DIGITAL ADAPTIVE POLE SHIFTING REGULATORS

A thesis presented to

THE VICTORIA UNIVERSITY OF MANCHESTER

for the degree of

DOCTOR OF PHILOSOPHY

by

John Meirion Edmunds

Control Systems Centre
The University of Manchester
Institute of Science and Technology

July 1976

TIGHTLY
BOUND
COPY

ABSTRACT

This thesis is a study of a self-tuning regulator which has been formed by combining a recursive least squares estimator with a general pole shifting control law. The least squares estimator is used to estimate the system parameters at the same time as the control law is used to control the system. The combination will give stable control for most linear systems.

The stability and conditional stability which may occur with this regulator are investigated with the aid of root locus diagrams.

Two ways of introducing setpoint variations and so turning the regulator into a controller are suggested. The initial properties of the self-tuning regulator are investigated with simple illustrations. It is shown that the bias in the parameter estimates, caused by certain types of coloured noise on the system, does not effect the asymptotic behaviour of the resulting regulation control. It is also shown that the lack of uniqueness of the parameter estimates, due to the closed-loop identification, does not effect the control law chosen.

A brief theoretical extension to multivariable systems is included. This multivariable approach reduces to the pole shifting control given in the rest of this thesis for single input systems.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to

Dr. F. P. Stainthorp, for his supervision, guidance and encouragement;

Dr. P. E. Wellstead, for his suggestions and assistance;

The Science Research Council, who provided a grant for this research.

CONTENTS

CHAPTER 1 INTRODUCTION

For many years there have been attempts to make control systems which tune themselves. The resulting adaptive controllers have been used in controlling systems which have slowly varying parameters and systems with constant parameters. The time varying systems are the most obvious candidates for adaptive control since it might be expected that the best performance would be obtained if the controller changed with the system. However, the analysis of time varying systems is complicated and has not progressed as far as the analysis of systems with constant parameters, so at present it may be better to try to analyse self tuning controllers on unchanging systems. It would then be hoped that if the controller could adapt well to constant systems, it may be able to deal with slowly changing systems. The analysis for constant systems is also worthwhile because a self tuning controller would eliminate or reduce the sometimes lengthy process of identifying the system and calculating the required controller parameters. Hence a good self tuning controller may decrease the work involved in tuning the control loop, and reduce the time for which the system has to be run before good control is achieved. The self tuning controller should not need as long to achieve good control because as the control tunes itself the data it is using for the tuning becomes more like the normal operating record. This contrasts with the more usual techniques of identifying systems which do not necessarily produce an accurate description under conditions of normal operation.

There are two main objectives in designing self tuning controllers. The first is to design a scheme which is easy and quick to implement, the second is to design a scheme which is stable, at least for a well defined class of systems, and which gives good control for many systems.

Many different approaches have been used to design self tuning control schemes. One of the simplest ways is to assume that the system is known approximately, use this to estimate the required control,and then use some hill climbing method on the control parameters to produce an optimal output. One example of this is a method proposed by P. Allen [1]

A second approach is to estimate the system parameters, and then use a table of best standard controllers to decide the required control parameters.

There have also been analogue computer methods proposed for tuning controllers.

Another approach was proposed by C. McGreavy and P.J. Gill.[2] They used a Kalman filter to estimate the system's parameters, combined with a P.I.D. control law. This approach of using Kalman Filters could be extended to general linear systems but was not in that paper.

Since many systems can be approximated by a set of linear differential equations the two main objectives of adaptive control would be satisfied by a self tuning controller which could deal with most such linear systems.

In recent years there have been several adaptive control schemes proposed for general linear systems. In 1973 Astrom and Wittenmark suggested a self tuning regulator[3] which asymptotically approaches a minimum variance regulator for linear systems whose z-transforms have no zeros with magnitude greater than one. This consists of a recursive

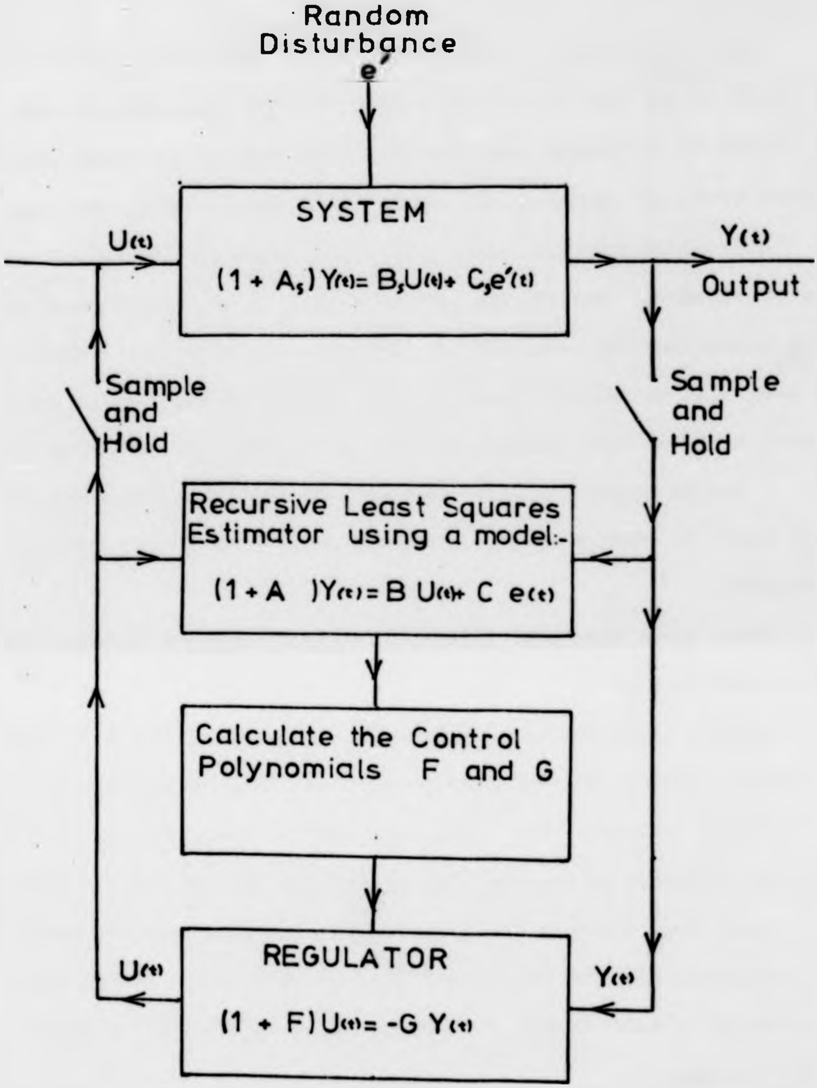


Figure 1.1

Diagram of self tuning controller studied in this thesis. The notation for the models and control is introduced in Chapter 2.

least squares estimator to identify the system coupled with a very simple control law. One problem with this regulator is that one of the system's parameters has to be estimated prior to the start of the control, and completely wrong estimates of this parameter can lead to instability. However, they demonstrated that these stability conditions are not too restrictive, by successfully controlling at least two industrial systems.

In 1974 Astrom and Wittenmark suggested a regulator which asymptotically gives minimum variance control for any linear system,[4] if the system estimates settle. This consists of a recursive least squares identifier followed by a polynomial factorization to find the zeros of the system which are outside the unit circle; the whole being combined with a complicated control law.

In 1975 D. Clarke† produced a self tuning controller which minimizes a function of the input and output variances.[5] This is an interesting scheme because it estimates the required control parameters directly, rather than estimating the system parameters and then calculating the control law.

The approach to self tuning controllers studied in this thesis is to use a recursive least squares estimator to obtain an estimate of the system's parameter values. Then at each sample time to calculate the control law which would move the closed loop poles to specified positions, assuming that the current parameter estimates for the system are correct. This control law is then used to calculate the next control signal. Figure 1.1 gives a diagram corresponding with this self tuning regulator. Both of the self tuning regulators which were suggested by Astrom and Wittenmark and mentioned above are special cases of this approach.

† and P. J. Gawthrop

The analysis of this control scheme begins in Chapter 2 with an introduction to z-transform models to show the likely positions for the poles and the zeros of the system.

In Chapter 3, the stable and conditionally stable regions are then investigated, assuming that the system has known fixed parameters. This case is investigated since it is unlikely that the asymptotic properties of a self tuning controller will be better than those if the control was calculated for and applied to a known fixed system. In Chapter 4 the least squares estimator is introduced, and the problem of non-unique estimates due to feedback are dealt with.

In Chapter 5 it is shown that the asymptotic properties of the self tuning regulator described are the same as would be obtained with the general pole shifting control law applied to a known system even if the disturbance is a coloured noise. Some simulated examples are given to demonstrate the initial behaviour of this regulator. In Chapter 6 several methods of computing this controller are then given together with their computational requirements. Chapter 7 gives a theoretical extension of this regulator to multivariable systems, although it is unlikely that it could be used on systems with more than two inputs and two outputs.

References for Chapter 1

1 P. Allen, D.C. Freshwater, H.W. Kropholler, D.J. Spikins, "A self adjusting digital controller", J. Chem. E., Symposium Series No. 24 (1967 Instn. Chem. Engrs.)

2. C. McGreavy and P.J. Gill, "Self-tuning state variable methods for D.D.C.", Second I.E.E. Conference on "Trends in on-line computer control systems", Sheffield, 1975

3 K.J. Astrom and B. Wittenmark, "on self-tuning regulators", Automatica, Vol. 9, 1973, pp. 185-199

4 K.J. Astrom and B. Wittenmark, "Analysis of a self-tuning regulator for non-minimum phase systems", I.F.A.C. Symposium on Stochastic Control, Budapest, 1974

5 D.W. Clarke and P.J. Gawthrop, "A self-tuning controller", I.E.E., Vol. 122, No.9, September, 1975, pp. 929-934

CHAPTER 2 SYSTEM MODELS

In this chapter several forms of linear dynamic system models are introduced and the relationships between Laplace and z-transforms investigated.

2.1 Dynamic Models for Linear Systems

The subject of dynamic models of systems is covered in detail by A.G.J. MacFarlane in "Dynamic System Models".[1] He deals mainly with continuous system models, but includes sections on sampled systems. A more extensive introduction to sampled systems is given by K.J. Astrom in "Introduction to Stochastic Control Theory".[2] This book investigates linear models which have a random unknown disturbance. A more simple approach to sampled systems is given by K. Steiglitz in "An Introduction to Discrete Systems".[3] This book is written as a textbook for an introductory course on circuit theory, and does not use differential equations or δ-functions. The author claims because it uses only simple mathematics that it can be used before the more difficult theory for continuous systems.

Dynamic System models can be divided into those which have a finite number of parameters, and those with an infinite number of parameters. Two commonly used models with an infinite number of parameters are impulse responses and frequency responses. These can describe all the systems which can be described by the finite models however the larger number of parameters makes them harder to handle, and means that more data is usually required to fit them to a system.

Therefore for control the finite models are usually used. For systems described by linear differential equations two common forms for the models are Laplace Transfer functions[4] and state space representations.[5] The Laplace Transfer function can be very easily obtained from the differential equations by substituting the 's' for the differential operator. A state space representation can also be obtained from the differential equation. For systems described by linear difference equations the corresponding models are z-Transfer functions[3] and discrete state space forms.[6] The z-Transfer function can be obtained from the difference equation by substituting z^{-1} for the delay operator. The difference equation models are useful for sampled systems since they provide an accurate description of the behaviour at the sample times of a system described by linear differential equation, provided that the input to the system remains constant between sample times.

When using a digital computer to control a Linear system the z transforms and discrete state space forms usually hold because the control input to the system will normally be constant between the sample times. For this thesis the z-Transfer function has been used rather than the discrete state space because many of the simpler recent identification methods produce z transfer function models rather than state space ones. However, if some of the states of the system were accessible it may be better to approach the problem of self-tuning controllers by using the discrete state space models.

2.2 Approaches to z Transfer functions

One of the oldest and most common ways of introducing z-transforms is to start with a continuous system described by a Laplace transform and then use an impulse train to sample the input.[7] This approach is not very satisfactory since it implies that the z-transforms only apply when impulses are put into the system, whereas they can apply when no impulses are used. However, it does show a relationship between the Laplace and z-transforms mapping the 's' plane to the 'z' plane by $z = e^{s\Delta t}$ where Δt is the sample period.

MacFarlane describes a different approach to z-transforms in "Dynamic System Models".[6] He considers the sampled systems using discrete data sequences instead of continuous functions. He associates a weighting sequence with each dynamic system; this sequence is very similar to the system's unit impulse response. The system output can then be obtained by convolution of the weighting sequence and the input sequence. This approach is more satisfactory than the previous one since it does not involve the use of impulses. Here, z^{-k} is associated with a member of a set of standard basis sequences; it corresponds with a data sequence which is zero for all except the kth sample, when it is unity.

The approach to z-transforms which shall be used in this thesis is to start from difference equations or recurrence relationships and use z^{-1} as a delay operator. This is suggested by P. Eykhoff,[8] and is also used by K. Steiglitz.[3] This approach has been used in this thesis because it does not use the impulse samplers, and it can be seen that the z-transforms apply to any system that can be

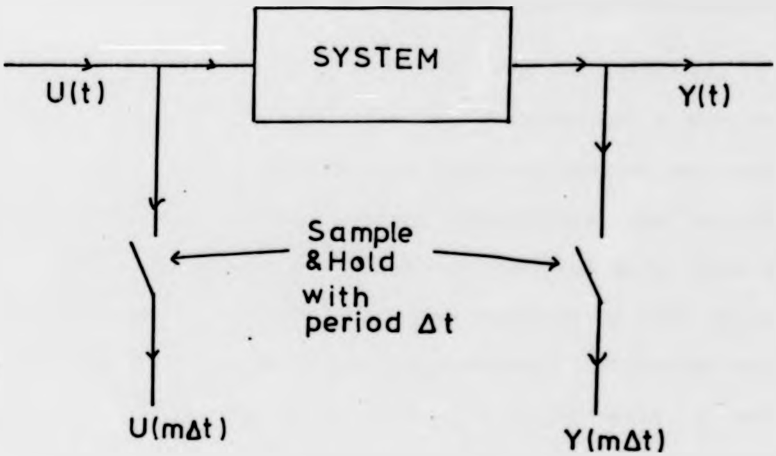


Figure 2.1

A sampled system.

described by a difference equation. It is also nearer to reality since many of the recent identification methods for dynamic systems just fit coefficients to difference equations, or identify a recurrence relationship between the sampled input to a system, and the sampled output.

