unimodal in these parameters if b is limited to only positive values. The above first term matrix is again positive definite by the same reasoning as before.

It is implied by equations (5.18) to (5.20) that the positive definite condition requires the signals f_k and h_k to be finite for all values of N and hence that the sum S is finitely bounded. This requires the filters to be stable and that their roots lie within the unit circle. Again this is a repeat of Åström's Lemma 1 in that the estimated polynomials shall have roots of less than unit magnitude. The condition is covered for the methods of this thesis by using the X' transformation of section 3.10.

Although these results show the first matrix term in (5.17) above to be positive definite when using the estimation procedure, no information is revealed about the condition of the matrix. If it were nearly singular, the climbing routines would have difficulty in finding the optimum. This has been covered before in section 4.10, where further tests were suggested to discover or avoid such a situation.

5.11 The second matrix term.

The arguments of section 5.10 can not be applied to show the positive definiteness of the second term of the second derivative matrix in equation (5.17). This second matrix is again a matrix of cross-correlation products between two signals, in this case $\hat{e}_k$ and $\frac{\partial^2 \hat{e}_k}{\partial \hat{\delta}_i \partial \hat{\delta}_j}$. However there are no convenient auto-correlation elements in the matrix such as there were in (5.20). As a result positive definiteness cannot be proven. If we consider $\hat{\eta}_i$ and $\hat{\delta}_i$ to be equal to the true values η_i and δ_i, for all i, then $\hat{e}_k$ will be a 'white' or independant stochastic signal. Thus any correlation products such as the second half of (5.17) will have an expectation of zero at the exact matching condition.

This result would also apply as before to the second term matrix for the complete system $\underline{\hat{\theta}}$ as given in (4.50), since the 2nd derivatives of $\hat{e}_k$ are again very similar to those in (5.17).

After the estimation or climbing process had been completed we would hope that $\hat{e}_k$ would be effectively white for the finite data set N. In this case the hill top is convex, i.e. a positive definite full second derivative matrix, since the second half contribution would tend to a null matrix. This condition would only occur locally about the top of the estimation hill. Using the methods of chapters 3 and 4, the various first and second term matric~~i~~es could be calculated given $\eta, \delta, \hat{\eta}$ and $\hat{\delta}$ or $\underline{\hat{\theta}}$ and $\underline{\theta}$. Thus some idea of the size of the convex region about the matching point could be obtained for particular cases.

The first term matrix is positive definite globally i.e. under any mis-match conditions. There would be some justification therefore in ignoring, at some cost of climbing efficiency, the second term matrix altogether and only employing the first term matrix in the climbing procedure. This was done by Åström for his Newton-Raphson method.

CHAPTER 6

EXAMPLES AND RESULTS

6.1 The Estimation Program.

A program was written in Fortran IV for use on an IBM 7090/94 computer as an implementation of the work described in this thesis. This program is shown as a flow chart in figure 21 as a number of subroutines each of which provide an individual utility. The purpose of the program was to estimate a system described by Z polynomials as in (1.38) from a data record y_k, u_k ; k=1, ... N derived from experimental work on a plant. It is well known that data collection in the field is difficult and time consuming. All of the following examples therefore have artificial data records which have been generated digitally within the computer. There is the natural advantage that the true generating 'plant' is thus known exactly and the estimation procedures can be critically assessed for bias etc.

A permanent record of 50,000 random numbers was generated by a digital random number generator[72] with a Gaussian distribution and kept on magnetic tape. The distribution and independance of these variates were checked and are summarised in table 1. The numbers were later used for all the examples to generate the y_k, u_k records by applying them as inputs to digital models of the form of equation (1.38). One advantage of this approach is that since the entire set of numbers is stored, runs may be repeated to compare different methods. Alternatively ensembles of runs may easily be made. A second advantage arose here as it was found approximately 8 times faster to read the

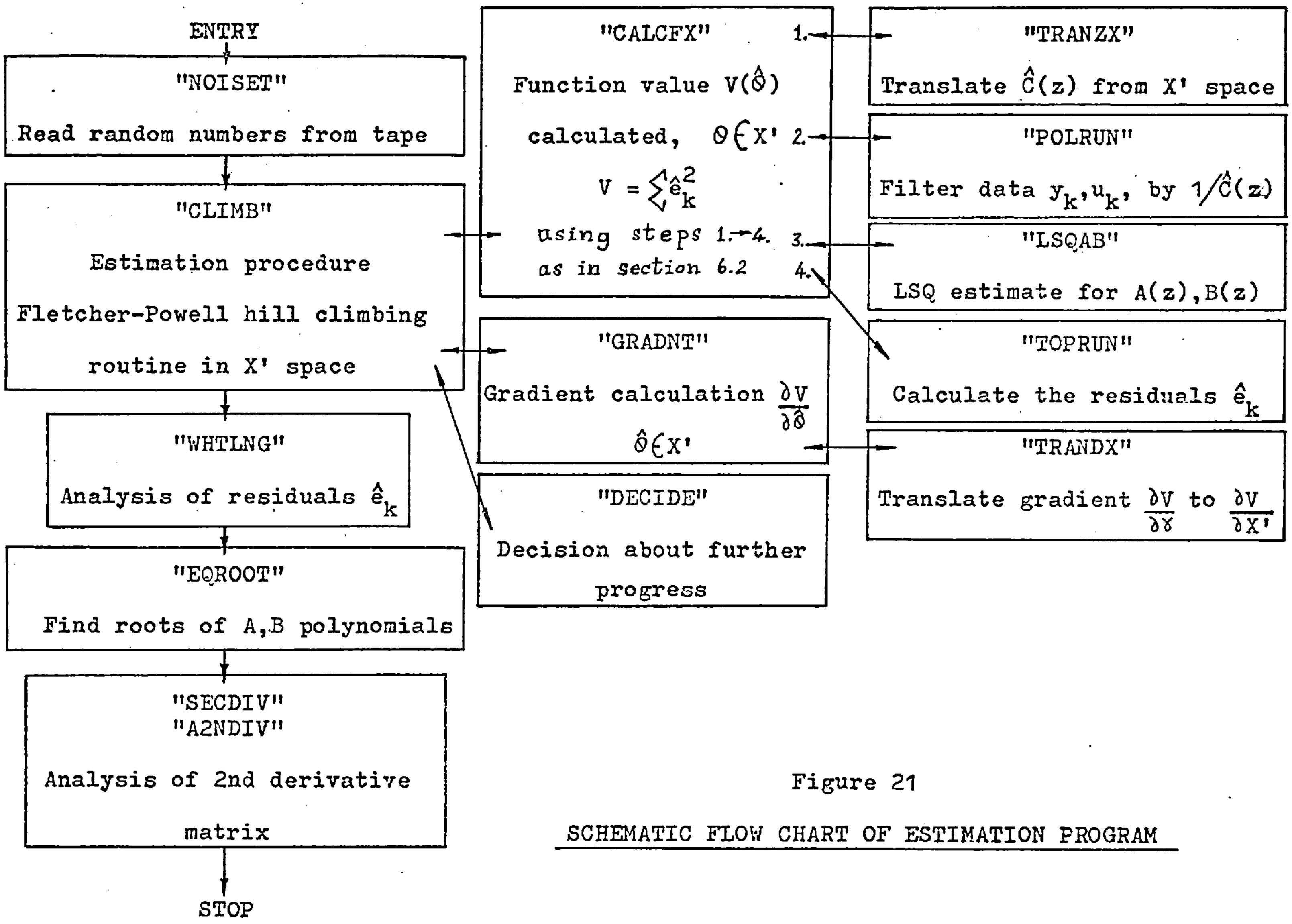

ENTRY
"NOISET"
Read random numbers from tape
"CLIMB"
Estimation procedure
Fletcher-Powell hill climbing
routine in X' space
"WHTLNG"
Analysis of residuals $\hat{e}_k$
"EQROOT"
Find roots of A,B polynomials
"SECDIV"
"A2NDIV"
Analysis of 2nd derivative
matrix
STOP
"CALCFX"
Function value $V(\hat{\theta})$
calculated, $\theta \in X'$
$V = \sum \hat{e}_k^2$
using steps 1.→4.
as in section 6.2
1.
2.
3.
4.
"GRADNT"
Gradient calculation $\frac{\partial V}{\partial \hat{\theta}}$
$\hat{\theta} \in X'$
"DECIDE"
Decision about further
progress
"TRANZX"
Translate $\hat{C}(z)$ from X' space
"POLRUN"
Filter data y_k, u_k, by $1/\hat{C}(z)$
"LSQAB"
LSQ estimate for A(z),B(z)
"TOPRUN"
Calculate the residuals $\hat{e}_k$
"TRANDX"
Translate gradient $\frac{\partial V}{\partial \gamma}$ to $\frac{\partial V}{\partial X'}$

Figure 21

SCHEMATIC FLOW CHART OF ESTIMATION PROGRAM

TABLE 1

SUMMARY OF ANALYSIS OF THE RANDOM NUMBER RECORD

Each block of 10,000 numbers were analysed for amplitude distribution and sample autocorrelation. A typical set of result is given below for one such block.

AUTOCORRELATION ANALYSIS

DELAY	r	0	1	2	3	4	5	6	7	8	9
VALUE	ϕ_r^N	.9922	.0024	-.0071	-.0037	.0012	-.0111	.0002	.0087	-.0169	.0069

DELAY	r	10	11	12	13	14	15	16	17	18	19
VALUE	ϕ_r^N	.128	-.0071	.0172	.0008	-.0032	.0105	-.0090	-.0054	-.0126	.0119

Number of values of ϕ_r^N outside ± 1.65 is 2

± 2.0 is 0

For the complete set of 50,000 numbers 4 out of 100 values of ϕ_r^N ; $r>0$, were found outside $\pm 2.\sigma$ limits, and 10 out of 100 were outside $\pm 1.65\sigma$. The expected levels are 5% and 10% respectively. It was concluded that the sequence was therefore sufficiently 'white'.

AMPLITUDE DISTRIBUTION ANALYSIS

CELL NUMBER	UPPER CELL LIMIT	EXPECTED CONTENTS	ACTUAL CONTENTS
1	-4.75	0.0369	0
2	-4.25	0.283	0
3	-3.75	2.010	3
4	-3.25	11.17	12
5	-2.75	48.60	40
6	-2.25	165.4	194
7	-1.75	440.6	440
8	-1.25	918.5	891
9	-0.75	1498.8	1491
10	-0.25	1914.6	1911
11	0.25	1914.6	1960
12	0.75	1498.8	1492
13	1.25	918.5	896
14	1.75	440.6	452
15	2.25	165.4	153
16	2.75	48.60	54
17	3.25	11.17	8
18	3.75	2.010	3
19	4.25	0.283	0
20	4.75	0.0369	0

SAMPLE MEAN 0.0070,STANDARD DEVIATION OF SAMPLE MEAN 0.0100

SAMPLE VARIANCE 0.9919,STANDARD DEVIATION OF SAMPLE VARIANCE 0.0140

VALUE χ^2 13.39 FOR 19 DEGREES OF FREEDOM values from tape than generate them from the digital routine[72] each time.

The estimation program employed the Fletcher-Powell[9] algorithm(2.68) for hill climbing. This method requires both function values and derivatives, and forms an estimate of the inverse of the second derivative matrix. This estimate converges for quadratic hills to the true value. For non-quadratic hills the estimate is forced to be positive-definite, which enables the routine always to proceed in a beneficial direction. Other routines such as the Newton-Raphson can easily get into difficulties on surfaces with non-positive definite second derivatives.

Fletcher and Powell recommend cubic minimisation as used by David~~s~~on to obtain a minimisation along a line. In our experience quadratic minimisation[6] remains better conditioned in difficult cases although theore~~c~~tically less efficient. Such a quadratic minimisation[8] only requires function values and saves some computation compared to the cubic method which also requires derivatives. After a minimum along a line has been achieved, the local first derivative can be computed and used to update the Fletcher-Powell algorithm. A worthwhile addition to this minimisation has been found to check the orthogonality of the initial and final gradients by comparing their projections on the line of search. Local minimisation by costing this orthogonality condition is helpful here as it ensures that the Fletcher-Powell estimated matrix is updated with correct information and does not become nearly singular in difficult situations.

6.2 The least squares procedure.

Throughout this thesis we have so far advocated the estimation of a system in terms of the roots of its component Z polynomials. This implies, for the system of (1.38), hill climbing in 3n+2 parameters, which are the roots of the A,B, and C polynomials together with G and χ' as in section 4.6. The value of λ can be calculated after the climbing has finished as shown in section 2.5. The n initial conditions have been ignored as described in chapter 5. If we regard the estimation process as purely one in hill climbing, the dimensionality (3n+2) is rather large for n greater than 2, and it would be desirable if this could be reduced to give a more practical scheme.

As mentioned in section 2.3, one advantage of using a coefficient description for the $\hat{A}$ and $\hat{B}$ polynomials is that a least squares solution can readily be obtained for these coefficients for a chosen value of $\hat{C}(z)$. This scheme was adopted to reduce the dimensionality of the hill to only (n+1). At each iteration the n roots of $\hat{C}(z)$ were decided by the climbing algorithm using the X' transformation, and the data set y_k, u_k was filtered by $1/\hat{C}(z)$ as in equation (2.37). The least squares procedure of (2.11) was then applied to the new filtered data set y_k^*, u_k^* to give estimates of the coefficients $a_1 \ldots\ldots a_n, b_o, b_1 \ldots\ldots b_n$ The gain term $\hat{G}_o$ has been implicitly included in the $\hat{b}_i$ coefficients and the $\hat{b}_o$ term provides the extra degree of freedom. After the hill climbing procedure has finished, the roots of the $\hat{A}$ and $\hat{B}$ polynomials of degree n can be found. This process is in itself non-linear and time consuming for large n, and is therefore not done at each iteration.

From a simple viewpoint the D.C. term χ' can be estimated from the mean values of the signals y_k and u_k. This estimate cannot be improved by any value of $\hat{C}(z)$. Thus the χ' term may well be estimated and extracted from the data set before the main procedure starts. This means the climbing dimensionality can be reduced further, from (n+1) to n. A fuller discussion of the worth of this approach is given later in section 6.9.

6.3 The total cost derivative.

If the decision is taken to adopt the above methods of section 6.2, the total derivative of the cost with respect to $\hat{C}(z)$ should be examined (6.1), since $\hat{C}(z)$ is fixed when estimating A and B polynomials. Define the cost $V(\underline{\hat{\theta}})$ as in (6.1)

$$\text{Cost } V \triangleq V\left[\hat{A}(\hat{C}),\hat{B}(\hat{C}),\hat{C}\right] = V\left[\hat{C}\right]$$

$$\text{then } \frac{dV}{d\hat{C}} = \left.\frac{\partial V}{\partial \hat{A}}\right|_{\hat{C}} \cdot \frac{d\hat{A}}{d\hat{C}} + \left.\frac{\partial V}{\partial \hat{B}}\right|_{\hat{C}} \cdot \frac{d\hat{B}}{d\hat{C}} + \left.\frac{\partial V}{\partial \hat{C}}\right|_{\hat{A},\hat{B}} \qquad (6.1)$$

Because of the least squares algorithm the partials $\left.\frac{\partial V}{\partial \hat{A}}\right|_{\hat{C}}$ and $\left.\frac{\partial V}{\partial \hat{B}}\right|_{\hat{C}}$ will be zero for a given value of $\hat{C}$. The derivative $\frac{dV}{dC}$ is also zero at the optimum value of $\hat{C}$ i.e. at the top of the hill in $\hat{C}$. The maximum likelihood estimate is only achieved when the all the partials in $\hat{A},\hat{B}$, and $\hat{C}$ are zero, and then a small movement in any direction in the space of $\hat{A},\hat{B}$, and $\hat{C}$ will give zero estimation cost change.

Consider a simple hill shown in figure 12. The contours are drawn for equal estimation cost V for a system consisting of one parameter in $\hat{C}$ and one in the $\hat{A}$ polynomial. The full dimensional

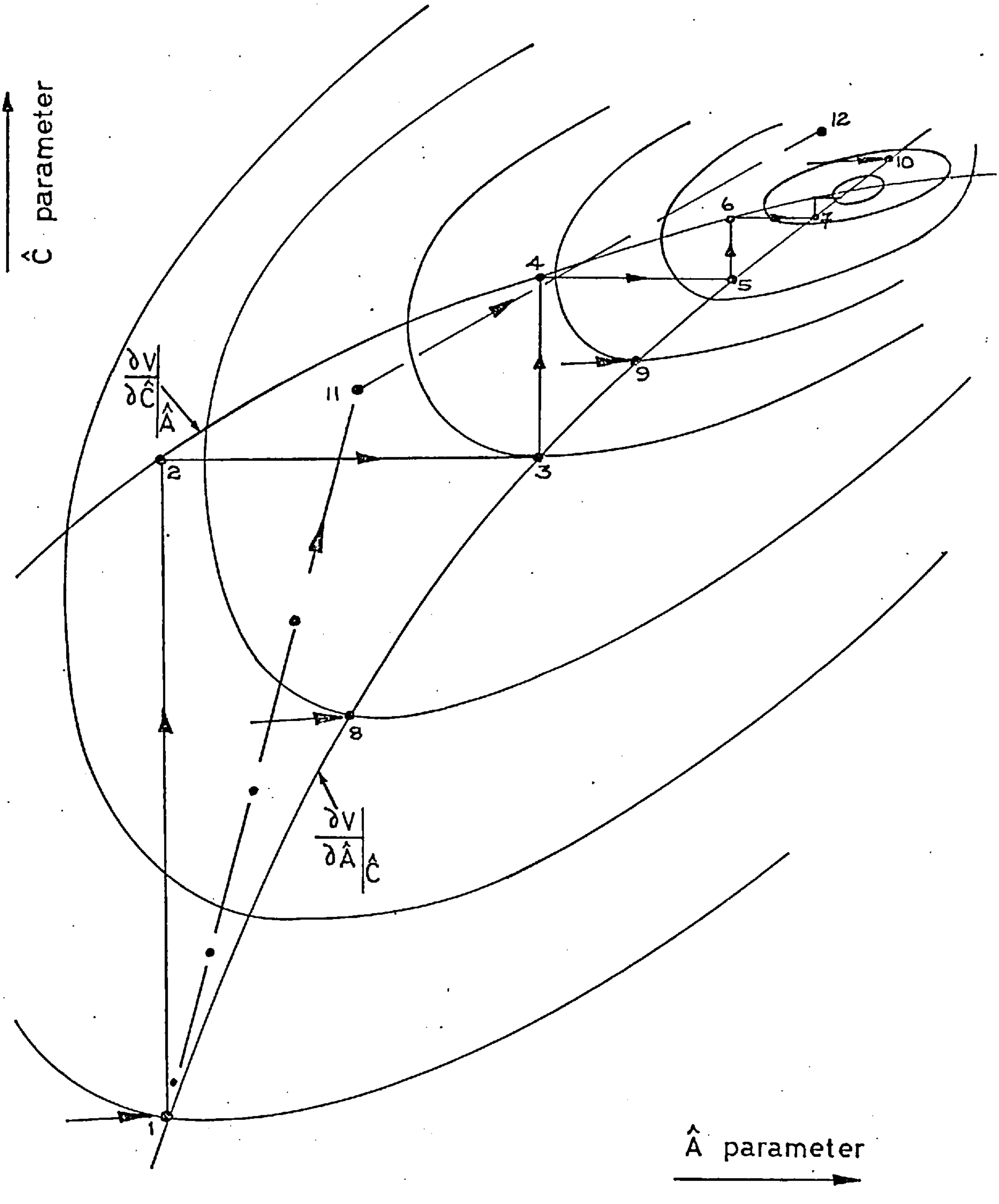

$\hat{C}$ parameter
$\left.\frac{\partial V}{\partial \hat{C}}\right|_{\hat{A}}$
$\left.\frac{\partial V}{\partial \hat{A}}\right|_{\hat{C}}$
1
2
3
4
5
6
7
8
9
10
11
12
$\hat{A}$ parameter

FIG. 22 Simple hill in $\hat{C}, \hat{A}$ space

climbing method might take a path such as 1, 11, 12, etc. using a conjugate gradient method in all the dimensions at once. Clarke's method in section 2.3 solves first for the least squares solution in $\hat{A}$ for constant $\hat{C}$ to get to point 1. He then treats $\hat{C}$ as if it described an auto-regressive process and uses the least squares algorithm to solve for point 2. The iterations continue to switch as shown along the path 1,2,3,4,5,6, etc. between the lines $\left.\frac{\partial V}{\partial \hat{A}}\right|_{\hat{C}} = 0.$ and $\left.\frac{\partial V}{\partial \hat{C}}\right|_{\hat{A}} = 0.$ The problem alternately has the dimensionality of $\hat{A}$ and then $\hat{C}$.

The method of Steiglitz^5 and also of section 6.3 chooses a value of $\hat{C}$ and then solves for $\hat{A}$ at point 1 using the least squares algorithm. The hill climbing method then rechooses $\hat{C}$ and the process repeats along a path such as 1,8,9,10, lying on the line $\left.\frac{\partial V}{\partial \hat{A}}\right|_{\hat{C}} = 0.$ When the cost gradient with respect to $\hat{C}$ is required, strictly only $\left.\frac{\partial V}{\partial \hat{C}}\right|_{\hat{A},\hat{B}}$ is evaluated in place of the full version of (6.1). However such an evaluation is made under the condition $\left.\frac{\partial V}{\partial \hat{A}}\right|_{\hat{C}} = 0.$ and $\left.\frac{\partial V}{\partial \hat{B}}\right|_{\hat{C}} = 0.$ and is valid locally. The situation clearly is improved if all the contours have a common centre, i.e. a pure quadratic surface.

In practise we have found that if the $\hat{C}(z)$ polynomial is only very roughly similar to the true C(z), the estimates $\hat{A}$ and $\hat{B}$ are reasonably close to their true values A and B. This so called 'ball park' effect has been noted by other research workers and supports the bias reduction methods described in chapter 2. The strict maximum likelihood estimate is only achieved when all the 1st derivatives including that in $\hat{C}$ are zero i.e. at the exact top of the hill. A practical application might stop short of this condition in the sense that once within the 'ball park' only very small cost improvements are likely.

6.4 The number of multiplications.

The major computation work in estimating parameters from a long data set N is the filtration of the records y_k, u_k by the Z polynomials $\hat{A}, \hat{B}$, and $\hat{C}$. After the cost and its derivatives have been formed the rest of the climbing and other routines require relatively little computer time. We should therefore be interested in using efficient methods such as in (4.6) to reduce the filtering work to a minimum. Naturally to this end all the short ways of calculating derivatives should also be employed as in equations (4.2), (4.3) and (4.11).

For the full dimensional hill climbing approach, i.e. expressing all the polynomials by their roots, the w_k and v_k signals formed in (4.29) and (4.30) require 4nN multiplications and 8nN additions. This arises through using the method of (4.6) for filtering. Advantage can also be taken of the fact that computers such as the IBM 7094 and 360 take virtually no extra time for complex arithmetic than for real. The estimation cost V can be totalised for a further N multiplications and additions. The total work is thus a linear function of n and N.

For the alternative approach via the least squares algorithm in section 6.2, we first form y_k^* and u_k^* as (2.37) but using (4.6). This requires nN multiplications and 2nN additions. To form the matrix in (2.11) we apparently need $\frac{1}{2}(2n+1) * (2n+2)$ N multiplications and additions, since (2n+1) is the number of free coefficients in $\hat{A}(z)$

and $\hat{B}(z)$. A further (2n+1)N multiplications and additions are also required to calculate the estimation cost V in (2.53) using (2.7). The computation appears to be a function of n^2 for this case. However by careful inspection of the form of the matrix M^tM in (2.11) it can be seen that there are numerical dependances between the terms which reduce the number of multiplications so that the total work varies linearly with both n and N.

The first derivatives of the cost for the full root description method are calculated with the methods of section 4.5 and require 2N additions and 2N multiplications for each of the 3n+1 components. The alternative mixed approach has it's 1st derivatives in M^tY of (2.11) and requires 2.(2n+1) shifts and additions for the $\hat{A}$ and $\hat{B}$ coefficients. The $\hat{C}$ root derivatives then are calculated as in section (4.5) with 4nN additions and 2nN multiplications. The work for the derivatives is linear in n and fairly equal for the two schemes.

6.5 Example No.1.

This problem was taken as a standard example similar to those in the literature[10,11,27,28,37], so that sensible comparisons could be made. The example was tried both with the new estimation program described in sections 6.1 and 6.2, and with Åström's method which was also available. The random number tape was beneficial here as an ensemble of results were easily obtained and could be precisely repeated using both methods. The polynomials used in the true process generating the plant data are given in (6.2)

$$y_k = G_o \ast \frac{B(z)}{A(z)} u_k + \lambda \frac{C(z)}{A(z)} e_k + \chi' \ ; \ k=1, \ \ldots\ldots \ 200$$

$$A(z) = 1.0 - 1.5z^{-1} + 0.7z^{-2} \quad \text{,Roots at } 0.75 \pm j0.3708$$
$$B(z) = 1.0 - 1.0z^{-1} + 1.0z^{-2} \quad \text{,Roots at } 0.50 \pm j0.866$$
$$C(z) = 1.0 - 1.65z^{-1} + 0.695z^{-2} \text{,Roots at } 0.825 \pm j0.12$$

(6.2)

The values of G_o(also named b_o) and λ were both 1.0,and χ' was 0.0 The sequences e_k and u_k were taken from the random number tape for each member of the ensemble, and scaled to have zero mean over the data length of 200. The summary of an ensemble of 10 runs for both estimatio methods is given in tables 2 and 3.

For this example the data length is sufficient by the criteria of chapter 4 that the difficulties mentioned there do not arise to any degree. Both methods give essentially the same estimates for the respective ensemble members within errors due to the computer word length. The stopping criteria for the two methods were not easily made compatible as they are working in different space descriptions. As a result the cost minima of table 2 are slightly lower than those of table 3. For the climbing procedure used by Åström, several iterations have to be repeated at half the step length when a failure has occured in the Newton-Raphson algorithm. Occasionally steps have also to be reversed in direction due to the non-positive definiteness of the second derivative matrix. The number of both these occurrences are given in table 2.

Using the schemes described in section 4.7 the matrix of second derivatives of the cost, $Q = \frac{\partial^2 V(\hat{\theta})}{\partial\hat{\theta}_i \partial\hat{\theta}_j}$ was calculated at the final

Final	Number of			Sum	Value of estimated coefficients							Value of
Cost	Iterations	H	R	Sq.	$\hat{c}_1$	$\hat{c}_2$	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_0$	$\hat{b}_1$	$\hat{b}_2$	$\hat{\chi}'$
=Nλ/2.				Slopes								
96.386	10	10	2	.0007	-1.722	0.7759	-1.484	0.6938	0.9988	-0.9854	1.050	$0.344*10^{-3}$
96.487	12	22	2	.060	-1.674	0.7074	-1.546	0.7443	0.9865	-1.1107	1.091	$0.528*10^{-3}$
96.834	14	12	2	.0002	-1.677	0.7158	-1.514	0.7026	0.9895	-0.9890	0.927	$-0.509*10^{-4}$
93.905	13	28	2	.02	-1.724	0.7750	-1.513	0.7159	0.9986	-1.0120	1.031	$-0.135*10^{-2}$
98.032	14	22	4	.002	-1.676	0.7191	-1.514	0.7133	0.9878	-0.9383	0.919	$-0.601*10^{-3}$
97.717	16	22	1	.001	-1616	0.6698	-1.504	0.6978	1.0415	-1.0789	1.043	$-0.246*10^{-3}$
98.069	10	8	2	.0006	-1.636	0.6836	-1.498	0.7035	1.0581	-1.1763	1.134	$0.348*10^{-3}$
98.593	12	22	1	.001	-1.622	0.6813	-1.525	0.7163	0.9245	-0.8650	0.894	$-0.339*10^{-4}$
96.136	12	25	2	.000004	-1.649	0.7207	-1.537	0.7287	0.9236	-0.9773	1.001	$-0.283*10^{-3}$
97.075	15	31	3	.002	-1.621	0.6652	-1.1503	0.6948	0.9921	-0.8778	0.868	$0.233*10^{-3}$

Table 2: Summary of ensemble of 10 runs for example 1, Åström's method.