2.3 Difference equation models

If a continuous system has its input and output observed simultaneously at discrete sample times (Fig. 2.1) then a difference equation describing the system at the sample times would be a relationship between the input and output sequences in the form

$$Y(m\Delta t) + a_1 Y([m-1]\Delta t) + a_2 Y([m-2]\Delta t) + \ldots + a_{na} Y([m-n_a]\Delta t)$$

$$= b_o U(m\Delta t) + b_1 U([m-1]\Delta t) + \ldots + b_{nb} U([m-nb]\Delta t) \qquad (2.1)$$

Where na and nb are constants, Δt is the sample period, and this relationship holds for any m. There is not bound to be such a relationship for a given system but it is shown in section 2.3.1 that there will be such a difference equation if the system can be described by a linear differential equation, and the input is constant between sample times. Introducing z^{-1} as a time delay operator, equation 2.1 can be rewritten

$$(1 + a_1 z^{-1} + \quad + a_{na} z^{-na})\, Y(mt) = (b_o + b_1 z^{-1} + \quad + b_{nb} z^{-nb}) U(mt) \qquad (2.2)$$

or $$(1 + A(z^{-1}))Y(mt) = B(z^{-1})\, U(mt) \qquad (2.3)$$

or $$Y(mt) = \frac{B(z^{-1})}{1 + A(z^{-1})}\, U(mt) \qquad (2.4)$$

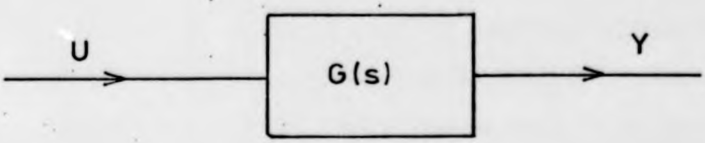


Figure 2.2 a

3 representations of a linear system.

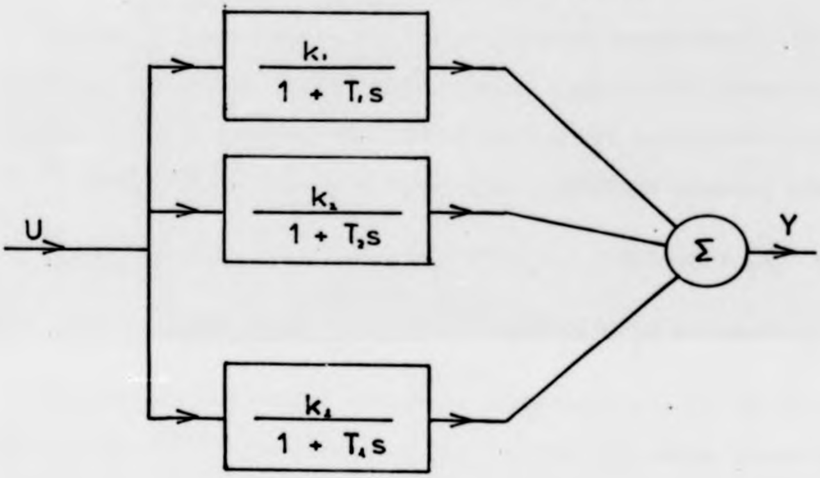


Figure 2.2 b

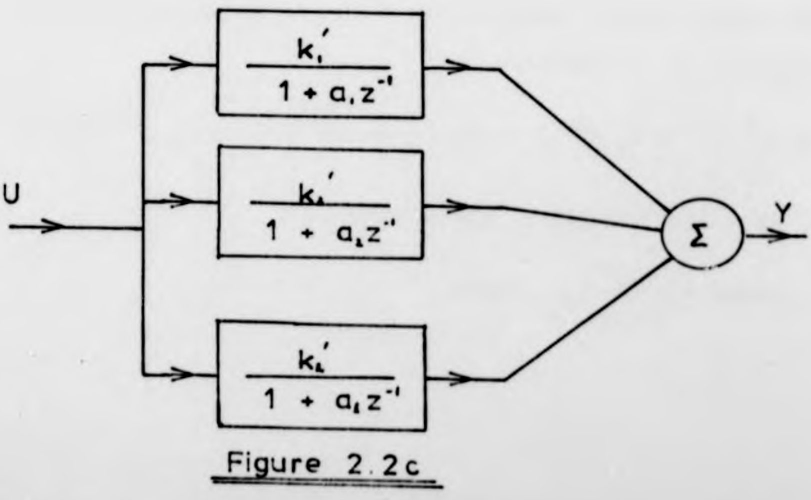


Figure 2.2c

In most cases b_o will be zero since a nonzero value implies that a change on the input of the system immediately produces a change on the output. This would imply that the system is improper, i.e. its Laplace transform has as many zeros as poles. When b_o is zero the next output can be very easily calculated given the past inputs and outputs by substituting in the difference equation 2.1.

2.3.1 Systems described by difference equations

Theorem If a system which is proper and which can be described by a linear differential equation, is sampled at regular intervals then there is a difference equation which relates the observed inputs and outputs if the input is constant between samples. A difference equation can also be found if there is a time delay in the system. Similarly if difference equation 2.1 holds for the sampled inputs and outputs of a system, and a_1 is not zero and $nb \leq na$, and the input to the system is constant between the sample times, and the $(1+A(z^{-1}))$ polynomial does not have roots with negative real part and zero complex part, then a linear differential equation can be found which will describe the system output at the sample times.

These conditions will usually be satisfied by a linear system controlled by a digital computer, since the control signal would commonly be constant between the sample times.

Proof. If the system (Fig. 2.2a) has distinct poles a partial fraction expansion can be used to split it into a sum of simple poles (Fig. 2.2b). Each single pole in this representation can then be converted to a z-transform which gives the same output at the sample intervals. These separate z-transforms can then be recombined because of the linearity of the z operator. The same method can be used if there is a time delay present. The converse can also be handled the same way by using a partial fraction expansion of the z-transform and converting each term separately. The condition nb $\leq$ na ensures that there is a partial fraction expansion. When the poles are not distinct the result can be obtained by taking the limit of a sequence of transfer functions with distinct poles.

Note that the conversion from the z-transform to Laplace transform is not unique since it involves taking logarithims. A digital computer program was written to convert between Laplace and z-transfer functions by using the method adopted for proving this theorem.

a) <u>Conversion of a single pole without time delay</u>

Since the input to the system is constant between sample intervals, and the system is linear, the output can be considered as a sum of step responses. Therefore if the z-transfer function and the Laplace Transfer function both give the same result for a step response they are equivalent for any allowed input.

The unit step response of a system described by a Laplace Transfer function (2.5).

$$\frac{Y}{U} = \frac{1}{1 + Ts} \qquad (2.5)$$

is

$$Y(t) = 0 \qquad t \leq 0$$

$$Y(t) = 1-e^{t/T} \qquad t \geq 0 \qquad (2.6)$$

Suppose the sample period is Δt

$$Y\ (n_{\Delta}t) = 1 - e^{-n\Delta t/T} \qquad n \geq 0$$

$$= 1 - e^{-\Delta t/T} + e^{-\Delta t/T} - e^{-n\Delta t/T}$$

$$= 1 - e^{-\Delta t/T} + e^{-\Delta t/T}\ (1 - e^{-(n-1)\Delta t/T})$$

$$= 1 - e^{-\Delta t/T} + e^{-\Delta t/T}\ Y((n-1)\Delta t) \quad n \geq 1$$

$$Y\ (n\Delta t) = (1-e^{-\Delta t/T})\ U((n-1)\Delta t) + e^{-\Delta t/T}\ Y((n-1)\Delta t) \quad n \geq 1 \qquad (2.7)$$

But if $n \leq 1$ $\quad Y(n\Delta t) = 0$.

Therefore equation 2.7 holds for any n. Introducing z^{-1} as a delay operator gives the z-transform

$$Y(n\Delta t) = z^{-1}\ (1-e^{-\Delta t/T})\ U(n\Delta t)$$

$$+ z^{-1}\ e^{-\Delta t/T}\ Y(n\Delta t) \qquad (2.8)$$

or

$$Y = \frac{z^{-1}\ (1-e^{-\Delta t/T})\ U}{1 - z^{-1} e^{-\Delta t/T}}$$

So the Laplace transfer function 2.5 and the z-transfer function 2.9 correspond to the same system.

The same relationship can be obtained using standard tables of z-transforms,[7] but the approach used here illustrates the close connection between time responses and z-transfer function.

b) <u>Conversion of a single pole with a time delay from Laplace to z-transform</u>

Consider a system described by a Laplace transform.

$$\frac{Y}{U} = \frac{e^{-\tau s}}{1+Ts} \qquad (2.10)$$

Where τ is a time delay which is less than the sample period Δt.

The step response is

$$Y(t) = 0 \qquad t \leq \tau$$

$$Y(t) = 1-e^{-(t-\tau)/T} \qquad t \geq \tau \qquad (2.11)$$

following the same argument as in the case with no time delay

$$Y(n\Delta t) = 1 - e^{-n\Delta t/T} \qquad n > 1$$

$$= (1-e^{-\Delta t/T}) + e^{-\Delta t/T}\, Y((n-1)\Delta t) \qquad (2.12)$$

Also substituting in 2.11 gives

$$Y(0) = 0$$

and $\quad Y(\Delta t) = 1 - e^{-(\Delta t-\tau)/T}$

Therefore the same step response can be obtained at the sample period by a difference equation, when the constants b_1 and b_2 have been chosen to give the same initial response

$$Y(n\Delta t) = b_1\, U((n-1)\Delta t) + b_2\, U((n-2)\Delta t) +$$

$$+ a_1\, Y((n-1)\Delta t) \qquad (2.13)$$

where $\quad a_1 = e^{-\Delta t/T} \qquad (2.14)$

$$b_1 = 1 - e^{-(\Delta t-\tau)/T} \qquad (2.15)$$

and $\quad b_1 + b_2 = (1 - e^{-\Delta t/T})$

Therefore $\quad b_2 = e^{-(\Delta t-\tau)/T} - e^{-\Delta t/T} \qquad (2.16)$

Since this difference equation has the same step response as the Laplace transfer function it will have the same response as the Laplace transfer function with any input signal which is constant between the sample times, due to the linearity of the z-transform, and Laplace transform.

c) <u>Conversion of a single pole from z-transform to Laplace transform</u>

In part a) of this proof the equivalence between the Laplace transfer function

$$\frac{Y}{U} = \frac{1}{1 + Ts}$$

and the z-transfer function

$$\frac{Y}{U} = \frac{z^{-1}\,(1-e^{-\Delta t/T})}{1 - z^{-1}e^{-\Delta t/T}} = \frac{b_1 z^{-1}}{1+a_1 z^{-1}}$$

was demonstrated. This equivalence can be used to convert from the Laplace transfer function to the z-transfer function or from the z-transfer function to the Laplace transfer function. The only difficulty is that the conversion from the z-transfer function requires the evaluation of $\mathrm{Log}(-a_1)$ since

$$\frac{-\Delta t}{T} = \mathrm{Log}(-a_1)$$

which implies $$T = \frac{-\Delta t}{\mathrm{Log}(-a_1)} \qquad (2.17)$$

Evaluating $\mathrm{Log}(-a_1)$ is difficult since the logarithm is not a uniquely defined number, it is only defined to within the addition of a multiple of $2\pi j$. However, usually this ambiguity can be removed by adding or subtracting multiples of $2\pi j$ to the $-\Delta t/T$ obtained in order to minimize the modulus of the imaginary part of the number.

The approach which has been used in this section in converting Laplace transfer functions to z-transfer functions can be considered as a special case of that proposed by Edwards.[9] He obtained a state space representation for the continuous system, and then converts to a sampled state space, and thus obtains the z-transfer function. Time delays can be added to this more general method in a similar way to that used for a simple pole, since he does the conversion from the continuous state space using the time response.

2.4 Expected regions for the Poles and Zeros of z-transfer functions

Since the most commonly used system representation for control systems is the (s) transfer function the expected regions for the poles and zeros of the z-transfer functions have been investigated by considering Laplace transfer functions and finding the corresponding z-transfer functions.

2.4.1 Poles

There is a one to one correspondence between the Laplace transfer function poles and the z-transfer function poles, which can be seen from the method of conversion suggested in section 2.3. A pole with a time constant T, i.e. a Laplace transfer function pole at $-{}^{1}/T$, corresponds to a z-transfer function pole at $e^{-\Delta t/T}$ using equation 2.9.

Therefore poles in the left half plane in the Laplace transfer function correspond with poles inside the unit circle in the z-transfer function. So a system will be stable if all its poles are within the unit circle. Real poles in the Laplace transfer function become positive real poles in the z-transfer function. However complex pairs of poles in the Laplace transfer function can become a pair of real poles in the z-transfer function, although they usually become a pair of complex poles. They can become real poles due to the periodic nature of $e^{-\Delta t/T}$ for complex T.

2.4.2 Poles with short sampling periods

As the sample period approaches zero the z-transfer function poles will all approach +1. This implies that if the sample period becomes too short the rounding errors in the z-transfer function will become significant, particularly if there are several poles.

2.4.3 Zeros

Generally the zeros of the Laplace and z-transfer functions are not in one to one correspondence, since the number of zeros in the z-transfer function is usually one less than the number of poles, and hence independent of the number of Laplace transfer function zeros. The mapping of the zeros is not defined in a simple manner like the mapping of poles, but depends on the pole values as well as the zeros and the sample interval.