Final Cost $=N\lambda/2.$	Iter.	Final Slope	Value of estimated roots $\hat{C}$	$\hat{A}$	$\hat{B}$	Value of $\hat{G}_0$	Value of $\hat{\chi}'$	Value of χ^2
96.390	6	.007	0.8614 ±j0.1838	0.7564 ±j0.3786	0.5046 ±j0.8807	0.9969	$-0.415*10^{-2}$	8.124
96.506	7	.003	0.8363 ±j0.0765	0.7570 ±j0.3742	0.4724 ±j0.8448	0.9781	$-0.125*10^{-2}$	18.004
96.835	7	.005	0.8386 ±j0.1128	0.7424 ±j0.3778	0.4947 ±j0.8972	1.0032	$0.263*10^{-2}$	18.411
93.962	5	.006	0.8622 ±j0.1779	0.7730 ±j0.3829	0.5644 ±j0.8859	0.9954	$-0.140*10^{-2}$	34.697
98.079	9	.0013	0.8373 ±j0.1253	0.7561 ±j0.3591	0.4979 ±j0.8225	1.0022	$0.368*10^{-3}$	9.879
97.720	6	.004	0.8081 ±j0.1297	0.7519 ±j0.3642	0.5201 ±j0.8570	1.0435	$-0.212*10^{-3}$	4.840
98.078	6	.003	0.8187 ±j0.1193	0.7495 ±j0.3774	0.5576 ±j0.8700	1.0653	$-0.399*10^{-2}$	6.805
98.590	8	.003	0.8110 ±j0.1544	0.7626 ±j0.3674	0.4669 ±j0.8660	.9877	$0.436*10^{-3}$	8.597
96.139	6	.002	0.8249 ±j0.2019	0.7685 ±j0.3716	0.5289 ±j0.8942	1.0013	$-0.269*10^{-2}$	10.225
97.077	9	.0006	0.8107 ±j0.08864	0.7516 ±j0.3604	0.4423 ±j0.8241	.9908	$0.848*10^{-3}$	14.759

Table 3: Summary of ensemble of 10 runs for example 1: New method of section 6.1 and 6.3

estimate $\hat{\underline{\theta}}$ for each ensemble member of table 3, the new estimation method.

The classical statistical approach of section 4.13 allows us to define a confidence region, in this case an elipsoid, which would cover the true parameters $\underline{\theta}$ with a probability of 95%. These elipsoids are individual to each data set of the ensemble and would give a confused effect if plotted together. Since this example has fixed known parameters $\underline{\theta}$ defined in (6.2), the error vector $\tilde{\underline{\theta}} \triangleq \hat{\underline{\theta}}-\underline{\theta}$ can be calculated for each ensemble member. Thus as an alternative presentation of the results, we can calculate the statistic $\tilde{\underline{\theta}}^{t}Q\tilde{\underline{\theta}}$ which has a χ^2 distribution due to its quadratic form. These values are shown in the final column in table 3. The value of χ^2 is 16.9 for 8 degrees of freedom at a 95% confidence level and since the majority of the values of $\tilde{\underline{\theta}}^{t}Q\tilde{\underline{\theta}}$ lie within this limit, the estimation procedure can be regarded as statistically satisfactory.

The Bayesian theory of section 4.13 can also be applied to this problem. Conveniently we know the true value of $\underline{\theta}$ and we can therefore calculate the expected second derivative of cost matrix using the methods of section 4.8. This matrix can be used together with section 4.9 to define an elipsoid for the Bayesian approach which have a 95% probability of containing the random $\hat{\underline{\theta}}$ values derived later from the estimation process. Projections of such an elipsoid for two parameters $\underline{\theta}_2$ in $\underline{\theta}$ have been drawn in figures 23 to 25 together with the various estimates. These elipses of projection are derived by inverting the partition of the covariance matrix concerning the two parameters $\underline{\theta}_2$ to give Q_2. The magnitude of these two dimensional elipses is given

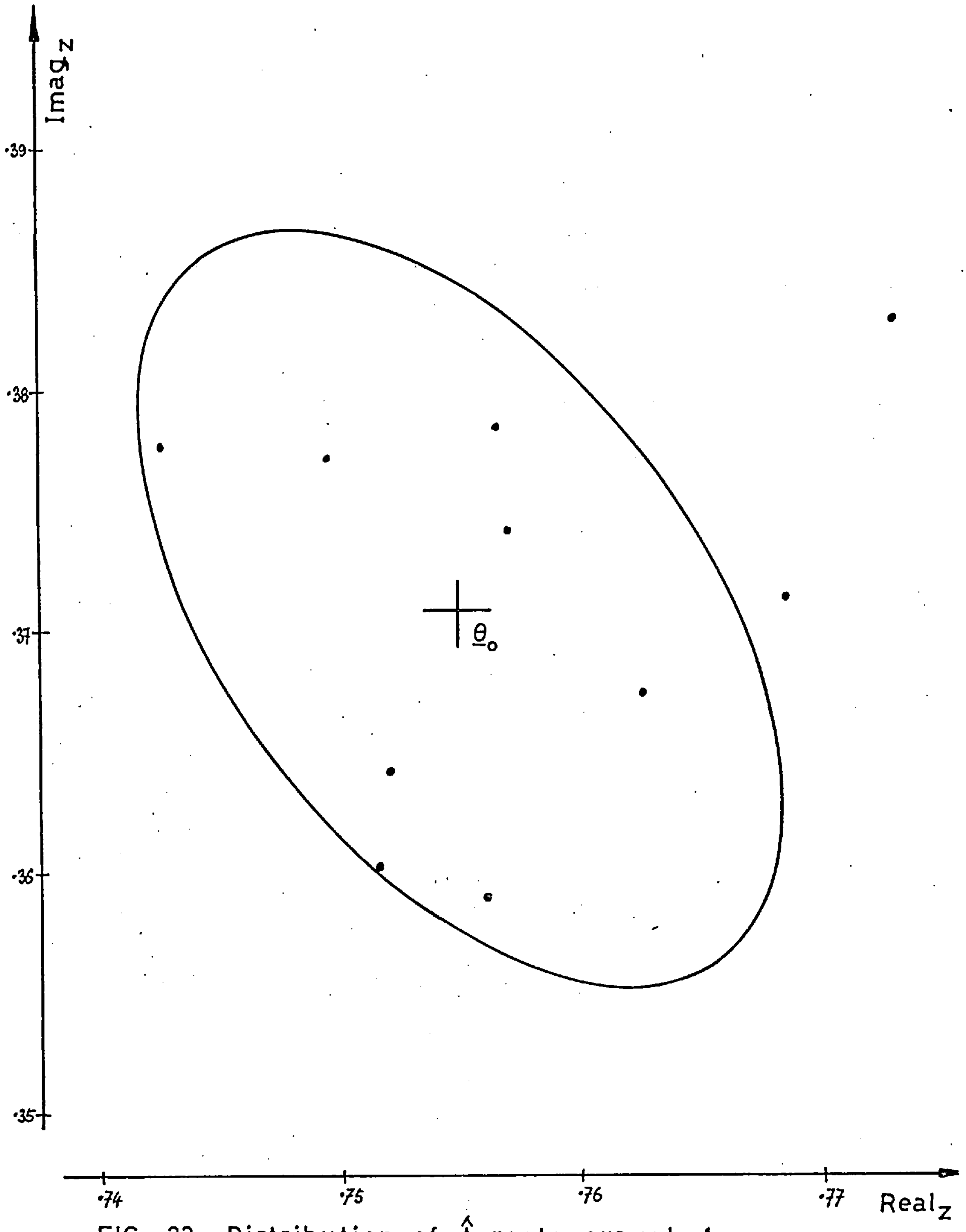


FIG. 23 Distribution of $\hat{A}$ roots, example 1

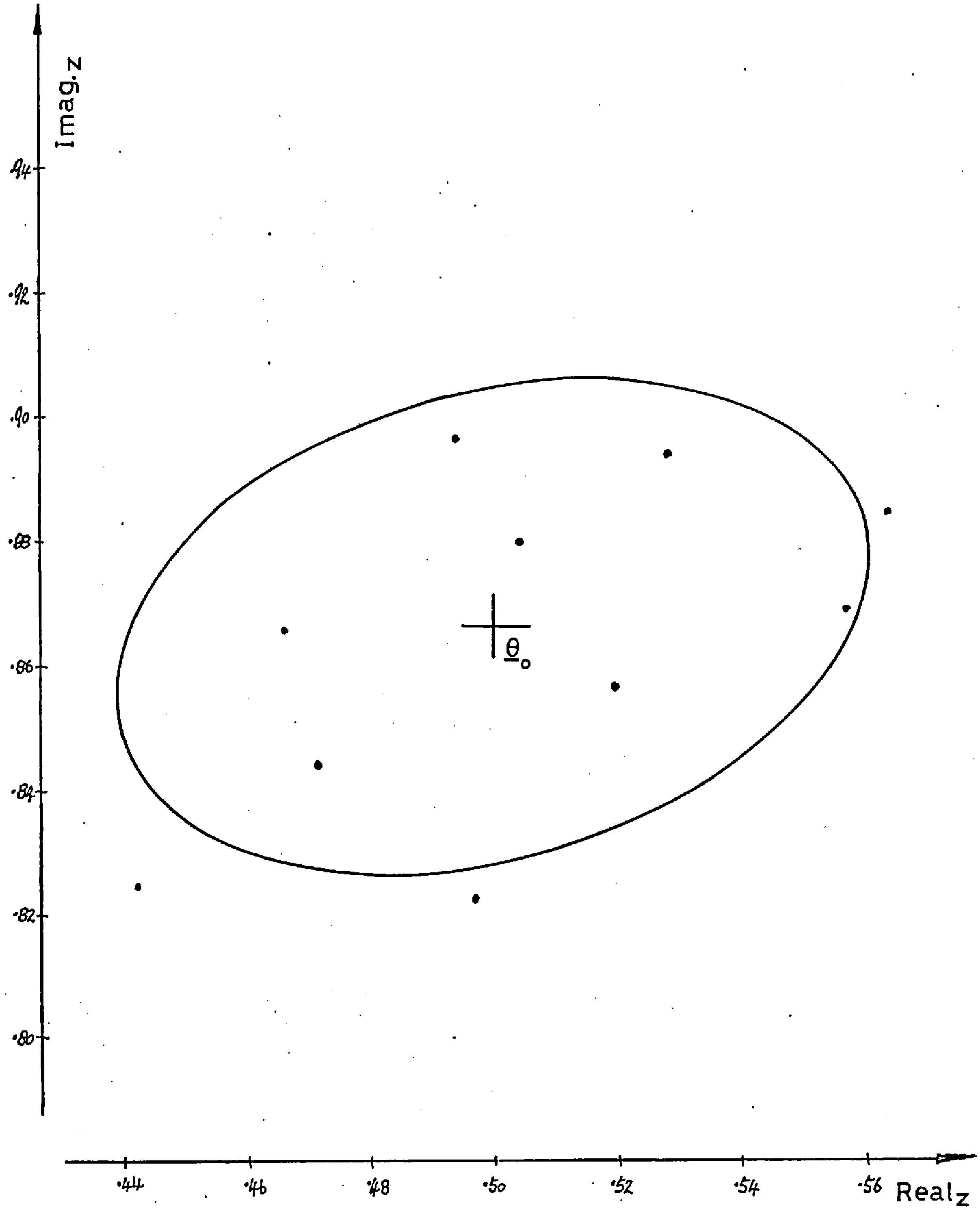


FIG 24 Distribution of $\hat{B}$ roots , example 1

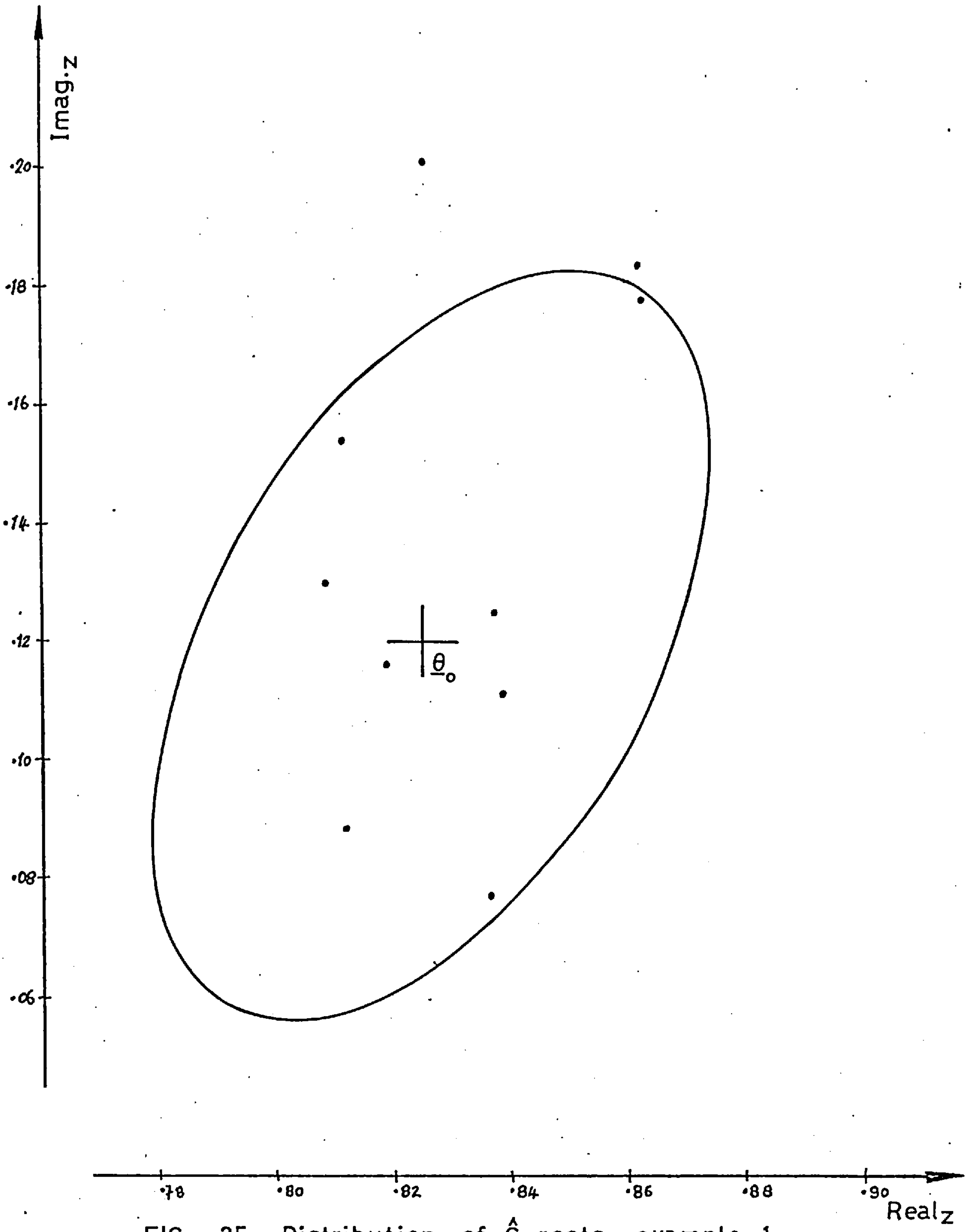


FIG. 25 Distribution of $\hat{C}$ roots example 1

by the value of χ^2 for 2 degrees of freedom at a 95% confidence level. Clearly the elipses contain most of the estimated values, and therefore confirm the estimation procedure as satisfactory in the Bayesian sense.

The sample correlation of the residuals $\hat{e}_k$ can also be examined to verify that the sequence is satisfactorily 'white' or independant. If this is so then it can be concluded that all the possible information has been extracted from the data. This criterion bears a relationship to the order of the model n which is fitted to the data, as discussed in section 1.11. We would expect that the residuals would not be white for n less than some value n^* and that a plateau of performance index would be achieved for $n \geqslant n^*$.

These effects are shown in figure 26. Here the residual 'colour' is shown for the 3rd member of the ensemble when the model is estimated with order n=1,2 and 3. Using the result of (A.14) in appendix 1 we would expect that 5% of the ordinates would be outside the limits ± 0.141 for delays $\tau \neq 0$. This corrosponds to the usual $\pm 2\sigma$ limit of a normal distribution with a variance σ^2 of $\hat{\lambda}^4/N$, where N=200. Obviously the model with n=1 does not satisfy this criterion, while those with n=2 and 3 are acceptable. A plateau of estimation cost appears to have been reached for $n \geqslant 2$. The total costs V are 307.13, 96.835 and 96.828 for n=1,2 and 3 respectively.

It will be noted that when the residuals are white following the estimation procedure, the value of $\hat{\lambda}$ is significantly less than the original 1.0 used when generating the data in (6.2). Now the N sequence for e_k was drawn from the long record of random numbers described by table 1, which had been scaled to have unit variance. The estimation

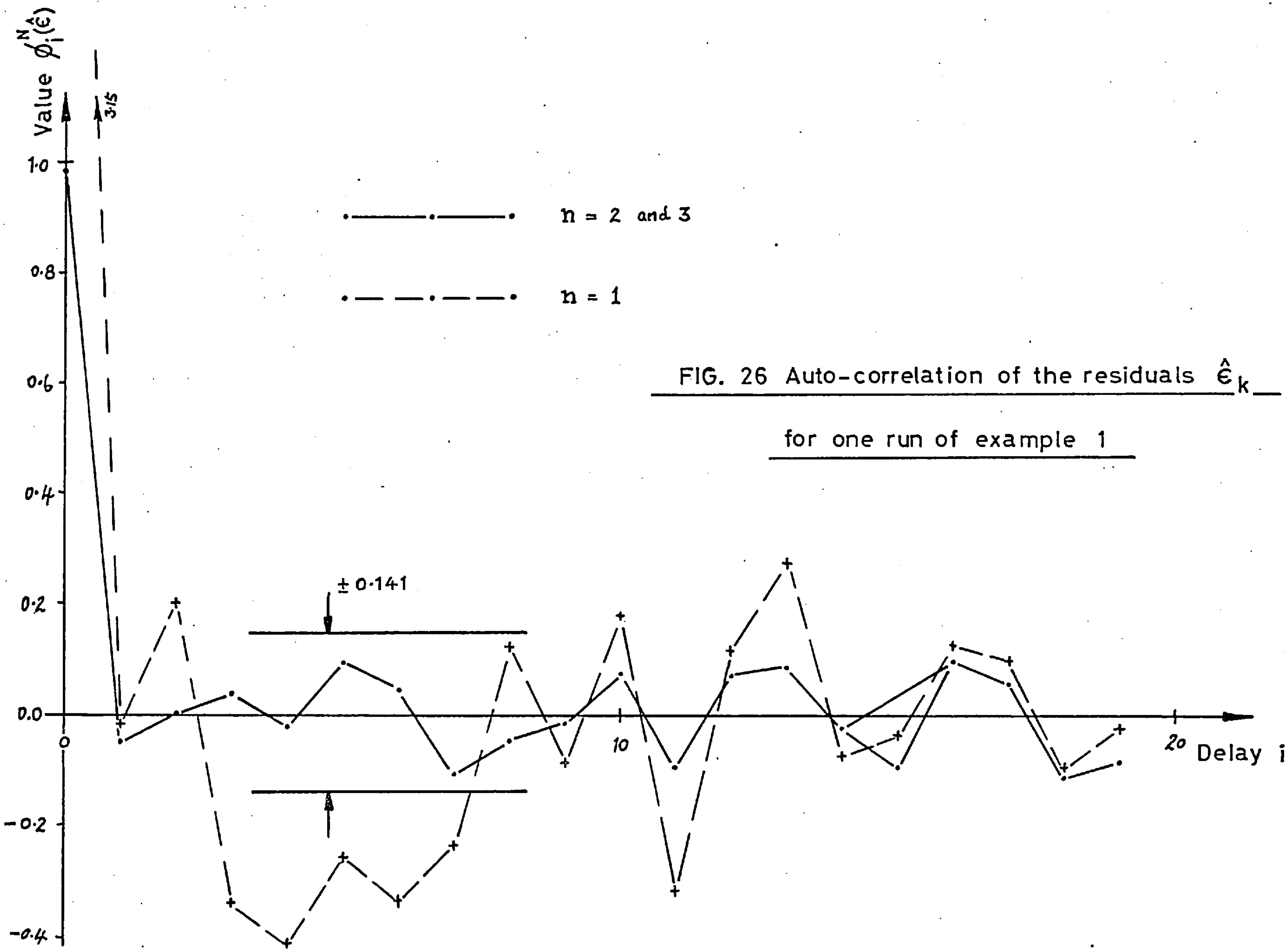


FIG. 26 Auto-correlation of the residuals $\hat{\epsilon}_k$ for one run of example 1

	n=1	n=2	n=3
a_1	0.0519	0.0153	0.0199
a_2	-	0.0143	0.0144
a_3	-	-	0.0113
$b_0,(G_0)$	0.126	0.0720	0.0720
b_1	0.145	0.0719	0.1330
b_2	-	0.0692	0.0971
b_3	-	-	0.0912
c_1	0.0973	0.0371	0.0508
c_2	-	0.0579	0.0609
c_3	-	-	0.0462
χ'	0.110	0.00278	0.00545

Table 4 : standard deviation of parameter estimates for 3rd ensemble member.

procedure appears to have the ability to further 'whiten' a relatively short sequence N drawn from a longer sequence which is already nominally 'white' i.e. random and uncorrelated. This is not unexpected as we cannot expect that a random sequence will have the same statistical properties for both short (N) and very long records (table 1).

If the model order n was higher than the plant order n^* producing the data, we would expect that the parameter estimates would be over-determined and have wider confidence limits. Experimentally we found that the matrix of second derivatives does tend towards singularity as n is increased larger than n^*. The estimated standard deviation of the parameters is shown in table 4 for n=1,2, and 3. From these figures it is obvious that the estimates for n=1 and n=3 have wider confidence limits than those for n=2. This could be regarded as a suitable indication that the original plant order was equal to two.

6.6 Example No.2

This example was chosen to be a difficult problem which should show the advantages of the estimation methods advocated in this thesis. The true plant parameters are given in (6.3). These were chosen to give a wide spread of roots for A(z), i.e. radii of 0.80 and 0.99 in the Z plane, and a pair of complex roots for C(z) close, (radius 0.99), to the unit circle

$$y_k = G_{o*} \frac{B(z)}{A(z)} u_k + \lambda_* \frac{C(z)}{A(z)} e_k + \chi' \quad ; \quad k=1, 200$$

where

$$A(z)=1.0-2.3766z^{-1}+2.0134z^{-2}-.63426z^{-3}\text{;Roots at } 0.6933 \pm j0.400,\ 0.99$$

$$B(z)=1.0-3.50z^{-1} + 3.0z^{-2} \qquad \text{;Roots at } 1.5 \text{ and } 2.0$$

$$C(z)=1.0-1.9572z^{-1} + .95766z^{-2} \qquad \text{;Roots at } 0.9786 \pm j0.1500$$

$$G_o = 1.0 \qquad \lambda = 1.0 \qquad \chi' = 0.0 \tag{6.3}$$

As a further difficulty the roots of B(z) lie outside the unit circle, thus giving a non-minimum phase system as described in section 3.14. The data length of 200 is considered short by the criteria of chapter 4, and the corrosponding effects arise when estimating the plant parameters.

The impulse response of B(z)/A(z) is shown in figure 27. This shows a non-minimum phase response typical of physical systems such as rear steered rockets and drum boilers. The spread of eigenvalues is demonstrated by the high frequency ripple on the longer term response. The impulse response of 1/C(z), which is significant in equation (2.49), is oscillatory with a period of k=41 and a damping such that the envelope is approximately $\frac{1}{4}$ of its initial magnitude at k=200.

The performance of the new estimation method on this problem was in general better than Åström's method which took generally more iterations and displayed poor convergence properties. A typical run is detailed in table 5. Åström's method failed to converge in 30 iterations and estimated the roots of C(z) as lying outside the unit circle. This occured in about 30% of the runs with this example and would be of little practical use as described in section 2.6. The large number of steps of the algorithm which were halved or reversed is evidence of the difficulty of convergence. The reason behind this

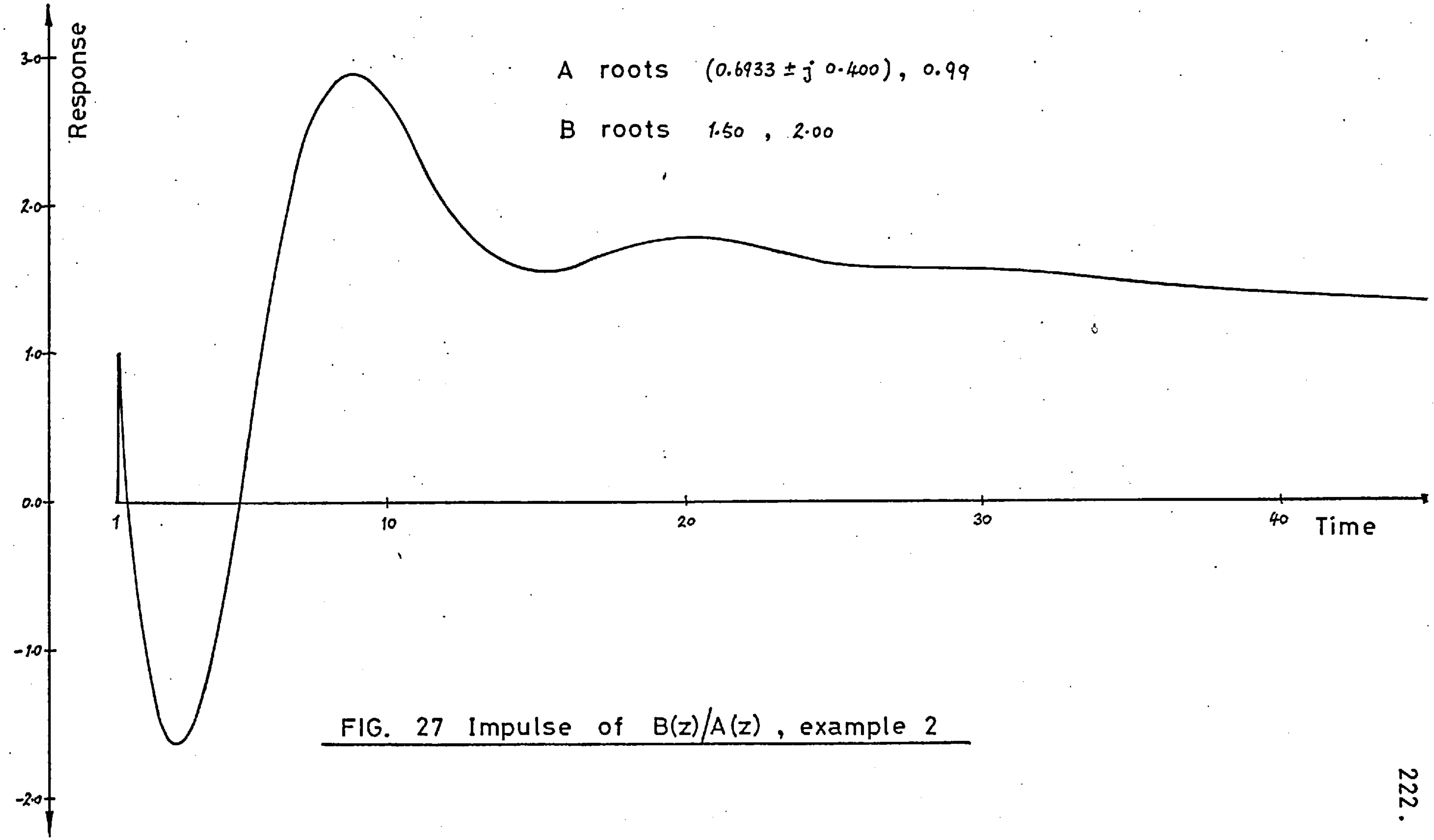


FIG. 27 Impulse of B(z)/A(z) , example 2

is more clearly shown by example 3.

As shown in table 5, the new estimation method was fairly efficient in comparison although having a higher final estimation cost. This was principally due to the automatic decision taken at a pole radius of 0.99809 that the data length and pole position were becoming incompatible under the criteria in chapter 4. The value of $\left|\hat{\gamma}\right|^{2N}=0.468$ This particular factor arises for a complex root pair from the $(\delta_i\delta_n)^N$ term in (3.41) when calculating the expected sample variance and occurs in most of the criteria of section 4.12. Such check calculations can easily be included in the algorithm at each iteration of the climbing routine.

A similar decision was previously made at a radius of 0.97613 to change from single precision working to double precision. Experimental evidence had previously shown in many cases that such a move was wise beyond a radius of 0.97. This action was taken to reduce the random noise introduced by the finite digital word length. There is otherwise a noise term introduced into the numerical calculations and similar effects are seen to those described in section 4.10. As the filter poles become stronger a 'roughness' is introduced in computing the cost V, and this affects the logic of the hill climbing routine which inherently assumes a smooth function. Naturally these effects are also present in Åström's method, but they cannot readily be checked without repeatedly solving for the polynomial roots.