2.4.4 Zeros of a z-transform of a system consisting only of integrators

Suppose a system has a Laplace transfer function

$$Y = \frac{U}{s^n} \tag{2.18}$$

The systems step response is

$$Y(t) = \frac{t^n}{n!} \tag{2.19}$$

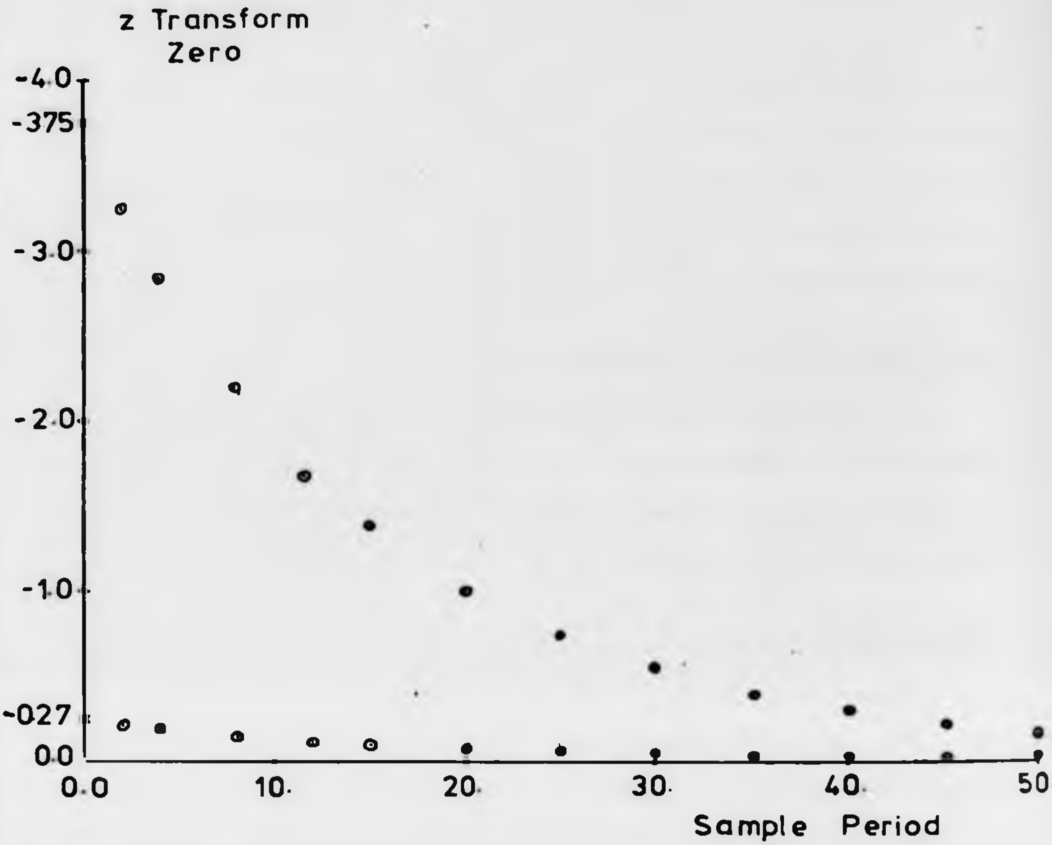


Figure 2.3

Zeros of the z-transform of a system

$$Y = \frac{U}{(1+10s)(1+11s)(1+12s)}$$

as a function of the sample period.

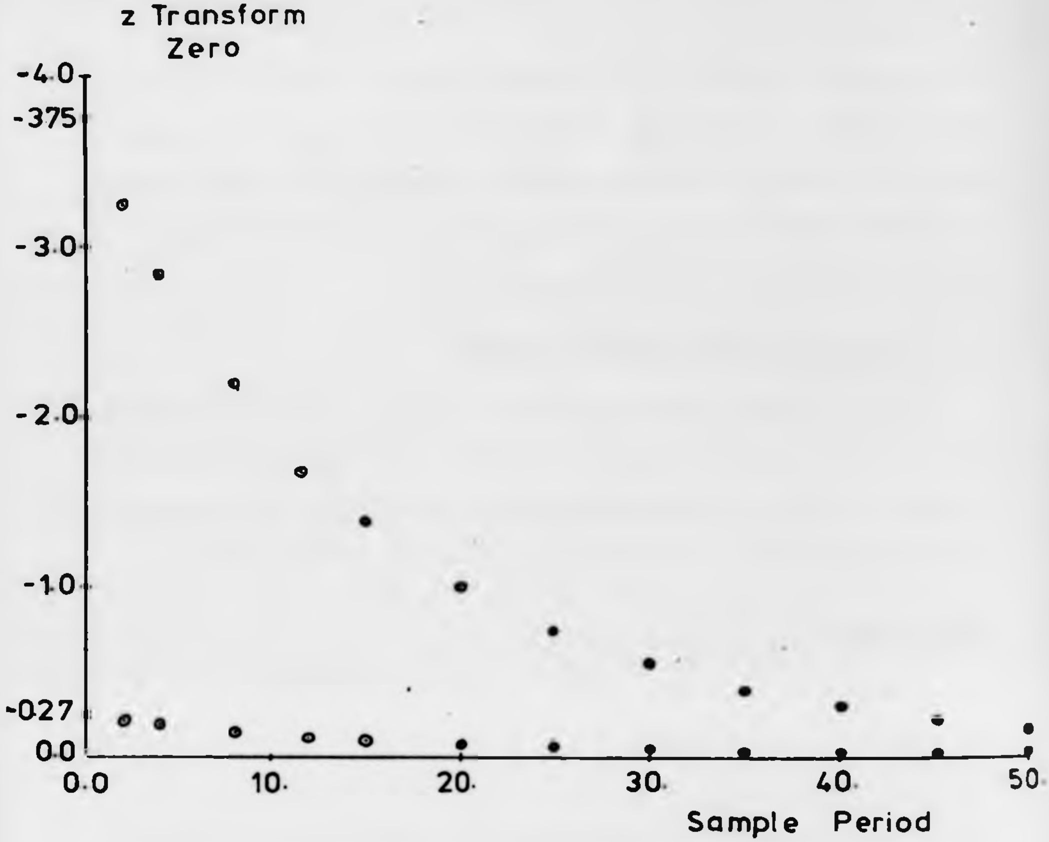


Figure 2.3

Zeros of the z-transform of a system

$$Y = \frac{U}{(1+10s)(1+11s)(1+12s)}$$

as a function of the sample period.

N	NUMERATOR POLYNOMIAL	ZEROS
1	z^{-1}	0.00
2	$(z^{-1} + z^{-2})/2!$	-1.00
3	$(z^{-1} + 4z^{-2} + z^{-3})/3!$	-3.73,-0.27
4	$(z^{-1} + 11z^{-2} + 11z^{-3} + z^{-4})/4!$	-9.90,-1.00,-0.10
5	$(z^{-1} + 26z^{-2} + 66z^{-3} + 26z^{-4} + z^{-5})/5!$	-23.20,-2.32,-0.43 -0.04
6	$(z^{-1}+57z^{-2}+302z^{-3}+302z^{-4}+57z^{-5}+z^{-6})/6!$	-51.29,-4.54,-1.00 -0.22,-0.02
7	$(z^{-1}+120z^{-2}+1191z^{-3}+2416z^{-4}$ $+1191z^{-5}+120z^{-6}+z^{-7})/7!$	-107.78,-8.17,-1.87 -0.54,-0.12,-0.01

TABLE 2.1

NUMERATORS OF THE 2 TRANSFER FUNCTIONS OF SYSTEMS OF JUST N INTEGRATORS as can be found in a standard table of Laplace transform pairs.

The z-transform of these functions $Y = t^n$ are given in one of the appendices of a book by E.I. Jury.[7] The z-transfer function of the system can then be obtained by dividing by $\frac{z}{z-1}$ to take account of the step input. The corresponding z-transform numerators are given in Table 2.1. This table also gives the zero values found by factorizing the numerator polynomials.

It will be noticed that in each case half of the zeros are outside the unit circle. This occurs because the coefficients of the polynomials are symmetrical about the largest values, and so if there is a zero with a value 'a' there is also a zero with value $^1/a$.

2.4.5 Zero positions as sample period approaches zero

Suppose that the Laplace transform of a system has np poles and nz zeros. Then as the sample period is made small the system looks more like np-nz integrators for the first few samples. However the z-transform is very closely linked with the time response for a few consecutive samples so it could be expected that the z-transform will be similar to that for np-nz integrators. This implies that nz of the z-transform zeros will nearly cancel poles. So it could be expected that nz of the z-transform zeros will approach +1 since all the poles approach +1, and the other np-nz-1 zeros will approach the position of the z-transform zeros of a system described by np-nz integrators.

Two cases were examined to confirm the above reasoning. Figure 2.3 shows the zeros of the z-transfer function corresponding with the Laplace transfer function

$$Y = \frac{U}{(1 + 10s)(1 + 11s)(1+12s)} \tag{2.20}$$

It can be seen that the zeros seem to approach the zeros of a system of just three integrators, as the sample time becomes small. However there were problems with the conversion becoming ill conditioned for short sample periods.

The second example uses a more complicated Laplace transform to test that some of the zeros move towards +1. The Laplace transfer function used was

$$\frac{Y}{U} = \frac{(1+2s)(1 + (1+j)s)(1+ (1-j)s)}{(1+s)(1+1.5s)(1+25s)(1+3s)(1+0.7(1+j)s)(1+0.7(1-j)s)} \tag{2.21}$$

The expected zero positions for small sample times are three at unity, one at -0.27 and one at -3.75.

With a sample period of 0.5 the zeros were

$.754 \pm 0.192\, j$, $.778$, -0.200 and -2.79

With a sample period of 0.1 the zeros were

$.953 \pm 0.047\, j$, 0.943, -0.25 and -3.51

Shorter sample periods were tried but the conversion had become too ill conditioned for the reverse procedure of converting the z-transform back to the Laplace transform, and so the results could not be checked. However it can be seen that the zeros are approaching the expected values.

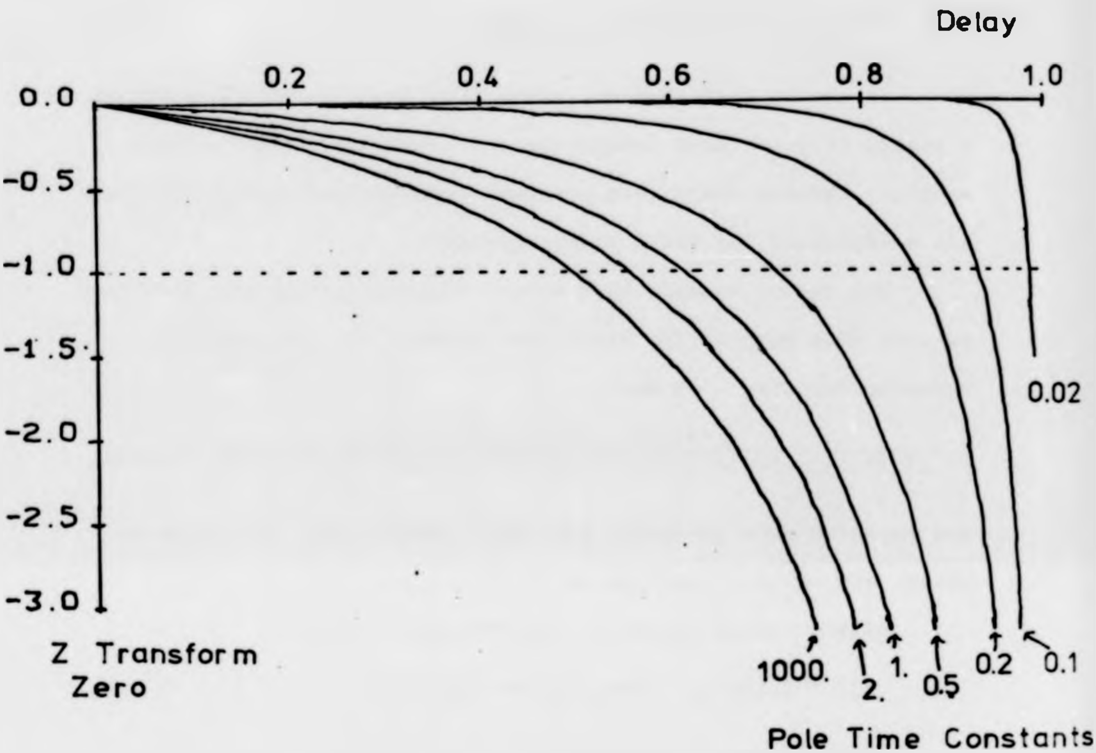


Figure 2.4

z-transform zero caused by the introduction of a delay, as a function of the delay for various pole time constants.

2.4.6 Zeros due to a time delay of less than one sample period

When a system is described by

$$Y = \frac{e^{-\tau s}\,U}{1+Ts}$$

section 2.3 shows that the sampled system is described by

$$Y(t) = \frac{\left[(1-e^{-(\Delta t-\tau)/T})z^{-1} + (e^{-(\Delta t-\tau)/T}-e^{-t/T})\,z^{-2}\right]U(t)}{1-e^{-\Delta t/T}z^{-1}} \qquad (2.22)$$

So the zero created by the time delay is

$$\frac{e^{-\Delta t/T} - e^{-(\Delta t-\tau)/T}}{1-e^{-(\Delta t-\tau)/T}} \qquad (2.23)$$

Figure 2.4 shows the zero value as a function of the time delay for various values of the pole time constant T. As the delay τ approaches the sample interval the zero approaches $-\infty$. It can be seen from the graph that the zero can be outside the unit circle if the delay is more than half the sample period. It may be expected that the zero created by a time delay on a more complicated system would behave in a similar manner to that created on this very simple system of a single pole.

2.5 Z-transfer functions with delays of more than one sample interval.

If a system has a time delay τ where

$$\tau = k\Delta t + \tau'$$

with $0 \leq \tau' < \Delta t$

Then the delay of $k\Delta t$ can be dealt with by multiplying the $B(z^{-1})$ polynomial by z^{-k} since z^{-1} is a time delay of one sample interval. The remainder of the time delay τ' can be dealt with by the method suggested in section 2.3.

Therefore if a system has a Laplace transfer function $(e^{-k\Delta ts}.e^{-\tau' s})\ G(s)$ and the z-transfer function of a system $e^{-\tau' s}\ G(s)$ is

$$Y = \frac{B(z^{-1})\ U}{1+A(z^{-1})}$$

Then the z-transfer function of the system with time delay is

$$Y = \frac{z^{-k}\ B(z^{-1})\ U}{(1+A(z^{-1}))} \qquad (2.24)$$

However the self tuning regulator suggested by Astrom for minimum phase systems requires a model of the form

$$(1 + z^{-k}\ A'(z^{-1}))\ Y = z^{-k}\ B'(z^{-1})\ U \qquad (2.25)$$

This form can be obtained by multiplying the top and bottom of the transfer function in 2.24 by a suitable polynomial $P(z^{-1})$

Where $\quad (1+A(z^{-1}))\ P(z^{-1}) = 1 + z^{-k}\ A'\ (z^{-1}) \qquad (2.26)$

and $\quad B(z^{-1})\ P(z^{-1}) = B'\ (z^{-1}) \qquad (2.27)$

and $\quad P(z^{-1}) = 1 + p_1 z^{-1} + \ldots\ldots + p_k z^{-k}$

The condition that the coefficients of $z^{-\ell}$ for $\ell = 1, \ldots k$ in the polynomial $(1 + A(z^{-1}))P(z^{-1})$ are all zero uniquely define the required polynomial $P(z^{-1})$.