		Åström's Method	New Method
Number of iterations		30	8
Number of steps halved		41	-
Number of steps reversed		9	-
Number of unstable costs evaluated		29	Nil
Final estimation cost		91.107	95.288
Final slope		1332.9	(157.05, 57.38)
$\hat{A}(z)$ polynomial	a_1	-2.3426	roots(0.67836 ± j0.40438),
	a_2	1.9657	and 0.98988
	a_3	-0.62040	
$\hat{B}(z)$ polynomial	b_0	1.0245	roots 2.1591
	b_1	-3.5581	and 1.4513
	b_2	2.96811	
$\hat{C}(z)$ polynomial	c_1	-2.0742	roots(0.98473 ± j0.15437)
	c_2	1.1529	

Table 5: Comparison of Åström's method and the New estimation method for one run of example number 2.

6.7 Example No.3

The true parameters for this example are given in (6.4)

$$y_k = G_o \cdot \frac{B(z)}{A(z)} u_k + \lambda \cdot \frac{C(z)}{A(z)} e_k + \chi' \quad ; \quad k=1,\ 50 \qquad (6.4)$$

$A(z)$ roots at $(-0.20, \pm j0.9288)$

$B(z)$ roots at $(-0.9288, \pm j0.20)$

$C(z)$ roots at $(1.01175 \pm j0.1500)$; radius 1.023

$G_o, \lambda = 1.0 \quad ; \quad \chi' = 0.0$

The data length is very short since the example was only used to investigate the effect of evaluating the estimation cost V at points where the roots of $\hat{C}(z)$ lay outside the unit circle. For one set of data y_k, u_k the cost V has been plotted before as figure 18 against pole position along a line of search in the Z plane outside the unit circle.. A basically smooth function is indicated which is perturbed by increasing am~~p~~ounts of added positive noise as the pole radius increases. This noise is due to the random round-off errors of the finite digital word length (8 decimal places), and disappears at these radii for double precision working. Any small roundoff error in the digital computation is amplified by the action of the unstable $1/\hat{C}(z)$ filter until its significance is much greater. The basic function is smooth as this has been recalculated in double precision arithmetic. The true optimum in this case for the given data set lies at $1.0976 \pm j0.2015$ with a cost of 18.140.

It will be noted that all the noise perturbations are additive due to the definition of V as $\sum_{k=1}^{N} \hat{e}_k^2$, with $\hat{e}_k$ derived using the

$1/\hat{C}(z)$ filter. This example vividly demonstrates the difficulties which can occur with poles outside the unit circle; many climbing routines would have their logic destroyed when the function V was far from smooth.

6.8 Examples No.4 and 5

Equation (6.5) gives the parameters for example 4. The parameters were chosen to give a working example with a long data length but with a strong complex pole pair in $1/C(z)$. Again the comparison was made between Åström's method and the new method.

$$y_k = G_{o*} \frac{B(z)u_k}{A(z)} + \lambda . \frac{C(z)}{A(z)} e_k + \chi' \quad ; k=1, 1000$$

$$\begin{array}{lll} A(z) & \text{roots at } (-0.20, \pm j0.9288) & \\ B(z) & \text{roots at } (-0.9288, \pm j0.20) & \\ C(z) & \text{roots at } (0.9740, \pm j0.1500) & , \text{Radius } 0.985 \end{array}$$

$$G_o = 3.0 \quad ; \quad \lambda = 1.0 \quad ; \quad \chi' = 0.0 \qquad (6.5)$$

Figures 28 and 29 illustrate the progress made by each method for a few typical estimation runs. In general the new method was superior and achieved faster convergence. This was again aided by switching to double precision working for radii beyond 0.970.

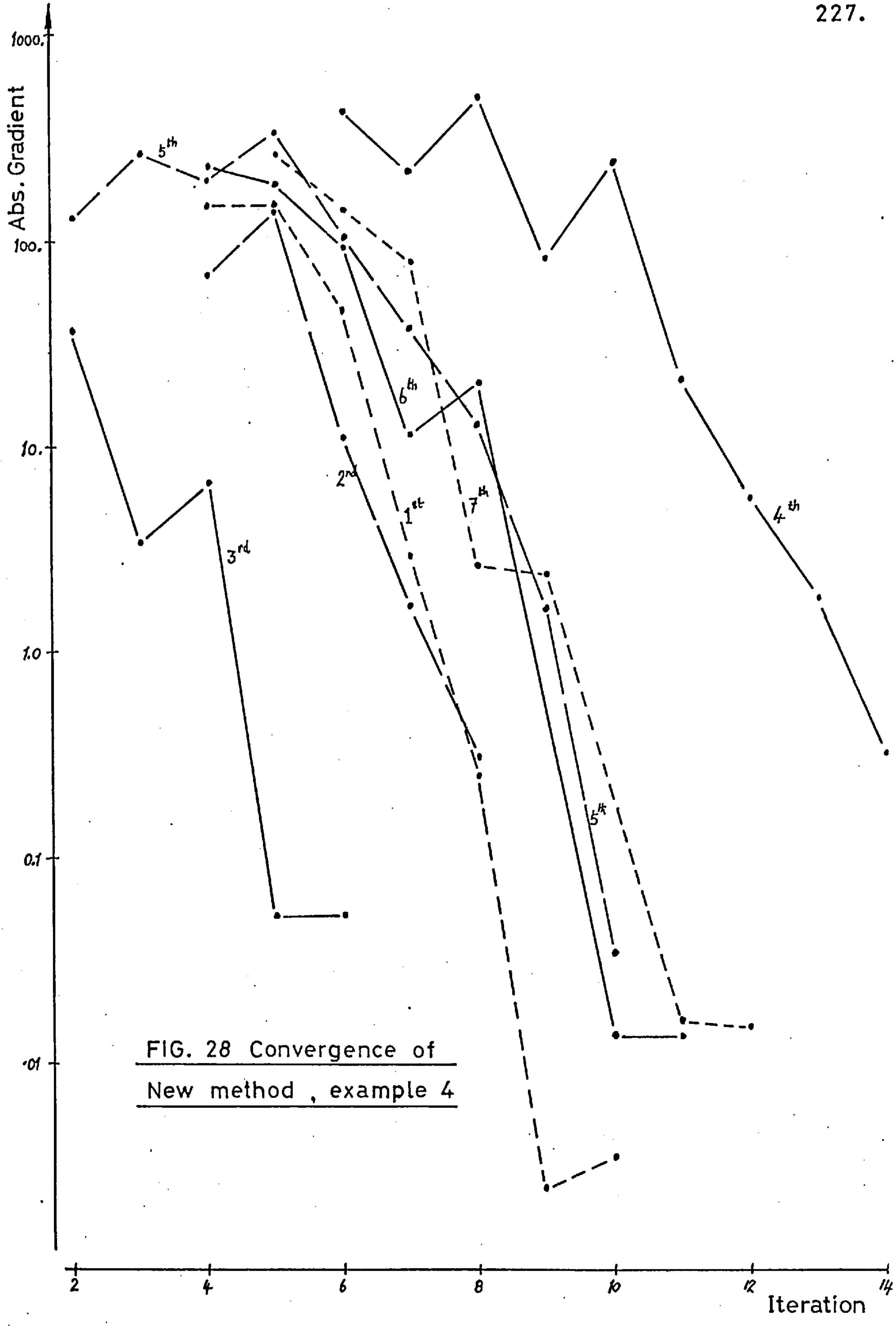


FIG. 28 Convergence of New method , example 4

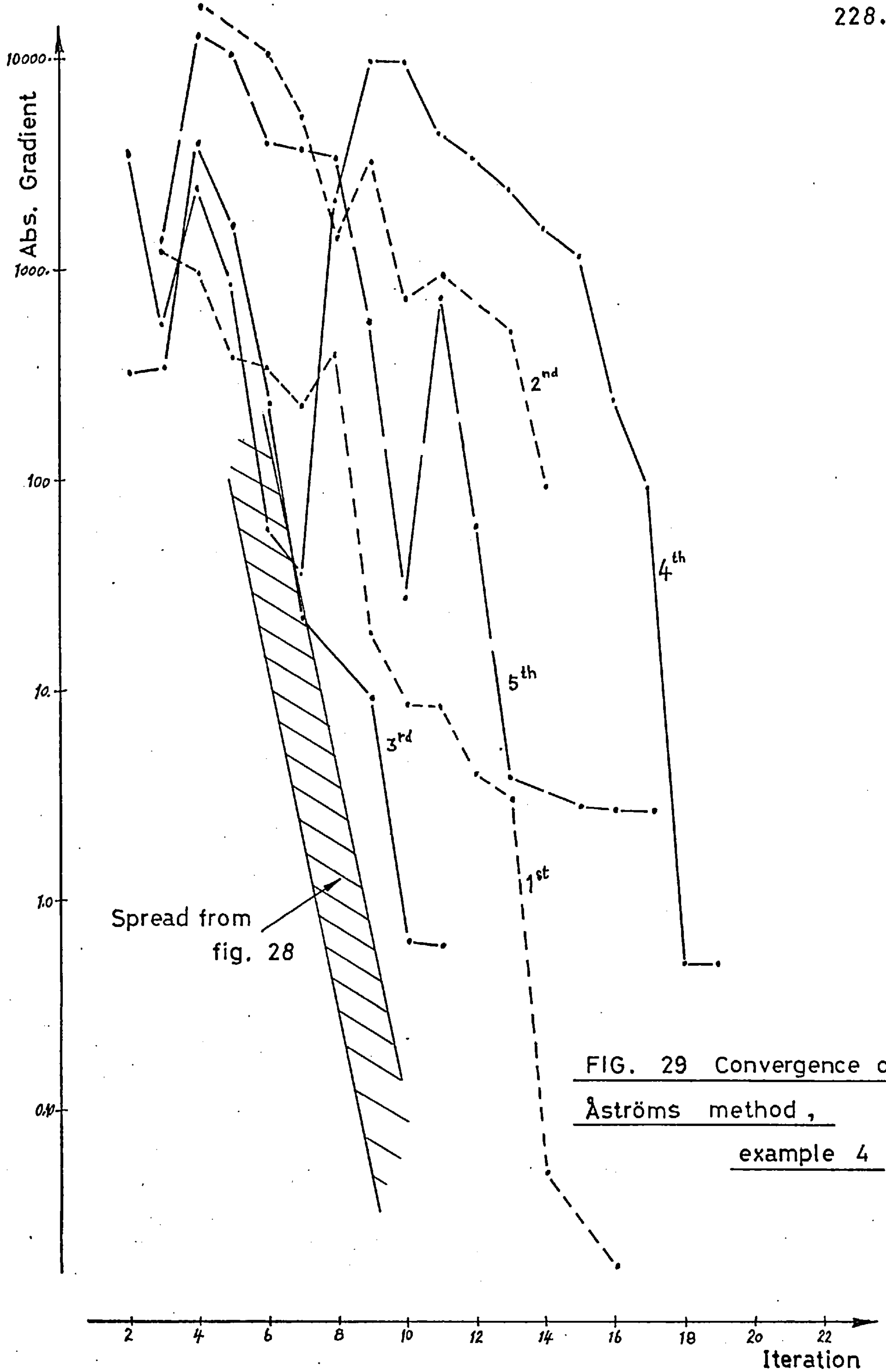


FIG. 29 Convergence of Åströms method, example 4

EXAMPLE 5

For this example the parameters of No. 4 were adopted i.e. those in (6.5). However C(z) was changed to a complex pair of roots (0.9792855 ± j0.200) at a radius of 0.99900. The data length was shortened to 500. Clearly these two items are incompatible in terms of chapter 4, and were used as an extreme test of the abilities of the new method.

During a typical estimation run, the algorithm reduced the cost from an initial 1272.5 to 246.146 in 9 iterations. The gradient at this point was (-319.4, 9.377) in the Z plane. The estimate of the roots γ of C(z) was then (0.979493 ± j0.199385) at a radius of 0.999580. The value of $|\gamma|^{2N}$ was 0.657. As in example 2 the decision was therefore taken to stop climbing radially due to the indicated incompatability. The climbing routine was permitted however to search angularly to see if a better minimum could be obtained. A further iteration gave an estimate of C(z) as (0.979880 ± j0.197472) with a gradient of (-296.955, -66.6461) and a cost of 246.085. This minimum was taken as the best achievable in the circumstances, and it will be noted that the gradient vector is virtually aligned with the radial through the final optimum point.

As a further check, the function was evaluated at the intersection of the unit circle and the above radial, and this gave a value of 245.958. This is significantly smaller than the optimum found and indicated that the true optimum in this case lay outside the unit circle. Åström's method ran into the same non-convergent difficulties with this example as were found for example 2, and described in reference 37.

6.9 Constant Bias terms.

The basic system model equations (1.1) contain a constant bias term γ' on the measurements y_k. Such a term might be due to a direct offset in the measuring instrument. If this is known it can be accounted for by subtracting the bias from the measured values. Usually the control signal u_k would have a constant bias about the measurement datum and as a result y_k would also have a constant bias of a related magnitude. The bias term γ is intended for some unknown value to be estimated and can represent a disparity between the input and output bias values on y_k and u_k.

The value of γ was transformed to appear in (1.38) as γ' and this is related to γ by the definition in (6.6).

$$\gamma' = \gamma * \sum_{i=0}^{n} a_i \qquad (6.6)$$

The maximum likelihood approach of section 2.5 obtains an expression for the residuals $\hat{e}_k$ and then seeks to minimise the estimation cost $V \triangleq \sum_{k=1}^{N} \hat{e}_k^{2}$. Equation (2.49) should be extended as shown in (6.7) to include the constant bias term. The notation for γ has been extended to γ_y to indicate that the bias is considered to be on the y_k signal. Thus γ_e is the bias on e_k etc.

$$\hat{e}_k = \frac{1}{\hat{C}(z)}\left[\hat{A}(z)\; y_k - \hat{B}(z)\; u_k\right] + \gamma_e \qquad (6.7)$$

$$\text{where} \quad \gamma_e = -\sum_{i=0}^{n} a_i \Big/ \sum_{i=1}^{n} c_i * \gamma_y \quad ; \quad \gamma_y \triangleq \gamma'$$

It should be clear that the choice is open to model this term $\mathcal{X}$ as convenient, on the y_k, e_k or u_k signals. The only factors required when transforming from one to the other are the steady state step responses of the A,B and C polynomials. These are given by either the sum of all the coefficients or the products of all the roots of the polynomials.

The full maximum likelihood solution is now given by climbing in all the previous parameters plus $\mathcal{X}_y$. This would be most convenientl done by correcting all the y_k values with a value of $\mathcal{X}_y$ chosen from the climbing routine to give (6.8)

$$\hat{e}_k = \frac{1}{\hat{C}(z)} \; (\hat{A}(z)\, y^{\delta}{}_k - \hat{B}(z)\, u_k) \qquad (6.8)$$

$$\text{where } y^{\delta}{}_k \triangleq y_k - \mathcal{X}_y$$

This method would extend the dimensionality of the space to 3n+2, which was already considered large and was the reason for the least squares modification of section 6.2. We introduced there the simple approach of dealing with $\mathcal{X}_y$ directly by using the means of the data on y_k and u_k to correct them to zero mean signals. If the roots of $\hat{C}(z)$ were very near the unit circle and N were not large, the cost response due to $\mathcal{X}_y$ i.e. a bias on y_k, would be similar to the response of $1/\hat{C}(z$ The simple method is then not ideal because an estimate $\mathcal{X}_y$ introduces components in the full second derivative matrix of section 4.7, and this matrix would tend to be singular.

As an alternative approach it will be seen from (6.7) that the estimation cost V is quadratic in $\mathcal{X}_e$ and therefore the least squares procedure can be extended to give an estimate of χ_e directly for any given $\hat{C}(z)$ value. Thus the dimensionality q of (2.11) and (2.7) is extended by one to 2n+2 and $\underline{m}_k$ has an extra component 1.0.

Example number 6 was studied to verify some of these ideas. The parameters are given in (6.9).

$$y_k = \frac{B(z)}{A(z)} u_k + C(z)\ e_k \quad ; \quad k=1,\ 1000 \tag{6.9}$$

$$\begin{aligned} \text{where } A(z) &= 1.0 + 1.0z^{-1} + 0.29z^{-2} \\ B(z) &= 2.5 - 2.5z^{-1} + 0.725z^{-2} \\ C(z) &= 1.0 + 0.9z^{-1} + 0.8z^{-2} + 0.7z^{-3} + 0.6z^{-4} \\ E(u_k) &= 1.0 \ ; \ E(e_k) = 5.0 \\ E\left[u_k - E(u_k)\right]^2 &= 1.0 \\ E\left[e_k - E(e_k)\right]^2 &= 1.0 \end{aligned}$$

One data run for u_k and e_k was used as before to generate the y_k, u_k signals for various trials. Initially C(z) was set to 1.0 only and thus the system was excited by only white noise from e_k with the bias shown. A least squares solution for $\hat{A}, \hat{B}$ and χ_e enabled the cost $V = \sum_{k=1}^{N} \hat{e}_k^{\,2}$ to be found for $\hat{C}(z) = 1.0$ only. This gave a value of 1001.531. Using the previous simple method of subtracting means first, the estimation cost was then 1015.256. The increase in cost of about 1.5% is significant compared with other sources of error. When the polynomial C(z) adopted the value shown in (6.9), the costs were 2219.397 for the complete least squares method and 2285.662 for the naive method i.e. about 4% worse.

This means that even for this trivial example method the simple method is not satisfactory and cannot be recommended. However if the y_k and u_k signals contain a large constant bias, numerical difficulties can easily arise in the full maximum likelihood approach or in the least squares method. This is principally due to the finite digital work length. Clearly in such a situation the above scheme of extracting the signal means beforehand would have a computational advantage. During the estimation process itself, this ought to be backed up by estimating the remnant constant bias.

Such a method counteracts the defect of not estimating the initial states on the signals at k=1. Difficulties can occur if all y_k, k= -n+1, -n+2, ... ,0 are taken as zero, whereas there is in fact a large D.C. bias on those values. The $1/\hat{C}(z)$ filter is then excited by quite the wrong initial conditions and this has a significant effect on the numerical analysis. The procedure of extracting the means first at least gives approximately correct initial conditions and enables the routines to work without word length troubles.

6.10 Delay in the System.

The full system described by equation (1.1) is not likely to describe a physical system in the sense that the model allows the y_k signal to respond directly to the control signal u_k. Certain coefficients such as b_o and b_1 should not be considered when the plant contains transport or storage times which are significant compared to the sampling period of the discrete time model. The study of this area can become quite extensive; however for present purposes we have

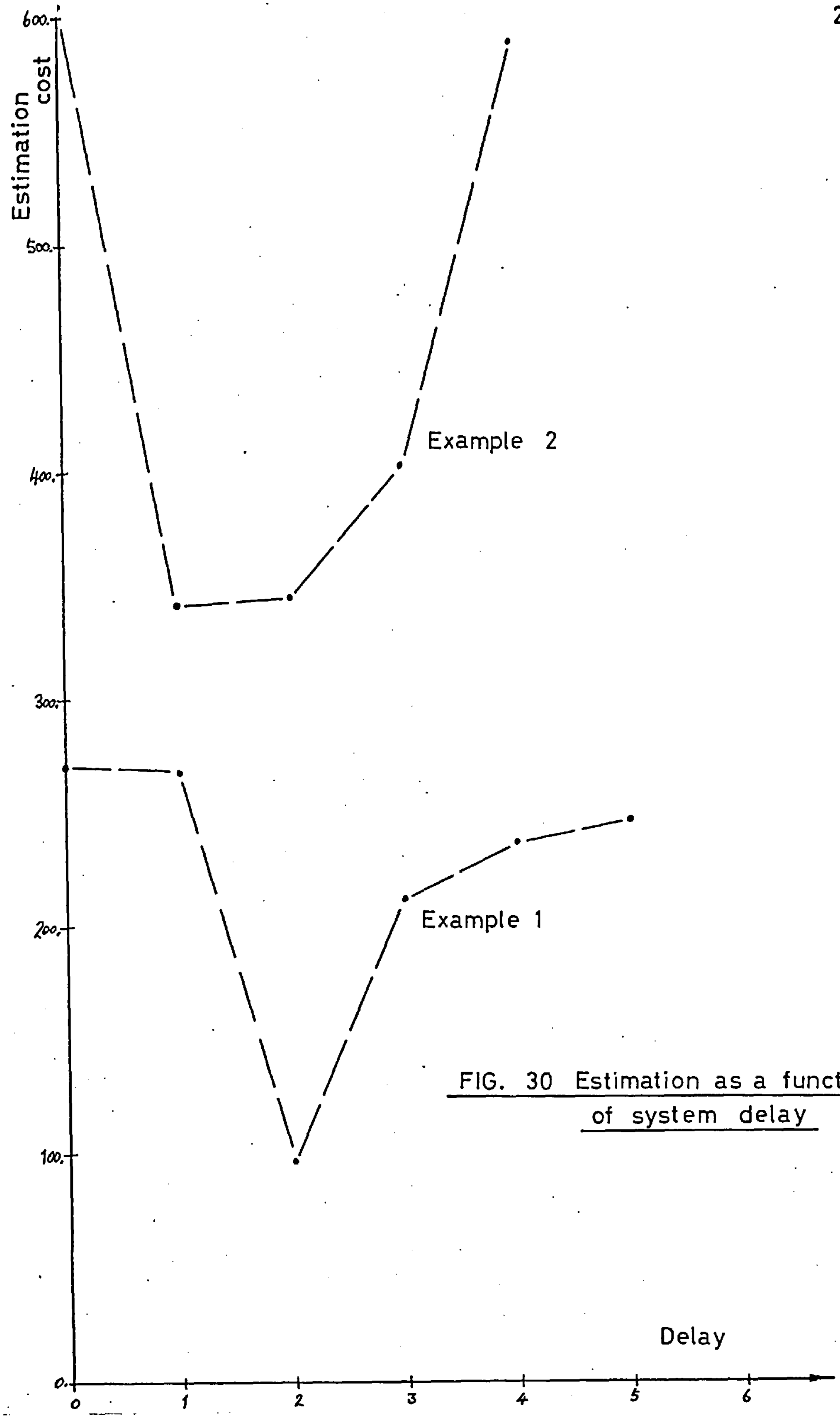


FIG. 30 Estimation as a function of system delay

considered a typical delay of two periods on the u_k signal together with examples 1 and 2.

The estimation runs were repeated for one member of the ensemble for various delays of the u_k signal. The resulting estimation costs are shown in figure 30. The system of example 1 gives a very clear minimum for a delay of two periods, while example 2 shows a flatter minimum centred on the same delay. These results are in line with Clarke's[13] work with a similar problem. The flatter minimum of the 2nd example is probably related to the non-minimum phase nature of the plant. Thus the coefficients of B(z) are 1.0, -3.50, +3.0 which are large and are dissimilar to those for a polynomial whose roots lie inside the unit circle.

Given a plant data record some trial idea of the order n of the assumed plant structure and any delay terms must be formed before the estimation routine is entered. From the results of section 6.5 and table 4, it is plain that other values of n must be tried as well before the true value of the order n* can be comfortably decided. From figure 30, it is also clear that more than one delay time should also be tried. This means that finding the correct value of n and the transport delays in the plant is itself a hill climbing procedure, but at a higher level. The entire estimation process could be considered as a hierarchy of hill climbing schemes. The topmost level would be deciding the possible structure of the plant, the next deciding the order n and delay terms, and the lowest would be the scheme for climbing in tne parameters as described in section 4.6. Naturally the higher levels can only take integer values for their variables. This reduces the number of effective combinations, but also implies that more specialised integer hill climbing routines are required, although it is likely that human beings would always be retained "in the loop" at that level.

CHAPTER 7

CONCLUSIONS AND EXTENSIONS

7.1 Summary of Contributions.

It is convenient to summarise the work of this thesis by a review of each chapter. The underlying structure of the systems studied was introduced in Chapter 1 and principally followed the work of Rowe[16]. A stochastic difference equation was derived from a basic system description in state variable form. Since the derived form is for a single input single output plant, the equations in sections 1.5 to 1.9 are a subset of Rowe's multivariable case. It should be clear that as it is possible to transform any description to any other, within certain conditions, we might as well choose a structure for which it is easiest to estimate the parameters. Åström and Rowe choose the coefficient polynomial description of (1.38), while we advocate here the root form description set out in section 3.7.

Chapter 2 described the properties of different estimation methods from the simplest least squares scheme to methods which made the best use of the available data. Åström's Maximum likelihood approach combines the property of being asymptotically efficient with a simple and elegant computation scheme which reduces to an iterative hill climbing problem. In practise this approach is not far removed from the Generalised Least Squares method of section 2.4, although the underlying philosophy is different. Åström's method also has the advantage of a theorectical proof of its properties at least asymptotically, which is lacking in more heuristic methods such as

Clarke's[13].

Section 2.6 demonstrated that Åström's solution of the estimation problem in fact solved the stochastic regulator problem as well. This demands that the system described by the polynomial $\hat{C}^{-1}(z)$ should have all its poles within the unit circle i.e. a stable system. It is feasible but inconvenient to solve for the roots when given the coefficients of low order polynomials during a hill climbing process. It is clearly more sound to climb in, and describe systems by their roots. This enables the stability criterion to be readily checked.

The method of calculating the response of a discrete time system with a rational polynomial Z transform was given in detail in Chapter 3 principally since this appeared to be lacking in the text books. This applies to both the pulse response and also to calculating the system output signal variance in a stochastic situation. Other authors[29.30,32] have employed the coefficient descriptions and have missed the simple and elegant results which come from the transformation into a root description. These results are not surprising in hindsight as similar root methods have been frequently used for continuous time systems described by rational S polynomials.

The approach from the root viewpoint does not appear to have been exploited before in the area of estimating discrete time systems from data records. Previous approaches have all[4,5,10,11,12,13,15,16,19,26,27,28,37,39,42,43,44] described systems in terms of polynomial coefficients and have on occasion run into convergence difficulties.

The X transformation method introduced in section 3.10 together with the system description in terms of roots allows the estimates to be constrained within the class of stable systems. This removes the most common source of non-convergence. The hill climbing routine with the X transformation operates in an unconstrained space and can therefore be chosen from a class of fairly sophisticated and efficient algorithms. Several authors[1,73,43,44] have shown that real plant disturbances frequently arise from finite random walks or white noise which has been passed through simple low pass filters. The system disturbance in such cases has a high correlation with itself at non zero time shifts. This implies that the estimated polynomial $\hat{C}(z)$ will have roots close to the unit circle and give convergence difficulties. The X transformation approach is most beneficial in this area. Not only are the roots constrained as required but the non-linearity of the Tanh transformation appears to match the sensitivity of cost to the root motion. Therefore as far as the unconstrained hill climbing algorithm is concerned the hill is quite regular and free from difficult regions.

Section 3.14 demonstrates more clearly than Åström[65], by using the root description, the significance of non-minimum phase systems. These can be regarded as having zeros outside the Z plane unit circle, and yet matched by internal reciprocal poles to give an all pass whitening effect. Thus non-minimum phase systems can be seen to be equally easy to estimate, although they will limit the performance of control systems[31]. Figure 17 demonstrated that such estimates will be strictly biassed for a finite data length and this can be seen by differentiating (3.41) under the above reciprocal pole condition. Section 3.14 also showed that a continuous time all-pass system loses that property on sampling, but the property does appear in discrete time for an inverse pole-zero relationship in the Z plane.

Sections 4.1 to 4.7 demonstrated that the first and second derivatives of the estimation cost in terms of a root description can be obtained in a way which is equally simple to Åström's shifting method[10,11] used for coefficient description systems. A trivially simple filter is required for each root and this can be implemented with an equivalent ammount of work to that of the shifting method. Further simplifications also arise for complex conjugate root pairs. Derivatives in the transformation X space can be obtained almost directly from the derivatives with respect to the roots, and can be used for a hill climbing routine[9] which requires gradient information.

The rest of Chapter 4 is devoted to studying the second derivative matrix of the estimation cost. This matrix enables statements to be made about the probable precision of the estimate which has been obtained and has been treated before by Åström[12]. However for several experiments the practical second derivative matrix calculated from a finite data set did not match the theoretical matrix for the same data length. This effect also appeared to be related to the speed of convergence of the estimate and the occurence of singularities. Section 4.10 gave an expression for the variance of the matrix elements.

The calculation method for such variances was given in sections 4.10 and 4.11 and shows that for a given quality of the second derivative matrix, the data length N and the pole 'strength' are related. Thus criteria can be developed as in section 4.12 to enable judgements to be made, either during the estimation procedure or when planning an experiment, about the length of data required in given circumstances in order to avoid convergence and other difficulties. For example the estimation procedure can be stopped by a simple test, such as (4.87) for want of a longer data length N.

Chapter 5 draws on Åström's work in order to prove consistency and efficiency when estimating in terms of a root structure. Lemma 1 is shown to hold since by means of the X transformation the stability of the A and C polynomials can be guaranteed and the estimate $\hat{\underline{\theta}}$ can only belong to the region $\mathcal{R}$ of stable systems.