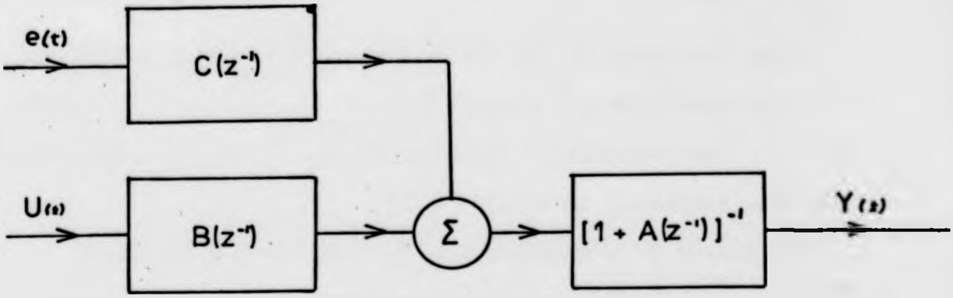


Figure 2.6

Introduction of a coloured random noise.

Figure 2.5c

ZEROS WITH TWO POLES AT 0.30

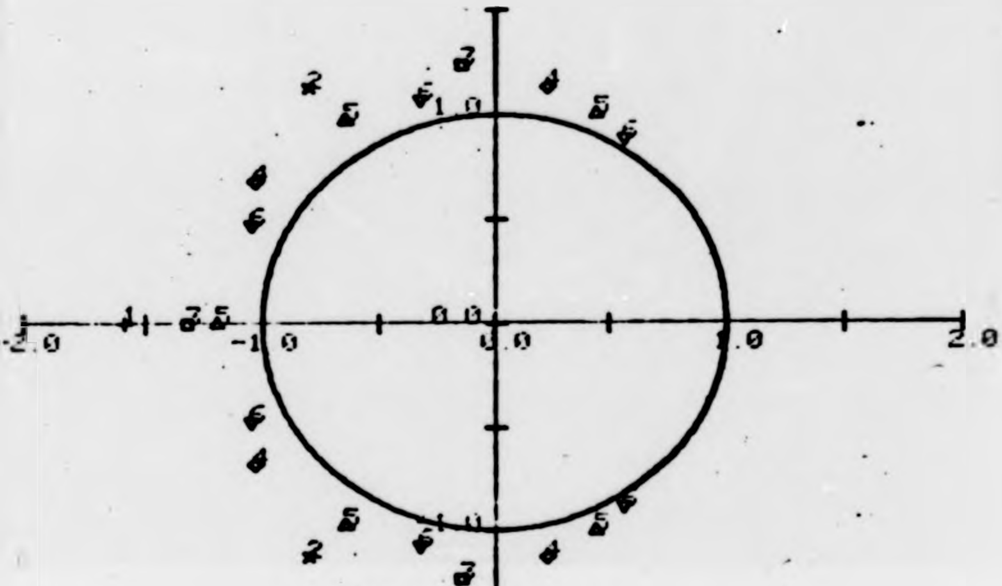


Figure 2.5d

ZEROS WITH TWO POLES AT 0.90

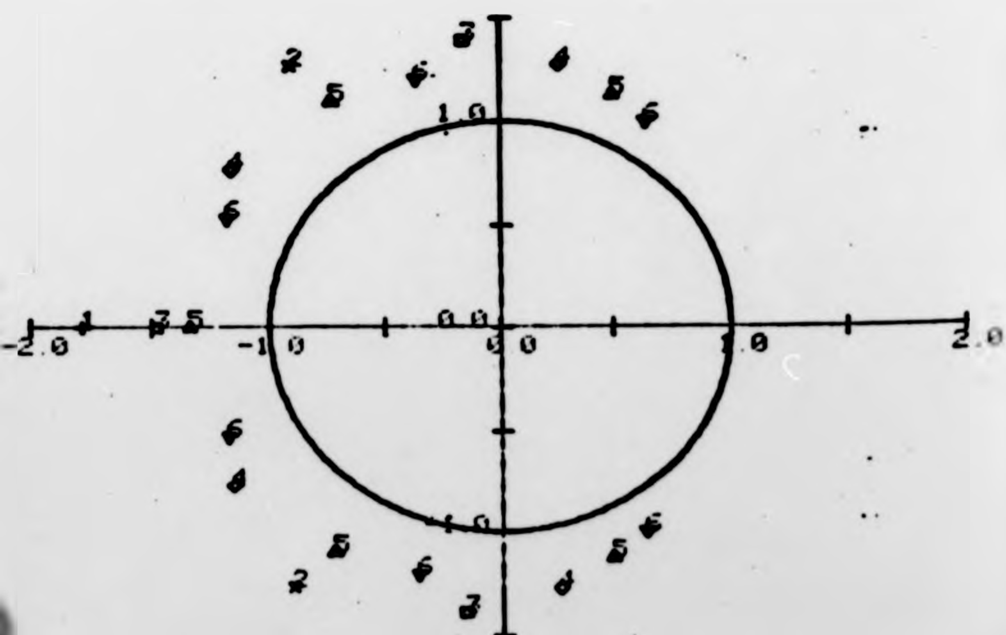


Figure 2.5a

ZEROS WITH TWO POLES AT 0.60

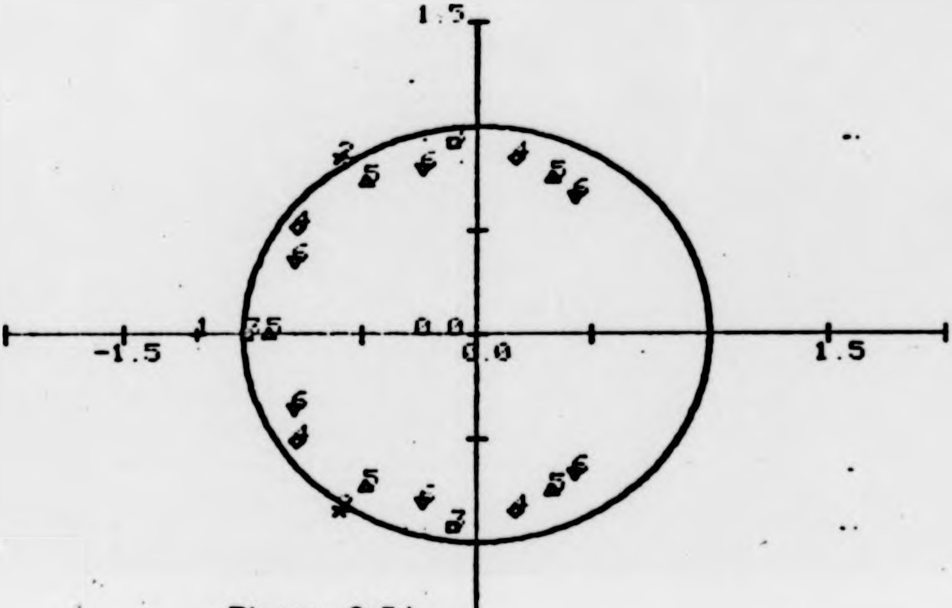


Figure 2.5b

ZEROS WITH TWO POLES AT 0.70

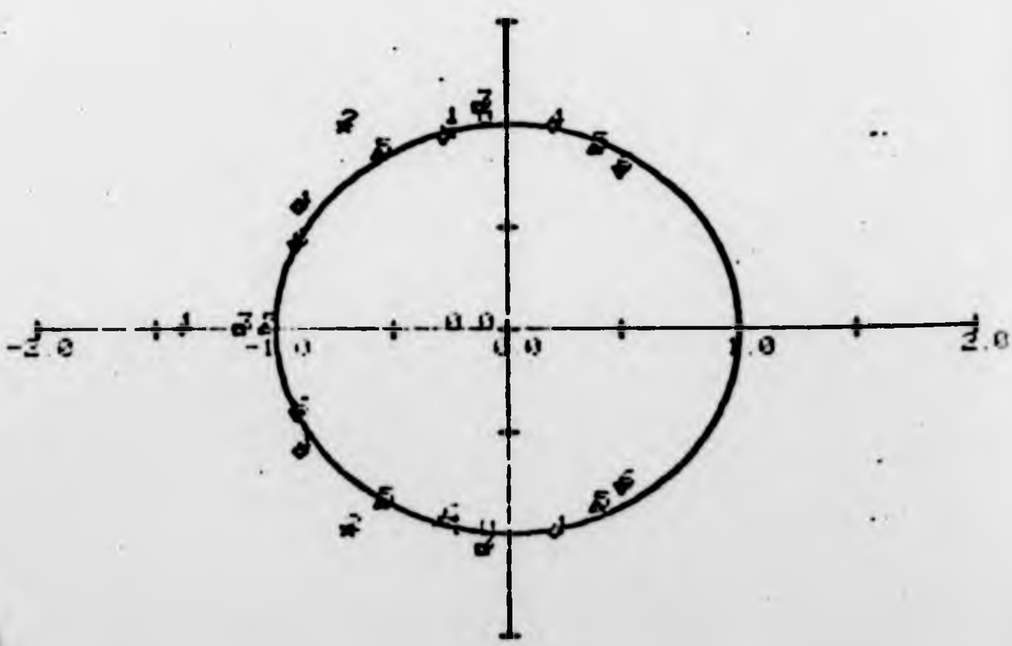


Figure 2.5 a, b, c, d show the positions of the roots of the polynomial $P(z^{-1})$ for systems with two z-transfer function poles at 0.6, 0.7, 0.8, and 0.9 respectively. On each plot the positions of the roots are shown for time delays k of 1, 2, 3, 4, 5 and 6. It will be noticed that the roots of $P(z^{-1})$ can frequently be outside the unit circle.

This same method of changing the form of the model can be used when there are random disturbances in the model, such as those introduced in the next section. In that case the disturbance polynomial, the $B(z^{-1})$ polynomial and the $(1 + A(z^{-1}))$ polynomial all have to be multiplied by the $P(z^{-1})$ polynomial.

The required $P(z^{-1})$ polynomial can be easily constructed from the $A(z^{-1})$ polynomial and the time delay k using the conditions which was mentioned above. More detail of the computation is given in Chapter 6.

2.6 Random disturbances

Many systems have random disturbances which effect their outputs. These can be included in the z-transfer function models by adding a noise model to the system. Equation 2.3 becomes

$$(1 + A(z^{-1}))\ Y = B(z^{-1})\ U + C(z^{-1})\ e \tag{2.28}$$

Where e is a white noise sequence and $C(z^{-1})$ is a polynomial in z^{-1} which colours the noise. Most forms of noise can be described in this way, ~~however the C polynomial may have to be very long~~. Fig. 2.6 gives a diagrammatic representation of this model of a disturbed system. For the least squares method of fitting parameter values to the A and B polynomials to give unbiased results the $C(z^{-1})$ polynomial should just be a constant. The residuals in the least squares fitting are then uncorrelated. This situation of having uncorrelated residuals should not be confused with the case which occurs more frequently of having white measurement noise on the observations of the system.

2.7 Constant offsets

Many systems give a non zero output for a zero input. This can be included in the equation (2.28) by adding a constant offset term d

$$(1 + A(z^{-1}))Y = B(z^{-1})\, U + C(z^{-1})e + d \qquad (2.29)$$

This form of model can also deal with the case of a non zero input being required for the system to give a zero output.

References for Chapter 2

1 "Dynamic System Models", A.G.J. MacFarlane. G.G. Harrap & Co. Ltd. 1970.

2 "Introduction to Stochastic Control Theory", K.J. Astrom. Academic Press, 1970.

3 "An Introduction to Discrete Systems", K. Steiglitz. John Wiley & Sons, Inc., 1974.

4 Chapter 3.1 of reference 1.

5 Chapter 6 of reference 1.

6 Chapter 6.4 of reference 1, and reference 2, section 3.2.

7 "Theory and Application of the 'z' Transform Method", E.T. Jury. John Wiley & Sons, Inc., 1964.

8 "System Identification", P. Eykhoff. John Wiley & Sons, Inc., 1974, section 4.2.2.

9 "Pulsed frequency response calculation for a digital control system", Edwards, Electronic Letters, Vol. 4, no. 1, 12th Jan. 1968, pp. 19-21.

CHAPTER 3 CONTROL LAWS

3.1 Regulators

When a system has no setpoint changes, the control scheme usually aims to counteract the effects of disturbances on the system. In this case the controller is called a regulator. The general form of a linear regulator is very similar to the form for a dynamic model of a linear system. There are two main differences. The first is that a change in the system output usually produces an immediate change in the control input, while on most systems a change on the input does not produce an immediate change on the output. The second main difference is that there is usually no random disturbance in the calculation of the control signal.

The general linear model for a sampled data regulator becomes

$$(1 + F(z^{-1}))\ U = -\ G(z^{-1})\ Y \tag{3.1}$$

Where $F(z^{-1}) = f_1\ z^{-1} + \ldots + f_{nf}\ z^{-nf}$

$$G(z^{-1}) = g_o + g_1\ z^{-1} + \ldots + g_{ng}\ z^{-ng}$$

c.f. equation 2.3.

(The minus sign has been introduced on the right of equation 3.1 since it is usual to have negative feedback rather than positive feedback.)

The general closed loop equation (3.2) can be obtained by substituting for U in the open loop equation 2.28, using the feedback law in equation 3.1.

$$(1 + A(z^{-1}))Y = \frac{-z^{-k}\ B(z^{-1})\ G(z^{-1})\ Y}{1 + F(z^{-1})} + Ce \tag{3.2}$$

rearranging gives

$$\frac{[(1 + A(z^{-1}))(1 + F(z^{-1})) + z^{-k}\ B(z^{-1})\ G(z^{-1})]\ Y}{1 + F(z^{-1})} = Ce \tag{3.3}$$

or

$$[(1+A(z^{-1}))(1+F(z^{-1})) + z^{-k}B(z^{-1})\ G(z^{-1})]\ Y = (1+F(z^{-1}))\ C(z^{-1})e \tag{3.4}$$

Therefore the closed loop system poles are the roots of the polynomial $T(z^{-1})$ where

$$T(z^{-1}) = 1 + t_1 z^{-1} + \qquad + t_{nt} z^{-nt} \tag{3.5}$$

$$T(z^{-1}) = (1 + A(z^{-1}))(1 + F(z^{-1})) + z^{-k}\ B(z^{-1})\ G(z^{-1}) \tag{3.6}$$

$T(z^{-1})$ is called the characteristic polynomial of the closed loop system. The closed loop system is stable if all the roots of $T(z^{-1})$ are within the unit circle.

3.2 Pole shifting regulators

Suppose that the required closed loop characteristic polynomial is $T'(z^{-1})$. The system can be made to have this characteristic polynomial by choosing F and G in the control law such that

$$T'(z^{-1}) = (1 + A(z^{-1}))(1 + F(z^{-1})) + z^{-k}\ B(z^{-1})G(z^{-1}) \tag{3.7}$$

since then $T(z^{-1})$ will equal $T'(z^{-1})$.

The orders of F and G will then usually satisfy 3.8.

$$nt \leq na + nf = k + nb + ng \tag{3.8a}$$

$$nf = nb + k - 1 \tag{3.8b}$$

$$ng = na - 1 \tag{3.8c}$$

The orders of F and G can be smaller in special cases, but generally they are defined by 3.8b and c, ensuring that there are parameter values giving the required characteristic polynomial. These orders can be obtained by considering equation 3.7 which gives a set of linear simultaneous equations defining F and G. The coefficients of each power of z^{-1} define one of the linear simultaneous equations. For example the coefficients of z^{-1} give

$$a_1 + f_1 + b_1 g_o = t_1 \qquad \text{if } k = 0$$

The closed loop equation then becomes

$$T'(z^{-1})Y = T(z^{-1})Y = (1 + F(z^{-1}))\ C(z^{-1})\ e \tag{3.9}$$

3.3 Minimum variance regulators *

3.3.1 No Time delays

If $T\ (z^{-1})$ is chosen to be equal to $(1 + F(z^{-1}))C(z^{-1})$ the closed loop equation would become

$$Y = e$$

This would clearly be the minimum variance feedback controller since the output is equal to the random disturbance which is driving the system. To decrease the output any more would mean that the random

* For simplicity the case with coloured noise has not been included since the result for coloured noise is not used elsewhere in this thesis.

disturbance on the system would have to be known before it affected the output. In the simple case of a white noise disturbance where $C(z^{-1})$ is unity the control is chosen such that

$$T'(z^{-1}) = (1 + F(z^{-1})) = (1 + A(z^{-1}))(1 + F(z^{-1})) + B(z^{-1})G(z^{-1})$$

rewriting gives $$A(z^{-1})(1 + F(z^{-1})) + B(z^{-1})G(z^{-1}) = 0 \qquad (3.10)$$

The solution of 3.10 can be seen to be

$$G(z^{-1}) = - \ell z\, A(z^{-1}) \qquad (3.11)$$

and $$(1 + F(z^{-1})) = \ell z\, B(z^{-1}) \qquad (3.12)$$

where ℓ is any scalar and the orders of A, B, G and F are as in section 3.2. This very simple control law was used by Astrom[1] in a self tuning regulator for minimum phase systems.

Substitution in 3.10 for $T(z^{-1})$ gives

$$T'(z^{-1}) = \ell z\, B(z^{-1}) \qquad (3.13)$$

for minimum variance control.

3.3.2 With time delays

If the closed loop equation 3.4 is multiplied by the polynomial $P(z^{-1})$ defined in section 2.5 it becomes

$$((1+A)(1+F)P + z^{-k}\, BGP)\ Y = P(1+F)C\ e \qquad (3.14)$$

In the simple case with white noise C is unity and 3.14 becomes

$$((1+F)(1+z^{-k}A') + z^{-k}BGP)Y = P(1+F)e \qquad (3.15)$$

Dividing by (1+F) gives

$$\therefore \quad Y = P(z^{-1})\, e - z^{-k}\left[A'(z^{-1}) + \frac{PBG}{1+F}\right] Y \qquad (3.16)$$

Writing the variance of Y as Var[Y]

$$\mathrm{Var}[Y] = \mathrm{Var}\left[P(z^{-1})\, e - z^{-k}\left(A'(z^{-1}) + \frac{PBG}{1+F}\right) Y\right]$$

$$= \mathrm{Var}\left[P_o\, e + P_1 z^{-1}\, e + \quad + P_k z^{-k}\, e - z^{-k}\left(A'(z^{-1}) + \frac{PBG}{1+F}\right) Y\right] \qquad (3.17)$$

However, the sequence of e's is a sequence of independent random variables which are also independent of the last term in the expression for the variance, since this last term depends only on previous output values. Therefore

$$\mathrm{Var}[Y] = P_o\, \mathrm{Var}[e] + P_1 \mathrm{Var}\,[z^{-1}\, e] + \quad + P_k\, \mathrm{Var}\,[z^{-k}\, e] + \mathrm{Var}\left[-z^{-k}\left(A'(z^{-1}) + \frac{P(z^{-1})\, B(z^{-1})\, G(z^{-1})}{1 + F(z^{-1})}\right) Y\right] \qquad (3.18)$$

Therefore the output variance will be minimized if the control is chosen to remove this last term, i.e.

$$A'(z^{-1}) + \frac{P(z^{-1})\, B(z^{-1})\, G(z^{-1})}{(1 + F(z^{-1})} = 0 \qquad (3.19)$$

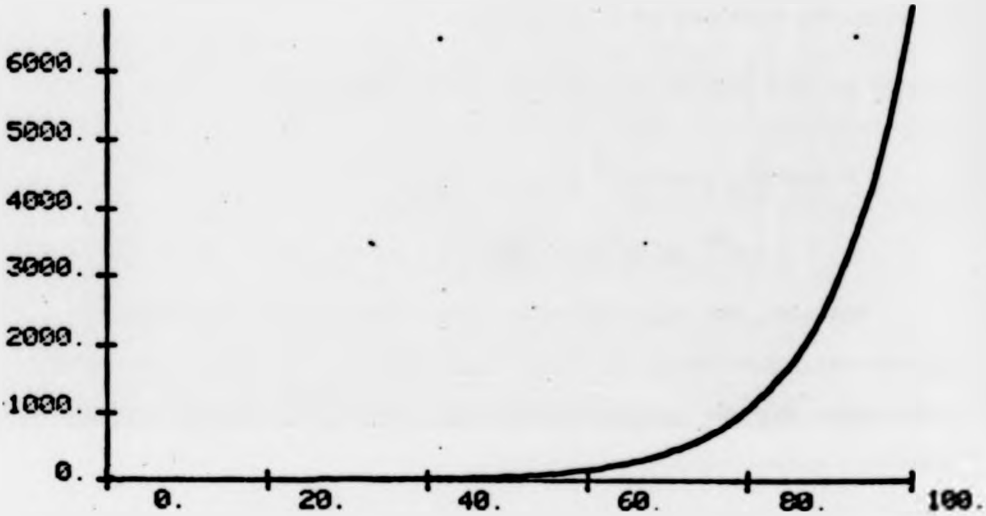
$$(1 + F)A' + PBG = 0 \qquad (3.20)$$

c.f. 3.10 without time delays.

The solution of 3.20 can be seen to be

FIGURE 3.1

INPUT TO A SYSTEM $(1-0.5Z^{-1})Y=(Z^{-1}-1.1Z^{-2})U$ WITH A MINIMUM VARIANCE CONTROLLER



$$G(z^{-1}) = -\ell z A'(z^{-1}) \tag{3.21}$$

$$1 + F(z^{-1}) = z P(z^{-1}) B(z^{-1}) = \ell z B'(z^{-1}) \tag{3.22}$$

This corresponds to choosing $T'(z^{-1})$ in 3.7 as

$$T'(z^{-1}) = (1+A(z^{-1}))(1+F(z^{-1})) + z^{-k} B(z^{-1}) G(z^{-1})$$

$$= \frac{1}{P(z^{-1})} \left(P(z^{-1})(1+A(z^{-1}))(1+F(z^{-1}))+z^{-k}P(Z^{-1})B(z^{-1})G(z^{-1})\right)$$

$$= \frac{1}{P(z^{-1})} \left((1+ z^{-k}A'(z^{-1}))(1+F(z^{-1}))+z^{-k} B'(z^{-1})G(z^{-1})\right)$$

using 3.21 gives

$$T'(z^{-1}) = \frac{1}{P(z^{-1})} \left(1 + F(z^{-1}) \right)$$

using 3.22 gives

$$T'(z^{-1}) = \ell z B(z^{-1})$$

$$\text{where } \ell = {}^{1}/b_1 \tag{3.23}$$

3.3.3 Stability of minimum variance regulators

It will be noticed that if the polynomial $B(z^{-1})$ has any roots outside the unit circle in the z plane the minimum variance regulator will produce an unstable system, since the closed loop poles are at the roots of $B(z^{-1})$ (Equations 3.13 and 3.23). A simple example of this is shown in Figure 3.1. This figure shows the input to a system described by

$$(1 - 0.5 z^{-1}) Y = (z^{-1} - 1.1 z^{-2}) U \tag{3.24}$$

with Y(o) set to -1 to act as an initial disturbance. The minimum variance control used is

$$U = \frac{-0.5Y}{1-1.1z^{-1}}$$

The output is zero after the first sample, however, the input grows exponentially. This control is clearly unacceptable.

3.3.4 Stable minimum variance control

The minimum variance control subject to the closed loop system being stable has been found by V. Peterka.[2] He found the minimum variance control strategy for several different forms of the system model.

Astrom and Wittenmark expressed the control in polynomial form[3] using a slightly different notation from that used in this thesis. In appendix B the closed loop characteristic polynomial is found starting from Astrom and Wittenmark's work. It is deduced there that the values of z which make the $T'(z^{-1})$ zero for this control are all the zeros of the system which have modulus less than unity, and the reciprocals of the zeros with modulus greater than or equal to unity.

3.4 A Stable Regulator

The stable regulator in section 3.3.4 requires a polynomial factorization which can be time consuming. However, if the control is chosen by setting $T'(z^{-1})$ in equation 3.7 to a constant polynomial with all its roots inside the unit circle in the z plane, the system will be stable if it is linear. The simplest such polynomial is $T'(z^{-1}) = 1$, but it will be shown in section 3.6 that this can easily result in conditionally stable systems. The conditional stability can frequently be removed by a suitable choice of $T'(z^{-1})$.

3.5 Variance of the output

With the regulator specified by equation 3.7 the closed loop output can be obtained from equation 3.9.

$$Y = \frac{(1 + F(z^{-1}))\ C(z^{-1})\ e}{T'(z^{-1})} \tag{3.25}$$

The output variance can be easily calculated using a method described by Astrom.[4] He gives a short Fortran program for a digital computer, which will calculate the output variance if the variance of the disturbance e is unity. This program uses a pair of tables one of which is very similar to the table used in Routh's stability test for linear dynamic systems.

3.6 Stability and Conditional Stability

In linear systems the stability depends only on the positions of the roots of $T'(z^{-1})$ since the characteristic polynomial $T(z^{-1})$ is equal to $T'(z^{-1})$. However, if the system is nonlinear, for example if there are limits on the allowed control signal, then other methods of determining stability have to be used. One case where instability could be expected is when the system is only conditionally stable, since then a limit on the control signal could effectively decrease the loop gain and so cause instability.

A closed loop system is called conditionally stable if a decrease in the feedback gain can produce instability. Conditional stability can be discovered by introducing a variable gain λ into the feedback law, and looking for values of λ between 0 and 1 which make some of the roots of the characteristic polynomial go outside the unit circle. Putting this λ in equation 3.6 gives

$$T(z^{-1}) = (1+A(z^{-1}))(1+F(z^{-1})) + \lambda\, z^{-k}\, B(z^{-1})\, G(z^{-1}) \tag{3.26}$$

A root locus diagram can then be used to track the roots of $T(z^{-1})$ for varying values of λ. The root locus diagram required is the one for a system

$$Y = \frac{z^{-k}\, B(z^{-1})\, G(z^{-1})\, U}{(1+A(z^{-1}))(1+F(z^{-1}))} \tag{3.27}$$

Similarly the Nyquist diagram of this system 3.27 can be used to check for conditional stability.

Root locus diagrams are described by Jury in reference 5.

3.6.1 Minimum variance regulator with no time delays

Substituting for $1 + F(z^{-1})$ in characteristic equation 3.26 using the control equation 3.12 gives

$$T(z^{-1}) = \ell\, z(1+A(z^{-1}))\, B(z^{-1}) + \lambda\, B(z^{-1})G(z^{-1})$$

$$= B(z^{-1})(\ell\, z(1+A) + \lambda\, G(z^{-1})) \tag{3.28}$$

Substituting for $G(z^{-1})$ in 3.28 using the other control equation 3.11 gives

$$T(z^{-1}) = B(z^{-1})(\ell\, z(1+A(z^{-1}))-\ell z\lambda A(z^{-1}))$$

$$T(z^{-1}) = \ell z\, B(z^{-1})(1 + A(z^{-1}) - \lambda A(z^{-1})) \tag{3.29}$$

Therefore the system is conditionally stable if any of the roots of $B(z^{-1})(1+A(z^{-1}) - \lambda A(z^{-1}))$ are outside the unit circle for λ less than unity. However, roots of $B(z^{-1})$ outside the unit circle make the system unstable, so the conditional stability can only be caused by the roots of $(1+A(z^{-1}) - \lambda A(z^{-1}))$. If the system is open loop unstable a value of $\lambda = 0$ makes this polynomial have an unstable root, and so the system would be conditionally stable. However, this is not an important case of conditional stability because it cannot be avoided by choosing a different controller. The more important cases are when the system is open loop stable but closed loop conditionally stable, since these can be avoided by suitable choice of regulator.

In the rest of this section an example will be given of conditional stability occurring with a system having three poles, and a proof will be presented that conditional stability with this regulator requires that the system has at least three poles.