Theorem 1 requires for a proof of consistency that the climbing routine finds the global maximum, rather than a local maximum of the likelihood function. Sophisticated climbing routines are of benefit here as in practise they are more capable of dealing with non-convex regions. The likelihood surface cannot be shown to be convex for the same reason that occurs in Åström's work, and hence a solid analytical backing is not available for global convergence statements. The reason lies in certain cross-correlation terms in the full expression for the second derivative matrix described in sections 5.10 to 5.12. These terms prevent the matrix from being proven to be positive definite except at the optimum itself.

Theorem 2 of Åström, described in section 5.6, applies to the root description approach with a proviso about the permutation of the roots. Such permutations can be shown to be of no consequence to the final system and such estimates are therefore consistent. Again similar to Åström, estimates of the system initial conditions are strictly inconsistent and have been ignored in later examples by choosing suitable data lengths. The identifiability theorem 3 is satisfied by the controllability assumptions of Chapter 1 about the original system, but also requires the control signal to be persistantly exciting as defined in section 5.7. All the previous lemmas and theorems are invoked in theorem 4 to show that the maximum likelihood estimate is asymptotically normally distributed as the data length tends to infinity. This leads to the estimates being shown to be asymptotically efficient.

Sections 6.1 to 6.4 described the computer program which was written to implement many of the ideas described in this thesis. Clearly it is possible to climb in the full root description of the A,B and C polynomials. However for a working engineering method, advantage was taken of opportunities to reduce the dimensionality of the space to only n instead of 4n+2, and also to reduce the work required for a cost evaluation. This means a loss of academic nicety, but a considerable gain of practical worth in the final program. The checks suggested in Chapter 4 were included, as this enables the estimation procedure to halt when the model and data length are incompatible.

The new method gave very similar results as Åström's program, with the standard example 1. This is a natural requirement of any new estimation method. Example 2 was chosen as a difficult problem likely to give convergence difficulties. The differences between the new method and Åström's now begin to be apparent. This is shown further in the later examples which were chosen to illustrate different features.

7.2 Extensions.

Various possibilities come to mind when considering the practical estimation of parameters from field data records. Section 6.9 briefly covered the treatment of signals which contain a constant bias. Such signals and ill defined transducer noise are common in practise and arise from both the variable measured and the instrumentation available. The bias term may well drift with time or otherwise limit the available measuring precision. These details can only be properly resolved with experience of field work, although hypothetical analysis can be used for background information. Thus the judgement of whether a sophisticated estimation method is good or not depends on practical experience as well as inherent computational advantages.

In a similar way, actual delays in the plant due to transport or storage are not necessarily amenable to the method described in section 6.10. There we described a hierachy of climbing operations involving decisions about structure, order and parameter estimation at successively lower levels. Such schemes could evolve to be very complicated and include pattern recognition algorithms to aid the decision processes. However it is quite likely that a human being would be retained 'in the loop' in all but a few fast time varying systems.

It is possible that the method described here might be extended, as in Åström's case[11], to an on-line adaptive situation. Strictly, this is not derived from the Maximum Likelihood method which is aligned towards a hill climbing procedure using all the available data. However given a value of $\hat{C}(z)$, a recursive form of the least squares method[19] can be used as in section 2.2. Each new data point would be filtered as in (2.37) and used to update the estimate using the computation of (2.25) and (2.26). The full Maximum Likelihood method could conceivably be treated by recursive updating in this manner if the hill in the $\hat{C}$ parameters could be shown to be suitable regular under certain conditions. For the present theory to hold, any recursive approach must be proven to be equivalent at each stage to the full maximisation over the total data set.

The state variable description assumed for generality in (1.1) may in certain practical cases prove too pendantic. It might be known for example that a system of (7.1) was nearer a true description of the actual system. This might be a fast dynamic system described by the A and B polynomials and a dynamically much slower disturbance C,D added to the system output.

$$y_k = \frac{B(z)}{A(z)} u_k + \frac{C(z)}{D(z)} e_k \qquad (7.1)$$

Although the standard description could be obtained by absorbing the D polynomial in A,B,C set, it might be prefered to rework the equations of section 4.6 and apply the method to the new structure. In such a case it would be wise to check that such a set of parameters could be estimated consistently in the sense of theorem 2. Thus the controlability conditions must still apply and this implies that pole-zero cancellation is not permissable.

It seems very likely that the work described here can be extended to the multivariate input-output case. As suggested by Åström[11] and developed by Rowe[16], a suitable canonical form is required which is a minimal representation in order to avoid the singularity effects described by Kalman[25]. The 'A' canonical form of Rowe is given here in (7.2). As mentioned in section 1.7, the matrix polynomial A has a number of possible zero elements for a minimal representation. The notation $(I_m)_j$ is used to donate that certain rows of a unit matrix of order m have been deleted during the transformation from the state description given in section 1.4 and 1.5.

$$\underline{y}_k + A_1\underline{y}_{k-1} + A_2(I_m)_2\underline{y}_{k-2} \cdots\cdots + A_p(I_m)_p\underline{y}_{k-p} = B_o\underline{u}_k + B_1\underline{u}_{k-1} \cdots\cdots$$

$$+ B_p\underline{u}_{k-p} + \Lambda^{\frac{1}{2}}\underline{e}_k + C_1\Lambda^{\frac{1}{2}}\underline{e}_{k-1} \cdots\cdots C_p\Lambda^{\frac{1}{2}}\underline{e}_{k-p} \tag{7.2}$$

A working estimation procedure can be constructed as a multivariate form of the program used for the examples in Chapter 6. Given some value of the $\hat{C}(z)$ matrix polynomial the vector signals $\underline{y}_k$ and $\underline{u}_k$ may be filtered by $\hat{C}(z)^{-1}$ as in section 2.4 to give $\underline{y}^*_k$ and $\underline{u}^*_k$. A multivariate form of the least squares algorithm[16] can now be used to estimate the A and B matrix polynomials. The procedure is iterated as before towards convergence using a hill climbing technique. As before the stability of the $\hat{C}(z)^{-1}$ system is important and its effects cannot be ignored.

Rowe retained the coefficient description i.e. the companion form of the A,B and C matrix polynomials,for the multivariate system. Computational difficulties are known to have arisen for the case when the eigenvalues of $\hat{C}$ were close to the unit circle. This is simply the multivariate form of the effects seen for scalar polynomials in this thesis. Clearly a more rational approach is to express the matrix polynomial $\hat{C}$ as a Jordan form in which the eigenvalues are displayed explicitly. A similar treatment should also be applied to the $\hat{A}$ matrix polynomial in view of the conditions required by the theorems in Chapter 5. These explicit eigenvalues can then be treated with the X transformation method as before to obtain similar advantages to the scalar version.

Consider the matrix polynomial $\hat{C}(z)$ to be of the form (7.3)

$$\hat{C}(z) = (I_m + \hat{C}_1 Z^{-1} + \hat{C}_2 Z^{-2} \,.....\, \hat{C}_p Z^{-p}) \qquad (7.3)$$

where I_m is an m*m unit matrix, and $\hat{C}_i$ are coefficient matricies.

The index p is the controllability index as discussed in Chapter 1 for minimal representation forms. If now $\hat{C}(z)$ is expressed as in the form of (7.4), there are still m^2p parameters, but each "root block" can be individually treated.

$$\hat{C}(z) = (I_m + \hat{C}_{r1}Z^{-1})(I_m+\hat{C}_{r2}Z^{-2}) \; \ldots\ldots \; (I_m+\hat{C}_{rp}Z^{-1}) \qquad (7.4)$$

Each matrix $\hat{C}_{ri}$ is m*m, and from these the coefficient matricies $\hat{C}_i$ of (7.3) could be calculated. Thus as in section 3.7, we may as well choose a form which has advantages for estimation purposes if alternative forms can always be derived later.

Suppose we are concerned with obtaining an m vector output signal $\underline{\hat{e}}_k$ from an input $\underline{v}_k$ to a filter $\hat{C}(z)^{-1}$ as in (7.5).

$$\underline{\hat{e}}_k = \hat{C}(z)^{-1}\underline{v}_k = (I_m+\hat{C}_{rp}Z^{-1})^{-1} \ldots\ldots \; (I_m+\hat{C}_{r2}z^{-1})^{-1}(I_m+\hat{C}_{r1}z^{-1})^{-1}\underline{v}_k \qquad (7.5)$$

It will be noted that the multivariate form does not commute and the block order is reversed on inversion. If the intermediate m vector signal $\underline{w}_k$ defined by (7.6) is introduced, then we can study the stability of the elemental form of (7.7) in isolation.

$$\underline{w}_k = (I_m + \hat{C}_{r(p-1)}z^{-1})^{-1} \; \ldots\ldots \; (I_m + \hat{C}_{r1}z^{-1})^{-1}\underline{v}_k \qquad (7.6)$$

$$\underline{\hat{e}}_k = (I_m + \hat{C}_{rp}z^{-1})^{-1} \; \underline{w}_k \qquad (7.7)$$

As shown in Chapter 1, an m*m non singular transformation matrix T can be chosen in (7.6) so that the matrix F_p is diagonal and explicitly showing the m eigenvalues of $\hat{C}_{rp}$.

$$\underline{\hat{e}}_k = T^{-1} \, (I_m + F_p z^{-1})^{-1} \, T \, \underline{w}_k \qquad (7.8)$$

The matrix T^{-1} is made up of m column eigenvectors each of arbitrary norm. Thus T^{-1} contains (m^2-m) non-arbitrary parameters and F_p contains m eigenvalues. This the same total of m^2 parameters as in $\hat{C}_{rp}$. The procedure can be repeated for each block of (7.4) to give the total description of (7.9)

$$\underline{\hat{e}}_k=T_p^{-1}(I_m+F_pz^{-1})^{-1}T_pT_{p-1}^{-1}(I_m+F_{p-1}z^{-1})^{-1}T_{p-1}\ \ldots\ T_1^{-1}(I_m+F_1z^{-1})^1T_1\underline{v}_k \tag{7.9}$$

To ensure stability of such a system, each of the m elements of each F_i matrix, i=1, ... p, should have a magnitude of less than 1.0, and are therefore candidates for the X transformation process of Chapter 3. For convenience we can set the m values on the main diagonal of each T_i to be unity, this leaves a total of $p(m^2-m)$ parameters to be specified by an unconstrained hill climbing process.

The above procedure has reduced the $\hat{C}(z)$ description of (7.3) to a form (7.9) which is more useful for estimation in the sense that the stability of $\hat{C}(z)^{-1}$ is ensured as was done in the scalar case, and any other form may be derived if required. Jordan forms of F, in which there are more than m non-zero elements, are not treated so easily, but have been assumed to be relatively rare. Complex conjugate eigenvalues or root blocks would probably simplify as in section 4.1 to give amenable forms.

The actual filter of (7.8) is as easy to implement as that in (4.6) since it reduces to the structure of (7.10), which is m simple decoupled scalar filters.

$$(I_m+F_iz^{-1})^{-1} = \begin{bmatrix} 1+f_1z^{-1} & & 0. \\ & 1+f_2z^{-1} & \\ & & \ddots \\ 0. & & & 1+f_mz^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{(1+f_1z^{-1})} & & 0. \\ & \frac{1}{(1+f_2z^{-1})} & \\ 0. & & \ddots & \frac{1}{(1+f_mz^{-1})} \end{bmatrix} \tag{7.10}$$

As in the scalar case, the derivatives $\frac{\partial \hat{\underline{e}}_k}{\partial q}$, where q is any parameter, are not difficult to calculate so that sophisticated hill climbing routines can be employed. Thus the multivariate problem, although complicated can be regarded and programmed as a set of scalar filter stages similar to the scalar polynomial systems of this thesis.

APPENDIX 1

The variance of sample variance of a 'white noise' sequence with a normal distribution.

Define the sample variance $\emptyset^N$ as :

$$\emptyset_o^N \triangleq \frac{1}{N} \sum_{k=1}^{N} \epsilon_k^2 \qquad (A.1)$$

where ϵ_k, k=1 ..... N is a sample of an independant random source with known zero mean, variance σ_e^2 and normal distribution. i.e. the 4th. moment, $E(\epsilon_k^4) = 3\sigma_e^4$

The variance of $\emptyset_o^N$ is given by its 2nd. central moment.

$$\text{var.}\ (\emptyset_o^N) = E\left[\emptyset_o^N - E(\emptyset_o^N)\right]^2 \qquad (A.2)$$

Now from (A.1) $E(\emptyset_o^N)$ is given by σ_e^2 then

$$\text{var.}\ (\emptyset_o^N) = E\left[(\emptyset_o^N)^2 + E(\emptyset_o^N)^2 - 2\emptyset_o^N * E(\emptyset_o^N)\right] \qquad (A.3)$$

$$= E\left[\frac{1}{N^2}\left\{\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\epsilon_i^2\epsilon_j^2 + \sum_{i=1}^{N}\epsilon_i^4\right\} + \sigma_\epsilon^4 - \frac{2}{N}\sum_{i=1}^{N}\epsilon_i^2 * \sigma_e^2\right] \qquad (A.4)$$

$$= \frac{1}{N^2}\left\{\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}E(\epsilon_i^2\epsilon_j^2) + \sum_{i=1}^{N}E(\epsilon_i^4)\right\} + \sigma_e^4 - \frac{2}{N}\sum_{i=1}^{N}E(\epsilon_i^2) * \sigma_e^2 \qquad (A.5)$$

Due to independance $E(\epsilon_i^2\,\epsilon_j^2) = E(\epsilon_i^2) * E(\epsilon_j^2)$ for $i \neq j$

$$\text{var.}(\phi_o^N) = \frac{N(N-1)}{N^2} \cdot \sigma_e^2 * \sigma_e^2 + \frac{N}{N^2} \cdot 3\sigma_e^4 + \sigma_e^4 - \frac{2N}{N}\sigma_e^2 * \sigma_e^2$$

$$\therefore \text{var.}(\phi_o^N) = \sigma_e^4 - \frac{\sigma_e^4}{N} + \frac{3\sigma_e^4}{N} + \sigma_e^4 - 2\sigma_e^4 = \frac{2\sigma_e^4}{N} \qquad (A.6)$$

This result can also be derived by considering ϕ_o^N as a sum of squares as in (A.1), and has a chi-square distribution with N degrees of freedom.

$$\frac{1}{\Gamma(\frac{N}{2})} \left(\frac{\chi^2}{2}\right)^{N/2 - 1} \exp \frac{(-\chi^2)}{2} \cdot \frac{\chi^2}{2} \qquad (A.7)$$

It is straight forward to show from this that ϕ_o^N has a mean and variance given by

$$\frac{1}{N}\, E(\chi^2) = \sigma_e^2 \qquad (A.8)$$

$$\frac{1}{N}\, \text{var.}(\chi^2) = \frac{2\sigma_e^2}{N}$$

For values of $N \geqslant 30$, the Chi-square distribution is closely approximated by a normal distribution with the same mean and variance.

<u>Variance of a sample autocorrelation.</u>

This can be derived in a similar way to above. Define a sample autocorrelation ϕ_1^N of shift 1 as

$$\phi_1^N = \frac{1}{N'} \sum_{i=1}^{N'} \epsilon_i\, \epsilon_{i+1} \qquad (A.9)$$

where N' = N-1 and is the maximum number of terms to be summed for a shift of 1 and a data length N

The variance of ϕ_l^N is given by (A.10)

$$\text{var.}(\phi_l^N) = E\left[\frac{1}{N'}\sum_{i=1}^{N'} \epsilon_i\epsilon_{i+l} - E(\phi_l^N)\right]^2 \tag{A.10}$$

Now $E(\phi_l^N)$ is zero due to the independance of $\epsilon_i, \epsilon_{i+l}$ for $l\neq 0$.

$$\text{var.}(\phi_l^N) = \frac{1}{N'^2}\sum_{i=1}^{N'}\sum_{j=1}^{N'} E(\epsilon_i\epsilon_{i+l}\epsilon_j\epsilon_{j+l}) \quad \text{for} \quad l\neq 0 \tag{A.11}$$

$$= \frac{1}{N'^2}\sum_{j=1}^{N'}\left\{\sum_{\substack{i=1\\ i\neq j}}^{N'}\left[E(\epsilon_i)*E(\epsilon_{i+l})*E(\epsilon_j)*E(\epsilon_{j+l})\right] + E(\epsilon_i^2\epsilon_{j+l}^2)\right\} \tag{A.12}$$

There will be terms in (A.12) such as $E(\epsilon_j^{\,2})*E(\epsilon_i)*E(\epsilon_{j+l})$ etc. but these with most of the others will be zero, since $E(\epsilon_k) = 0.0$

$$\therefore \text{var.}(\phi_l^N) = \frac{1}{N'^2}\sum_{j=1}^{N'} E(\epsilon_j^{\,2}\ \epsilon_{j+l}^2)$$

$$= \frac{1}{N'^2}\sum_{j=1}^{N'} E(\epsilon_j^{\,2})*E(\epsilon_{j+l}^2) \quad \text{since } l\neq 0 \tag{A.13}$$

$$\therefore \text{var.}(\phi_l^N) = \frac{N'}{N'^2} * \sigma_e^{\,2} * \sigma_e^{\,2} = \frac{\sigma_e^4}{N'} \tag{A.14}$$

The expected sample autocorrelation ϕ_l^N has therefore zero mean, and variance of σ_e^4/N'

We could now repeat the process between (A.9) and (A.14) to obtain the covariance between $\phi_{l_1}^N$ and $\phi_{l_2}^N$, the sample autocorrelations for different delays l_1 and l_2. Equation (A.13) would then contain the term $E(\epsilon_j^{\ 2}) * E(\epsilon_{j+l_2} \cdot \epsilon_{j+l_1})$ which would be zero for $l_1 \neq l_2$. Hence we would expect $\emptyset^N_{l_1}$ and $\emptyset^N_{l_2}$ to be independant and have a zero covariance for $l_1 \neq l_2$ (A.15)

ACKNOWLEDGEMENTS

I would like to acknowledge the helpful supervision by Dr. D. Q. Mayne of the research given in this thesis. The encouragement of many colleagu es in technical discussions is recognised with gratitude. Considerable labour and assistance has been freely given by my wife during the work and the typing of this thesis.

The work was done within the Centre of Computing and Automation at Imperial College, which was under the direction of Professor J. H.. Westcott,and the computing facilities used were provided under a generous contract by I.B.M. The research was financially supported by an Electricity Supply Scholarship awarded by a committee of the National Joint Advisory Council of the Electricity Supply Industry.

The support and contributions of all the above people are gratefully appreciated. Without their help this work would never have been accomplished.

REFERENCES

1. F. MORAN, C.S. BERGER, and D. XIROKOSTAS
"Development and application of self optimising control to coal fired steam generating plant."
PROC. IEE Vol.115 No.2 February 1968. p.307-317

2. R.E. KALMAN
"A new approach to Linear Filtering and prediction problems."
ASME Jnl. of Basic Engineering March 1960 p.35-45

3. R.E. KALMAN and R.S. BUCY
"New Results in Linear Filtering and Prediction theory."
ASME Jnl. of Basic Engineering March 1961 p.95-107

4. K. STEIGLITZ
"Power Spectrum Identification for Adaptive Systems."
IEEE Trans. App. and Industry Vol.83 No.72 May 1964 p.195-197

5. S.A. TRETTER and K. STEIGLITZ
"Power Spectrum Identification in terms of Rational Models."
IEEE Trans. Auto Control Vol.AC - 12 April 1967 p.185-188

6. R. FLETCHER
"Function Minimisation without evaluating derivatives - a review."
Computer Jnl. Vol.8 April 1965 p.33-36

7. H.H. ROSENBROCK
"An automatic method for finding the Greatest or Least value of a function."
Computer Jnl. Vol.3 p.175-184

8. M.J.D. POWELL
"An efficient method for finding the minimum of a function of several variables without calculating derivatives."
Computer Jnl. Vol.7 p.155-162

9. R. FLETCHER and M.J.D. POWELL
"A rapidly convergent descent method for minimisation."
Computer Jnl. Vol.6 p.163-168, 1963.

10. K.J. ÅSTRÖM
"Control Problems in Papermaking."
IBM Scientific Computing Symposium - Control theory and Application
Yorktown Heights NY. October 1964 p.135-170

11. K.J. ÅSTRÖM
"Numerical Identification of Linear Dynamical Systems from normal operating records."
IFAC Symposium on the Theory of Self Adaptive Control Systems, Teddington 1965

12. K.J. ÅSTRÖM
"On the achievable accuracy in identification problems."
IFAC Symposium on Identification in Automatic Control Systems
Paper 1.8 Prague, June 1967

13. D.W. CLARKE
"Generalised Least Squares estimation of the parameters of a dynamic model."
IFAC Symposium on Identification in Automatic Control Systems
Paper 3.17 Prague, June 1967

14. L. SHAW and G. ROBINSON
"Invariant estimation of stochastic system parameters."
IFAC Symposium on Identification in Automatic Control Systems
Paper 3.16 Prague, June 1967

15. D.Q. MAYNE
"Parameter Estimation."
Automatica Vol.3 p.245-255, 1966.

16. I.H. ROWE
"Statistical methods for the identification and control of multivariate stochastic systems."
PhD. Thesis London University November 1967

17. D.R. COX and H.D. MILLER
"The Theory of Stochastic processes."
Methuen 1965.

18. T.W. ANDERSON
"Introduction to multivariate statistical Analysis."
John Wiley and Sons, Inc. 1958.

19. R.C.K. LEE
"Optimal Estimation, Identification, and Control."
Research Monograph No.28 M.I.T. Press.

20. J.T. TOU
"Modern Control Theory."
McGraw-Hill 1964

21. R. DEUTSCH
"Estimation Theory."
Prentice-Hall, Inc. 1965

22. A. PAPOULIS
"Probability, Random variables and stochastic Processes."
McGraw-Hill 1965

23. A.M. MOOD and F.A. GRAYBILL
"Introduction to the Theory of statistics."
McGraw-Hill 1963

24. S.S. WILKS
"Mathemathematical statistics."
John Wiley and Sons. 1962

25. R.E. KALMAN
"On the structural properties of linear, constant, multivariable Systems."
Paper 6A Third Congress of IFAC, LONDON 1966

26. D.Q. MAYNE
"A method for estimating discrete time transfer functions."
Paper C2 Second UKAC Control Convention, Bristol April 1967

27. S.G. TEAFESTAS
"Estimation of parameters of multi-input and multi-output linear dynamic systems."
Report, Dept. of Electrical Engineering, Imperial College, London University September 1966

28. D.Q. MAYNE
"Estimation and Control of Stochastic Systems."
PhD. Thesis University of London 1967

29. E.I. JURY
"Theory and application of the z transform method."
Wiley 1964

"Sampled data control systems."
Wiley; Chapman and Hall, 1958

30. D.P. LINDORFF
"Theory of sampled Data control systems."
John Wiley 1965

31. G.C. NEWTON, L.A. GOULD, and J.F. KAISER
"Analytical design of Linear Feedback Controls."
John Wiley and Sons Inc. 1957

32. P. WHITTLE
"Prediction and Regulation by Linear least squares methods."
E.U.P. 1963

33. C.J. GREAVES and J.A. CADZOW
"The optimal discrete filter corrosponding to a given analog filte
IEEE TRANS. AUTO CONTROL June 1967. p.304-307

34. J.C. JAEGER
"An introduction to the Laplace transformation with engineering applications."
Metheuen Monograph, 1961

35. M.G. KENDALL and A. STUART
"The advanced theory of statistics."
Griffin, London 1961

36. P. EYKHØFF
"Process parameter and state estimation."
IFAC Symposium on Identification in Automatic Control Systems
Paper 2 Prague, June 1967

37. K.J. ÅSTRÖM, T. BOHLIN, and S. WENSMARK
"Automatic construction of linear stochastic dynamic models for stationary industrial processes with random disturbances using operating records."
IBM NORDIC LABORATORY technical paper TP 18.150 June 1, 1965

38. D.Q. MAYNE
"Computational procedure for the minimal realisation of transfer function matricies."
PROC. IEE Vol.115 No. 9 September 1968

39. K.R. GODFREY
"Dynamic analysis of an oil-refinery unit under normal operating conditions."
PROC. IEE Vol. 116 No. 5 May 1969 p.879-888

40. A.G. DEWEY
"A computational procedure for the minimal realisation of transfer function matricies."
Report CCA, Imperial College, London University May 1967

41. F.T. SMITH
"Advances in Control Systems."
Ed. C.T. LEONDES. Vol.2 Academic Press 1965

42. D.W. CLARKE
"Generalised Least Aquares estimation of the parameters of a dynamic model."
Report AUTO 26 National Physical Laboratory, Autonomics Division
November 1966

43. G.E.P. BOX and G.M. JENKINS
"Models for prediction and control; II linear stationary models."
Technical Report No.2, Dept. of Systems Engineering,
University of Lancaster.

44. G.E.P. BOX and G.M. JENKINS
"Recent advances in forecasting and control."
Technical report No.5, Dept. of Systems Engineering,
University of Lancaster.

45. D.W. NORRIS
"Investigations into practical parameter estimation problems."
Centre for Computing and Automation Research Report,
Imperial College, London University August 1967

46. V. PETERKA and P. VIDINCEV
"Rational-Fraction Approximation of Transfer Functions."
IFAC Symposium on Identification in Automatic Control Systems
Paper 2.1 Prague, June 1967

47. J. NEKOLNY and J. BENES
"Simultaneous Control of Stability and Quality of Adjustment."
Proceedings of the First IFAC Congress Moscow 1960

48. J.L. DOOB
"Stochastic processes."
Wiley New York 1935

49. E.T. COPSON
"An introduction to the theory of Functions of a Complex Variable."
Oxford University Press, London 1935

50. R.E. KALMAN
"On the general theory of control."
Proceedings of the First Congress of IFAC, Moscow 1960

51. D.G. LUENBERGER
"Canonical forms for linear multivariable systems."
IEEE Trans. on Auto Control Vol. AC-12 1967 p.290-293

52. F.R. GANTMACHER
"The Theory of Matricies."
Chelsea Publishing Co., New York 1959

53. M. ATHANS and P.L. FALB
"Optimal Control; an introduction to the theory and its applications."
McGraw-Hill 1966

54. B.L. HO and R.E. KALMAN
"Spectral factorization using the Riccati-equation."
4th Allerton Conference, Illinois. October, 1966

55. D.C. YOULA
"On the factorization of rational matricies."
IEEE Trans. on Information Theory Vol. It-7, 1961

56. M.C. DAVIES
"Factorizing the spectral matrix."
IEEE Trans. on Automatic control Vol.AC-8, 1963

57. B.D.O. ANDERSON
"Development and application of a system theory criterion for rational positive real matricies."
4th Allerton Conference, Illinois. October, 1966

58. B.D.O. ANDERSON
"An algebraic solution to the spectral factorization problem."
IEEE Trans. on Automatic control Vol.AC-12. 1967

59. V.S. HUZURBAZAR
"The likelihood equation, consistency and the maxima of the likelihood function."
Annals of Eugenics Vol.14 1947-1949 p.185-200

60. R.A. RUCKER
"Real time system identification in the presence of noise."
IEEE Wescon. Convention August, 1963

61. V.S. LEVADI
"Parameter estimation of linear systems in the presence of noise."
Proc. International Conference on Microwaves, Circuit theory and Information Theory TOKYO September, 1964

62. U. GRENANDER and M. ROSENBLATT
"Statistical Analysis of Stationary Time series."
John Wiley and Sons. New York. 1957

63. J. DURBIN
"Estimation of parameters in time series regression models."
Jnl. Royal Stat. Soc., B. Vol.22 1960 p.139-153

64. D.A.S. FRASER
"Statistics, an Introduction."
John Wiley and Sons. New York. 1958

65. K.J. ÅSTRÖM
"Computer Control of a paper machine; application of linear stochastic control theory."
Paper.

66. E.G. PHILLIPS
"Functions of a complex variable."
Oliver and Boyd 1963

67. B.J. WILLIAMS
"The computer control of a distillation column."
Report COM. SCI. 37 July, 1968 National Physical Laboratory.

68. M.J. BOX
"A comparison of several current optimisation methods and the use of transformations in constrained problems."
Computer Jnl. Vol.9 No.1 May, 1966

69. H. CRAMÉR
"Mathematical methods of statistics."
Princeton University Press, Princeton 1946

70. H.R. SIMPSON
"A sampled data nonlinear filter."
PROC. IEE, Vol.112 No.6 June, 1965 p.1187-1196

71. W. FELLER
"An introduction to Probability theory and its applications."
Vol.1 John Wiley, New York 1957

72. PROGRAM 'GAUS'
SHARE program library

73. ROBERTS
IFAC , STOCKHOLM