<u>Proof that conditional stability cannot occur with just two poles</u>

Suppose the denominator of the system's transfer function is $1 + A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2}$. The characteristic polynomial of the closed loop system with the minimum variance regulator is

$$\begin{aligned} T(z^{-1}) &= z\, B(z^{-1})\, (1 + A(z^{-1}) - \lambda A(z^{-1})) \\ &= z\, B(z^{-1})\, (1 + (1-\lambda)a_1 z^{-1} + (1-\lambda)a_2\, z^{-2}) \qquad (3.30) \end{aligned}$$

The polynomial $1 + A(z^{-1}) - \lambda A(z^{-1})$ can either have two complex roots or two real roots. If there are two real roots for $\lambda = 0$ then

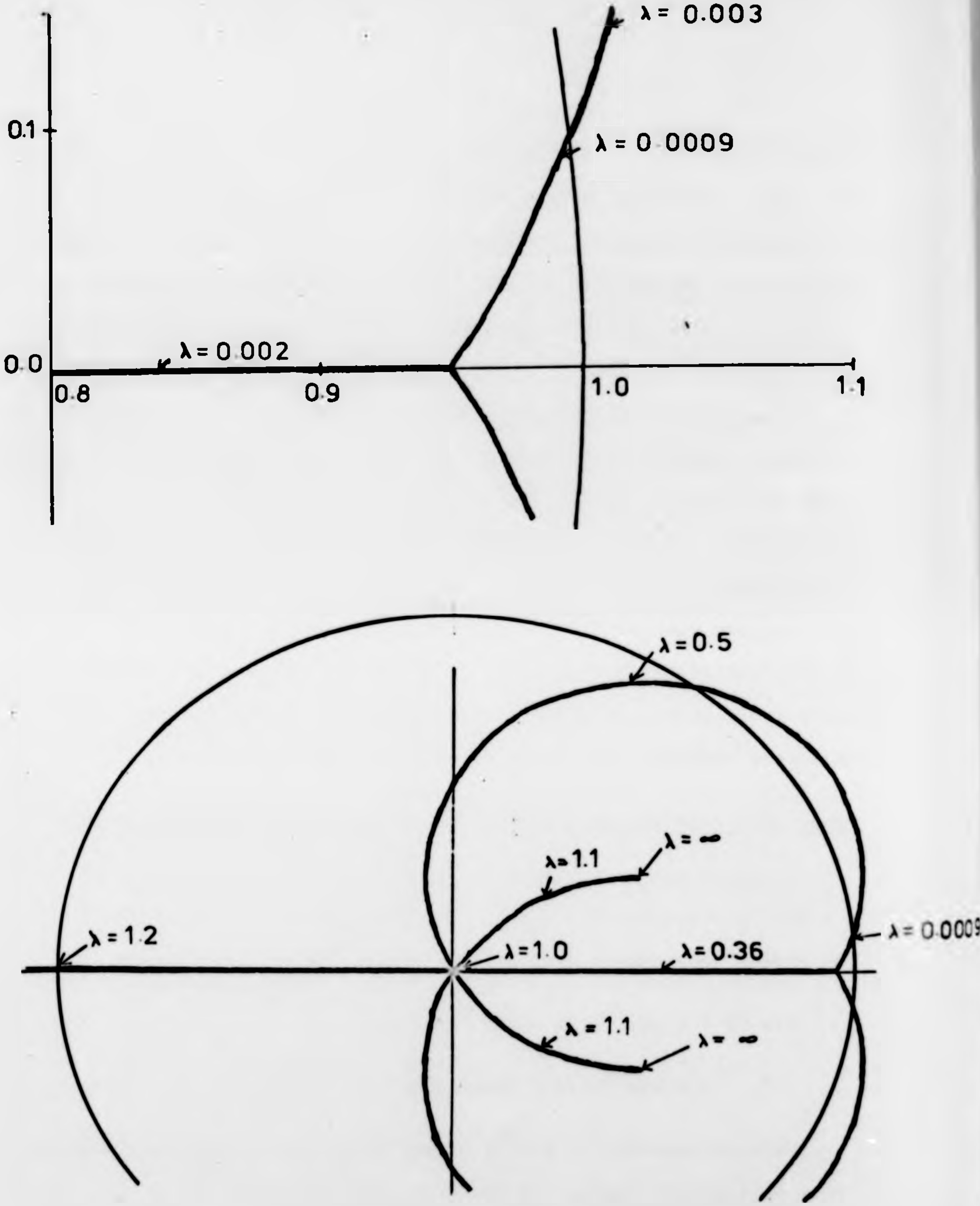


Figure 3.3

Root locus of $1 + A(z^{-1}) - \lambda A(z^{-1})$ with

$1 + A(z^{-1}) = (1 - 0.95z^{-1})^3$

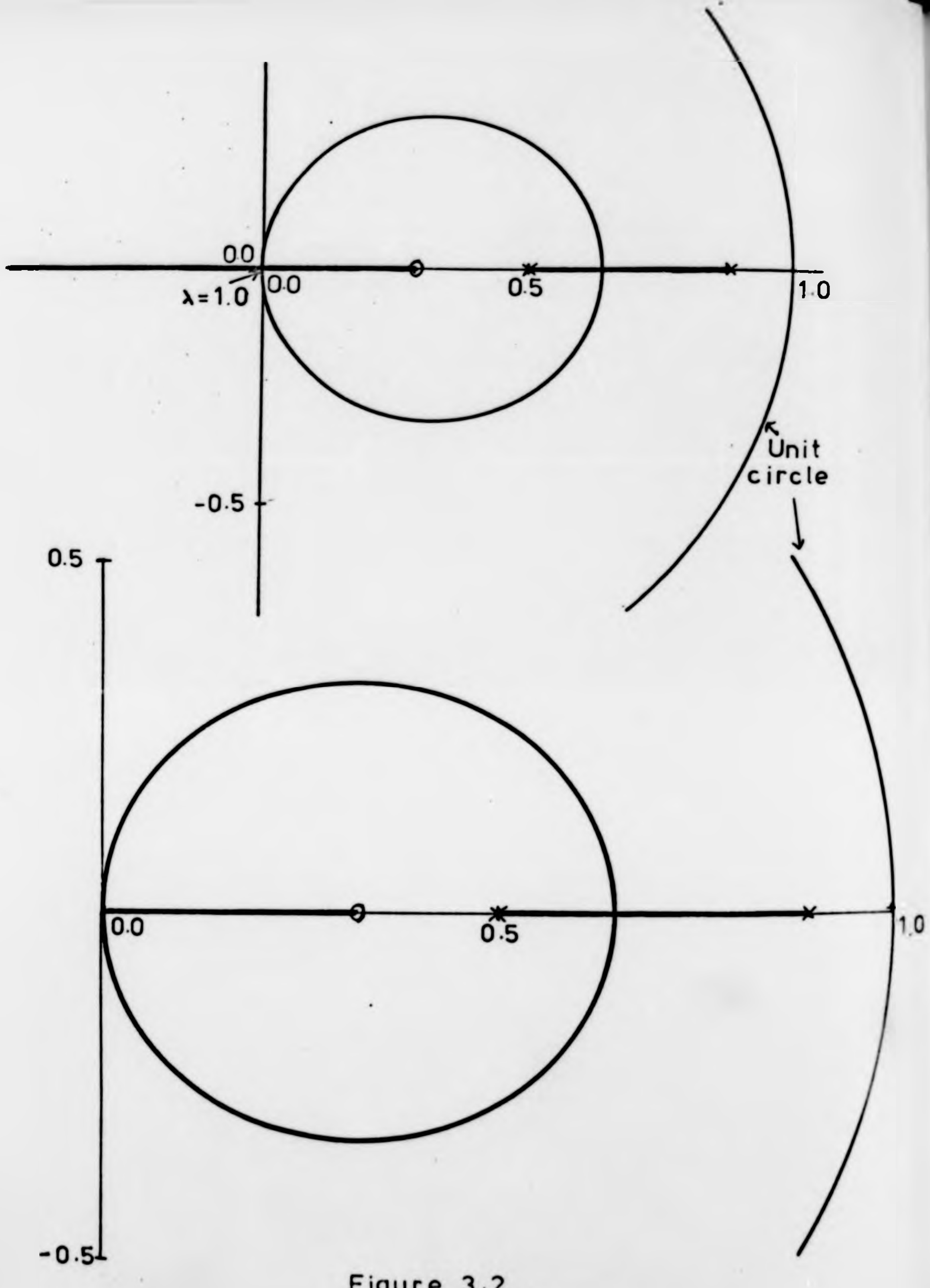


Figure 3.2

Root locus of $1 + A(z^{-1}) - \lambda A(z^{-1})$ for varying λ with $1 + A(z^{-1}) = (1 - 0.9z^{-1})(1 - 0.5z^{-1})$.

the rules for constructing root locus diagrams imply that the one with larger magnitude will initially move along the real axis towards the origin as λ increases. This larger root will decrease until it reaches the smaller root, which has also been moving; the two roots will then become complex as λ is increased. Figure 3.2 shows the root locus when

$$1 + A(z^{-1}) = (1 - 0.9z^{-1})(1 - 0.5z^{-1}) .$$

While the roots of $1 + A(z^{-1}) - \lambda A(z^{-1})$ are complex, they are conplex conjugates and their product is equal to the coefficient of z^{-2} in the polynomial. Hence the magnitude of each of the roots is equal to $((1-\lambda)|a_2|)^{\frac{1}{2}}$ and so decreases with increasing λ while λ is between 0 and 1. Therefore the larger of the two roots of $1 + A(z^{-1}) - \lambda A(z^{-1})$ decreases for increasing λ, and so the roots must stay within the unit circle if they start within it.

Example of conditional stability with three poles

Figure 3.3 shows the root locus diagram for $(1 + A) = (1 - 0.95z^{-1})^3$. It can be seen from this diagram that there are values of λ between 0 and 1 for which the polynomial $1 + (1-\lambda)A(z^{-1})$ has roots outside the unit circle. Therefore this system would be conditionally stable if the minimum variance regulator was used.

The initial behaviour of the roots in figure 3.3 is fairly typical of the behaviour with a group of poles near to each other, in that they tend to set off away from each other as soon as they become complex. Hence conditional stability would be expected if the system being controlled by the minimum variance regulator has a group of poles near to the unit circle in the z domain. It was

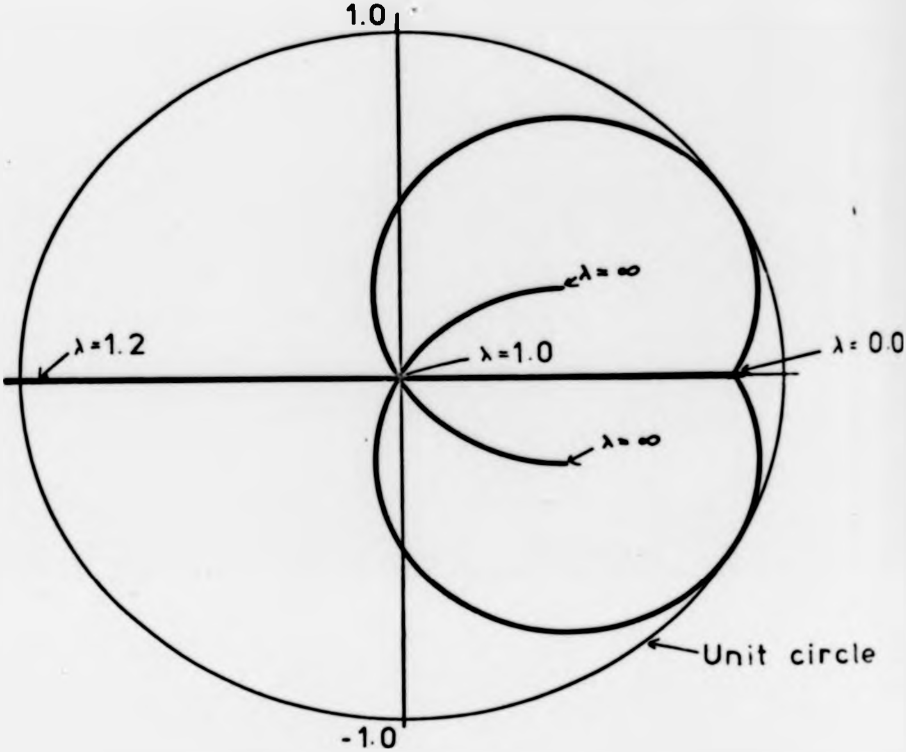


Figure 3.4

Root locus with $1 + A(z^{-1}) = (1 - 0.865z^{-1})^3$

shown in Chapter 2 that this is likely to happen if the system has several poles in its Laplace transfer function and the sample interval is short compared with the pole time constants. For the case of three equal poles the system will be conditionally stable if the time constants are more than 7 sample intervals. Figure 3.4 gives the root locus diagram for the controlled system when the open loop system has 3 poles in its z-transfer function each at 0.865 which corresponds to a time constant of 6.9 sample intervals. It can be seen that this is just on the limit between stability and conditional stability.

3.6.2 Minimum variance control with time delays

Substituting for the control polynomials $1 + F(z^{-1})$ and $G(z^{-1})$ in characteristic equation 3.26 using the control equations 3.21 and 3.22 gives that the characteristic polynomial of the closed loop system with extra gain λ added is

$$T(z^{-1}) = (1+A(z^{-1}))(1+F(z^{-1})) + \lambda z^{-k} B(z^{-1})\, G(z^{-1})$$

$$T(z^{-1}) = (1+A(z^{-1}))\, \ell z\, P(z^{-1})\, B(z^{-1}) - \lambda z^{-k} B(z^{-1})\, \ell\, z\, A'(z^{-1})$$

$$T(z^{-1}) = \ell z\, B(z^{-1})\, ((1+A(z^{-1}))\, P(z^{-1}) - z^{-k}\, \lambda A'(z^{-1}))$$

$$T(z^{-1}) = \ell z\, B(z^{-1})\, (1+z^{-k}\, A'(z^{-1})(1-\lambda)) \qquad (3.31)$$

This has the same form as the equation for the case without time delays (Equation 3.29). The only difference is that $1 + A(z^{-1})$ has been replaced by $(1+A(z^{-1}))P(z^{-1})$ or $(1+z^{-k}A'(z^{-1}))$. Hence the roots of $P(z^{-1})$ have been added to the roots of $(1+A(z^{-1}))$ as starting points for branches in the root locus diagrams which have been used

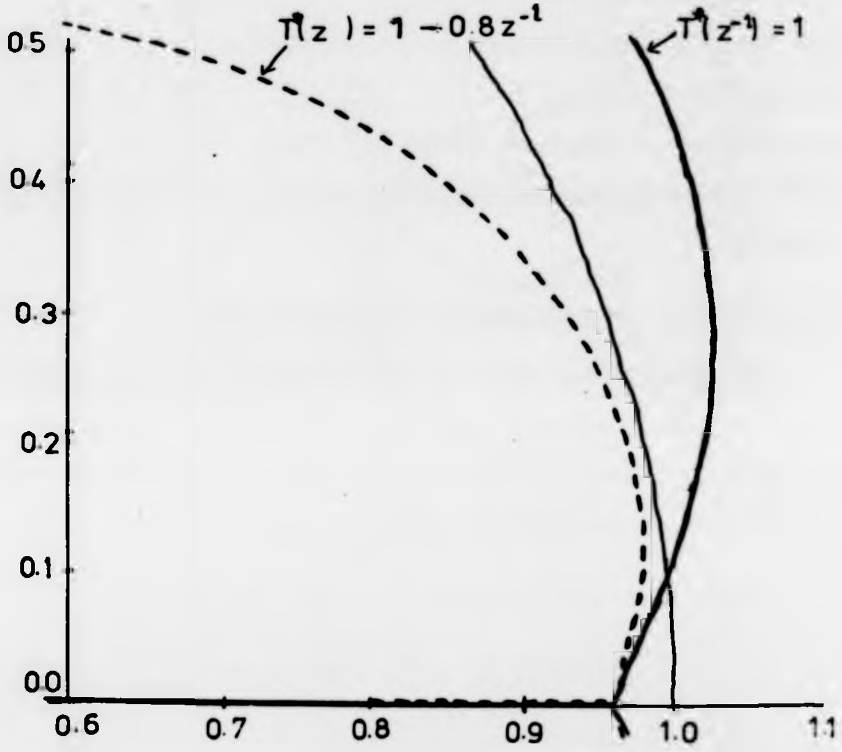


Figure 3.5

Root locii of $1 + A + \lambda(T - 1 + A)$

with $1 + A = (1 - 0.95z^{-1})^3$

for $T^* = 1$

and $T^* = 1 - 0.8z^{-1}$

to investigate conditional stability. This increase in the number of roots will increase the chances of one of the paths going outside the unit circle, so the time delay will increase the likelihood of the closed loop system being conditionally stable. In particular it was shown in section 2.5 that $P(z^{-1})$ can frequently have roots outside the unit circle, and this would automatically result in a conditionally stable system.

3.6.3 Removal of conditional stability by tailoring the response

The conditional stability can be removed by reducing the demands on the control system. For instance, if instead of aiming for the minimum variance control the control was chosen such that $T'(z^{-1}) = (1+F)(T^*)$. This would move the system poles to the roots of T^*. The characteristic equation with a gain λ introduced then becomes

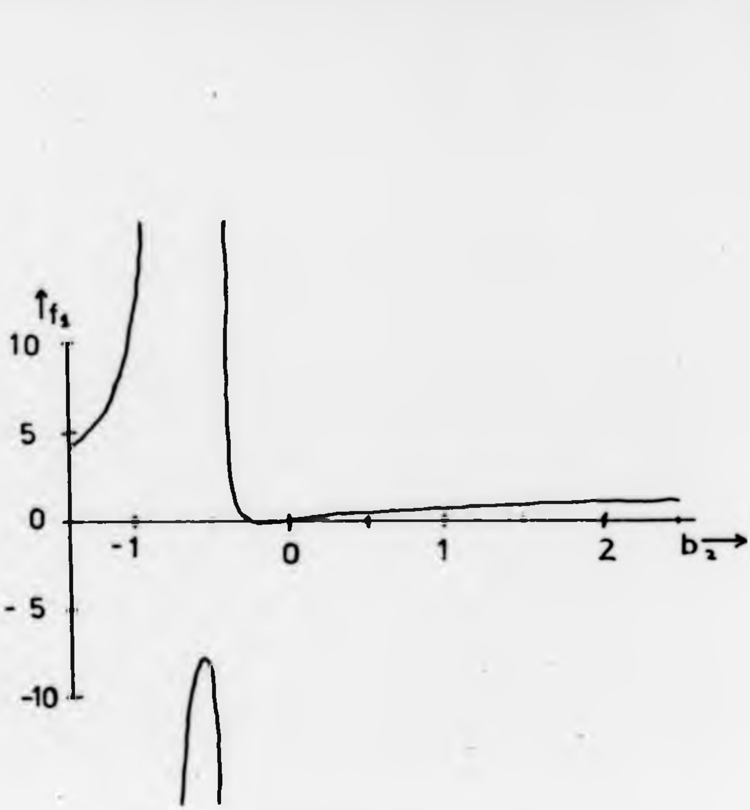
$$T(z^{-1}) = (1+A)(1+F)+\lambda BG$$

$$= (1+A)(1+F)+\lambda(T'-(1+A)(1+F)) \text{ using 3.7}$$

$$= (1+F)\left[1+A+\lambda(T^*-1-A)\right]$$

Clearly if T^* approaches $1+A$ the possibility of conditional stability is removed. The proof in section 3.6.1 implies that if the polynomial $1+A$ has m roots and T^* has just m-2 roots which all are equal to roots of $1+A$ then the possibility of conditional stability is removed. The m-2 roots can be taken out as factors of $1+A+\lambda(T^*-1-A)$ and then the polynomial reduces to $1+A^*-\lambda A^*$ where $A^* = a_1^*z^{-1} + a_2^*z^{-2}$. In section 3.6.1 it was shown that in this case the system is stable. Figure 3.5 shows the effect of adding a $T^* = 1 - 0.8z^{-1}$ to a system with $1+A = (1 - .95z^{-1})^3$ (Fig. 3.3).

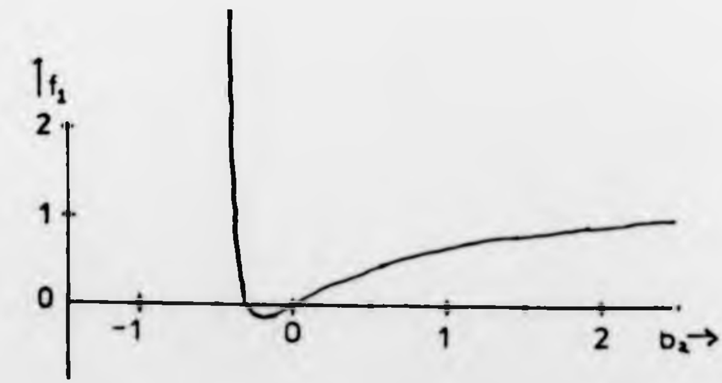
3.6a



<u>Figure 3.6</u>

f_1 as a function of b_2 in the equation 3.33.

3.6b
close-up of 3.6a



It can be seen that the conditional stability has been removed.

3.6.4 Conditional stability with the stable controller of section 3.4

A simple control law which obtains a stable system is obtained by using $T'(z^{-1}) = 1$ in 3.7.

The characteristic polynomial with a variable gain introduced becomes

$$T(z^{-1}) = (1+A)(1+F) + \lambda BG$$

Using 3.7 to substitute for BG gives

$$T(z^{-1}) = (1+A)(1+F) + \lambda(T'(z^{-1}) - (1+A)(1+F))$$

For the simple case with $T'(z^{-1}) = 1$

$$T(z^{-1}) = (1+A)(1+F) - \lambda((1+A)(1+F)-1) \qquad (3.32)$$

This has the same form as 3.29, so again the root locus diagrams can be used to indicate conditional stability. It will be noticed that this control law will be more prone to conditional stability than the simple minimum variance control since the 1+A polynomial has been replaced by (1+A)(1+F), and so there are more branches on the root locus diagrams.

A frequent cause of conditional stability with this control law is for (1+F) to have roots outside the unit circle. Examples can easily be found since if the system has a pole near to one of its zeros the control polynomial 1+F usually has a root outside the unit circle. Figure 3.6 shows the position of the root of 1+F as a function of the system zero with a system given by equation 3.33.

$$(1 - .9z^{-1})(1 - .5z^{-1})\ Y = (z^{-1} + b_2 z^{-1})\ U \qquad (3.33)$$

In this case the control will produce a conditionally stable system for a large range of values for b_2.

The general conclusion that the control will be conditionally stable if the system has a pole near to a zero can be reached by considering the equation 3.7 defining the control. Suppose the system has a zero at 'x' and a pole near to this point where x is not near the origin. Substituting for z^{-1} in 3.7 with $^1/x$ gives

$$(1+A(\tfrac{1}{x}))(1+F(\tfrac{1}{x})) + B(\tfrac{1}{x})\ G(\tfrac{1}{x}) = T'(\tfrac{1}{x}) \qquad (3.34)$$

but x is a zero of the system, hence $B(\frac{1}{x}) = 0$ (3.35)

$$\therefore \quad (1+A(\tfrac{1}{x}))(1+F(\tfrac{1}{x})) = T'(\tfrac{1}{x}) \qquad (3.35)$$

However, the system has a pole near to x, therefore $1+A(\frac{1}{x})$ will be small and so $(1+F(\frac{1}{x}))$ must be large to satisfy equation 3.35. This implies that the coefficients of the polynomial F must be large, hence 1+F has zeros outside the unit circle.

This source of conditional stability can be eliminated by choosing $T'(z^{-1})$ to have factors similar to the roots of $B(z^{-1})$ which have a modulus less than 1. The minimum variance control for non-minimum phase systems suggested by Astrom and mentioned in section 2.3.3 chooses these roots to be roots of $T'(z^{-1})$, and so will not have this extra source of conditional stability. However, it will still have the same conditional stability problems as the minimum variance control for minimum phase systems.

3.7 Lower order regulators

The control laws used in the previous sections are all complex control laws in the sense that there are almost as many variables in the control as there are parameters in the model of the system. Simpler control laws can be constructed using the closed loop equation 3.4, since this holds for any order of controller.

$$((1+A)(1+F) + z^{-k}BG)Y = (1+F)\ e \qquad (3.35)$$

For example, F and G could be chosen so as to make $((1+A)(1+F) + z^{-k}BG)Y$ as close as possible to $(1+F)Y$ by using a least squares fitting procedure. The control would then approximate to the minimum variance control.[6] However, since these control laws are only approximations it is difficult to predict their stability regions.

3.8 Systems with offsets

In this report the offset on a system is considered to be the steady state output of a system with zero input.

Suppose the system is described by equation 2.29 as described in section 2.7.

$$(1+A(z^{-1}))Y = z^{-k}\ B(z^{-1})U + Ce + d \qquad (3.36)$$

This offset in the system can be compensated for by adding an offset to the control law:

$$U = \frac{-GY}{1+F} - \frac{d}{B(1)} \qquad (3.37)$$

The closed loop equation becomes

$$(1+A(z^{-1}))Y + z^{-k}\ \frac{B(z^{-1})G(z^{-1})Y}{(1+F(z^{-1})} = Ce + d - z^{-k}\ \frac{B(z^{-1})\ d}{B(1)} \qquad (3.38)$$

But d is a constant, hence $B(z^{-1})d$ is equal to $d.B(1)$, so the closed loop equation reduces to

$$(1+A(z^{-1}))Y + \frac{B(z^{-1})G(z^{-1})Y}{1+F(z^{-1})} = Ce \qquad (3.39)$$

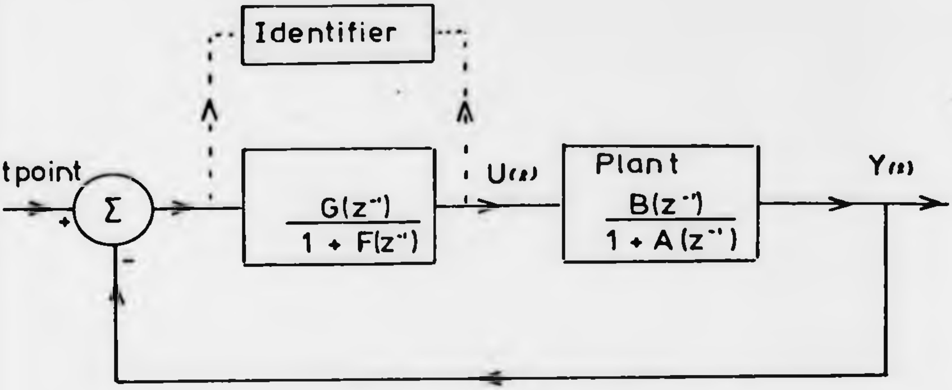


Figure 3.7

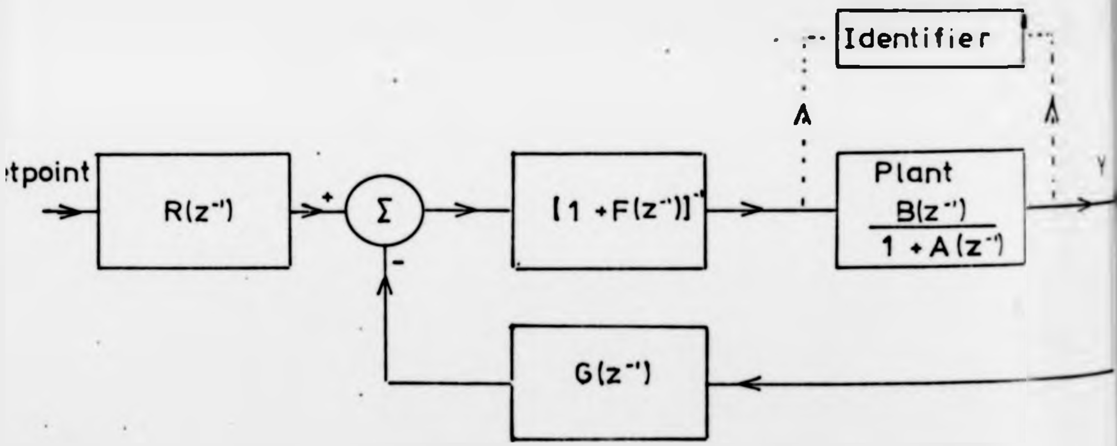


Figure 3.8

Regulator modified to include setpoints.

This is the same as equation 3.3 for the case with no constant offsets. Therefore if the regulator is modified to that described by equation 3.37 the results of the previous sections hold for linear systems which have constant offsets.

This modification to the control can also be used to improve the regulator when there is a constant non zero setpoint for the system.

3.9 Setpoints and setpoint changes

Setpoints can be introduced into the control by subtracting the setpoint from the measured output, and then using this modified output in the identification and control (Figure 3.7). This treats the setpoint changes as extra disturbances on the system, and so usually interferes with the identification algorithm. However, it does retain the proper feedback structure in the regulator. In the special case of there being no setpoint changes the setpoint acts like a constant offset, as described in section 3.8.

An alternative method of introducing setpoint changes is shown in Figure 3.8. Here the identification is done with the actual inputs and outputs and the control law is modified to include the setpoint.

$$(1+F(z^{-1}))U = -G(z^{-1})Y + R(z^{-1})\ \text{setpoint} \qquad (3.40)$$

Combining this with the open loop equation 2.28 gives the closed loop equation 3.41.

$$(1+A(z^{-1}))(1+F(z^{-1}))Y+z^{-k}G(z^{-k})B(z^{-1})Y=z^{-k}B.R(z^{-1})\text{setpoint}+(1+F(z^{-1}))Ce \qquad (3.41)$$

rewriting gives

$$Y = \frac{z^{-k}B(z^{-1})R(z^{-1})\text{setpoint}}{(1+A(z^{-1}))(1+F(z^{-1}))+z^{-k}G(z^{-1})B(z^{-1})} + \frac{(1+F(z^{-1}))C(z^{-1})e}{(1+A(z^{-1}))(1+F(z^{-1}))+z^{-k}G(z^{-1})B(z^{-1})}$$

(3.42)

If the regulator is chosen as in section 3.2 by using equation 3.7, the closed loop equation becomes

$$Y = \frac{z^{-k}B(z^{-1})\ R(z^{-1})\ \text{setpoint}}{T'(z^{-1})} + \frac{(1+F(z^{-1}))C(z^{-1})e}{T(z^{-1})}$$

Therefore the transfer function between the setpoint and the output is defined by:

$$Y = \frac{z^{-k}\ B(z^{-1})\ R(z^{-1})\ \text{setpoint}}{T'(z^{-1})} \qquad (3.43)$$

This leaves the $R(z^{-1})$ polynomial to be chosen to give suitable setpoint following. In particular, the polynomial $R(z^{-1})$ should be chosen such that the gain of the transfer function 3.43 is unity, i.e.

$$R(1) = \frac{T'(1)}{B\ (1)} \qquad (3.44)$$

For example, $R(z^{-1})$ may be chosen as a constant at this value. Examples are given in Chapter 5 of setpoint following using this approach. It was usually satisfactory if the noise was white; unfortunately the results were not always satisfactory if the disturbance on the system was coloured noise.

Poor results could be expected in some cases since, when the disturbance is coloured, the self tuning algorithm depends on the least squares parameter estimates having exactly the right bias, but this bias will be changed by the setpoint changes.

3.10 Behaviour between sample times

With some of the control schemes suggested in this chapter the system will tend to be very oscillatory between the sample times. This subject has been studied by Dr. D. Clarke (University of Oxford).

References for Chapter 3

1 Astrom, K.J. and Wittenmark, B. "On Self Tuning Regulators", Automatica, Vol. 9, 1973, pp. 185-99.

2 Peterka, V., "On Steady State Minimum Variance Control Strategy", Kybernetica, Vol. 8, 1972, pp. 219-31.

3 Astrom, K.J. and Wittenmark, B., "Analysis of a self tuning regulator for non minimum phase systems", I.F.A.C. Symposium on Stochastic Control, Budapest, 1974.

4 Astrom, K.J., "Introduction to Stochastic Control", Academic Press, 1970, pp. 117-127.

5 Jury, E.T., "Sampled Data Control Systems", John Wiley & Sons, 1958.

6 Wellstead, P.E. and Edmunds, J.M., "On Line Process Identification and Regulation", Second I.E.E. Conference on "Trends in On-line Computer Control Systems", Sheffield, April, 1975.

CHAPTER 4 LEAST SQUARES ESTIMATION

The theory of least squares estimation has been published many times and so has not been included in this chapter.[1,2]

Two of the forms given in Chapter 2 for a dynamic model of a system, are suitable for use with a least squares estimator. These are the two models with white noise disturbances given in equations 4.1 and 4.2.

$$(1+A(z^{-1}))Y(t) = z^{-k}B(z^{-1})\,U(t) + e(t) \tag{4.1}$$

and for cases with constant offsets

$$(1+A(z^{-1}))Y(t) = z^{-k}\,B(z^{-1})\,U + e(t) + d \tag{4.2}$$

The least squares estimator chooses the parameter values which minimize the variance of the estimated disturbance e(t). There have been two main approaches to the calculations involved in obtaining these parameter estimates. The first is to collect all the experimental data, and process it all at the same time. The methods produced by this approach are suitable for off line identification. The second main approach to the calculation is to use recursive formulations of the problem. These methods use each set of observations separately and give continually improving parameter estimates throughout the experiment, making them suitable for on line parameter estimation. The recursive methods are not too suitable for off line estimation because they only give the proper estimates asymtotically,* and they generally require more computation than the methods which process all the data at once. P.C. Young[1] described several recursive

* Unless $P_0 = \infty$

methods of calculating the least square estimates. V. Peterka[2] presented an interesting approach to the calculations, which is more difficult to understand but considerably reduces any rounding errors which occur with short word length computers. Both of these papers include methods of exponentially weighting the past data so that more emphasis is placed on recent data. This enables slowly varying parameters to be tracked.

In several of the examples in Chapters 5 and 6 an exponential weighting factor 'ϕ' has been used. 'ϕ' is a number between 0 and 1 and when each new sample is used the weight given to the past data is decreased by this factor.

4.1 Bias in Estimates

The parameter values found by the least squares estimator will only be the same as the system's parameters if the system is governed by an equation of the same form as the model. Hence if the model being used is that given in equation 4.1 the parameter estimates will only avoid bias if the system is described by equation 4.3.

$$(1+A_s(z^{-1}))Y(t) = z^{-k} B_s(z^{-1}) U(t) + e_s(t) \qquad (4.3)$$

Where the polynomials A and A_s have the same order, the polynomials B and B_s have the same order, and $e_s(t)$ is an uncorrelated or white noise sequence. If all these conditions are satisfied the polynomial A will approach A_s and the polynomial B will approach B_s. However if one of these conditions is not satisfied the estimates will not approach the true values. For example, if the disturbance is not white noise, the parameter estimates will be biased, i.e.

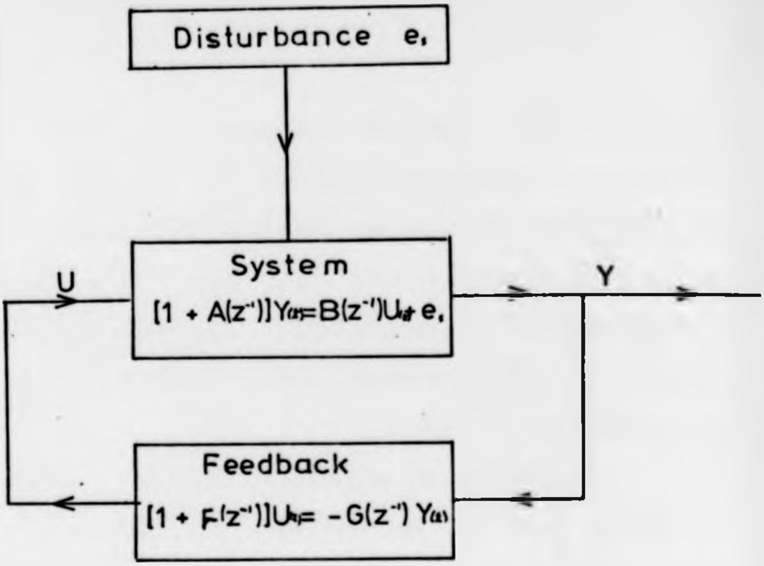


Figure 4.1

Diagram of a regulated system.

if the system is described by 4.4.

$$(1+A_s(z^{-1}))Y(t) = z^{-k} B_s(z^{-1}) U(t) + C(z^{-1}) e_s(t) \quad (4.4)$$

The estimated polynomials A and B will no longer approach the polynomials A_s and B_s describing the system. Proper estimates of the polynomials A_s and B_s can be obtained by using an identification method which identifies the parameters in a model with the same form as the system. One such method is the maximum likelihood method.[3] Unfortunately these better methods are more complicated and require more computation.

However, one of the surprising properties of a self tuning regulator made by combining a least squares estimator with the general pole shifting control law (equation 3.7) is that the final control obtained will be the same as that obtained by using unbiased estimates of the parameters in the pole shifting control law. The proof of this together with the conditions which must be satisfied are given in Chapter 5.2.

4.2 Bias and non-uniqueness due to feedback [5]

When there is feedback on the system the input U and the feedback Y are related by the feedback as well as the dynamics of the system (Figure 4.1). Therefore the estimates of the parameters of the system may be affected by the feedback since the estimator just discovers a relationship between the observed inputs and outputs. It has been shown[4] that if the feedback does not stop the estimates obtained by a least squares estimator from being unique, then the feedback will not cause the estimates to be biased. It was also shown[4] that the estimates will be unique if the order of one of the feedback polynomials F or G is sufficiently large, or if there are sufficient setpoint variations to stop the input and output being simply related by the feedback law. In the notation introduced in Chapters 2 and 3 the estimates will be unique if $nf > nb-1$ or $ng > na-1$.

Consideration of the error function minimized by the least squares estimator shows that the estimates will not be unique when neither polynomial has high enough order. The function minimized by the least squares estimator is the variance of e from equation 4.1.

Rewriting 4.1 gives

$$e(t) = (1+A(z^{-1}))\ Y(t) - B(z^{-1})\ U(t) \qquad (4.5)$$

The feedback law is:

$$(1+F(z^{-1}))\ U(t) = -G(z^{-1})\ Y(t) \qquad (4.6)$$

Multiplying 4.6 by any constant ℓ, and the delay operator z^{-1}, and then adding to 4.5 gives

$$e(t) = (1+A(z^{-1}) + \ell z^{-1}G(z^{-1}))Y(t) - (B(z^{-1}) - \ell z^{-1}(1+F(z^{-1})))\ U(t) \qquad (4.7)$$

Therefore if $nf \leq nb-1$ and $ng \leq na-1$ the error function e(t) will have the same value with parameter estimates $\ell A(z^{-1}) + \ell z^{-1}G(z^{-1})$ and $B(z^{-1}) - \ell z^{-1}(1+F(z^{-1}))$ as with the proper values $A(z^{-1})$ and $B(z^{-1})$. Hence the estimates are not unique.

If there is a delay of k sample periods, equation 4.5 becomes

$$e(t) = (1+A(z^{-1}))\ Y - z^{-k}\ B(z^{-1})\ U(t) \qquad 4.8$$

Now equation 4.6 must be multiplied by $\ell\ z^{-k-1}$ before adding it to equation 4.8, to ensure that no extra powers of z^{-1} are introduced before the U(t). This gives

$$e(t) = (1+A(z^{-1}) + \ell z^{-k-1}\ G(z^{-1}))Y(t) - z^{-k}(B(z^{-1}) - \ell z^{-1}(1+F(z^{-1})))Ut$$

So the condition for uniqueness becomes that either

$$nf > nb-1 \qquad (4.9)$$

$$ng > na-1-k \qquad (4.10)$$

It will be noticed that the control laws given in Chapter 3 have orders which imply that the estimates will not be unique if there is no delay.

References for Chapter 4

1. Young, P.C. "Applying Parameter Estimation to Dynamic Systems", Control Engineering, October-November, 1969

2 Peterka, V. "A square root filter for real time multivariate Regression", Institute of Information Theory and Automation, Czechoslovak Academy of Sciences, Prague

3 Reference 8 in Chapter 2

4 Wellstead, P.E. and Edmunds, J.M. "Least squares identification of closed-loop systems", Int. J. Control, 1975, Vol.21, No.4, pp. 689-699

5 Gustavsson I., Ljung L and Soderstrom T, 1974 Report 7401 Division of Automatic Control, Lund Institute of Technology

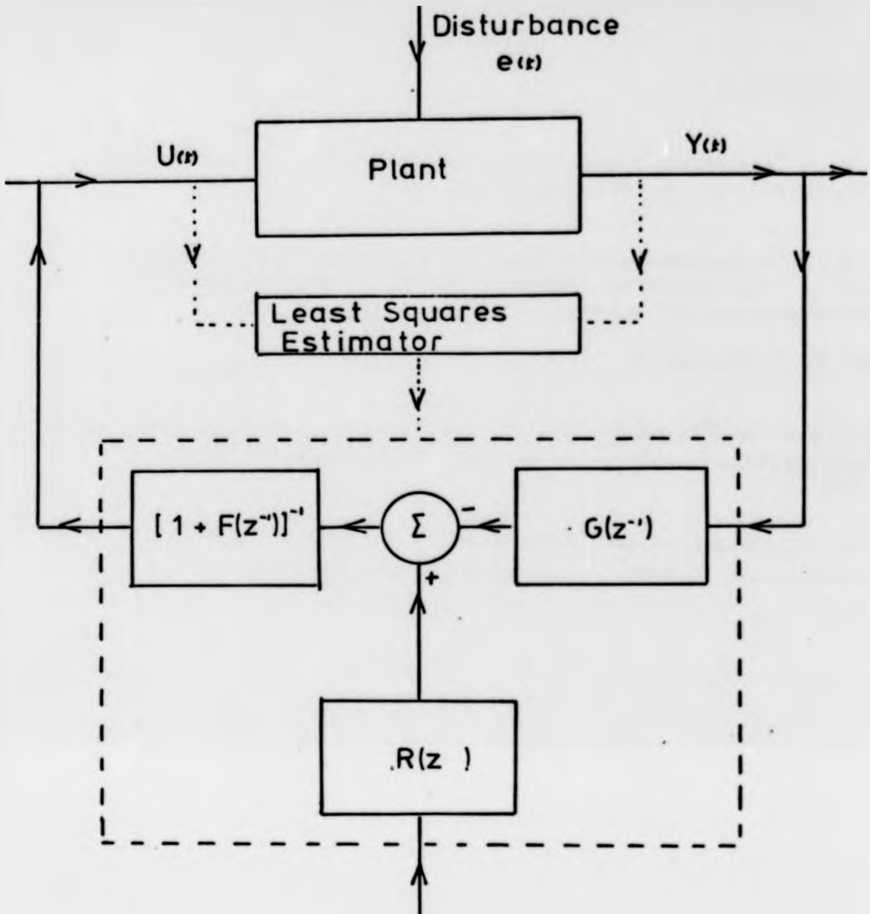


Figure 5.3

A form for a self tuning pole shifting controller.

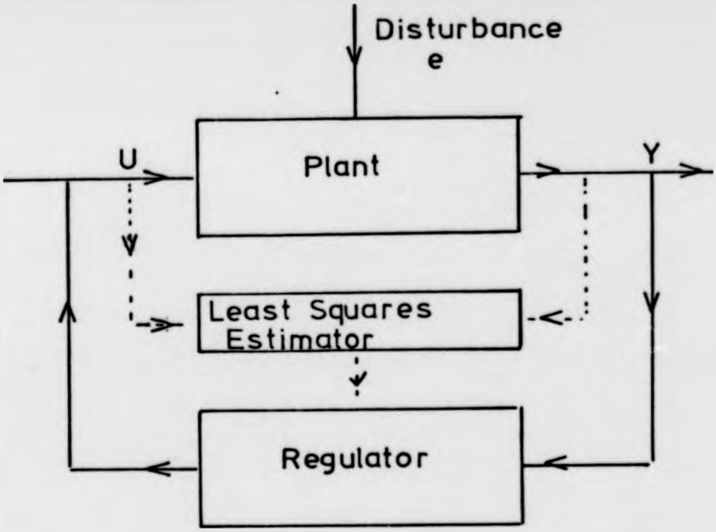


Figure 5.1

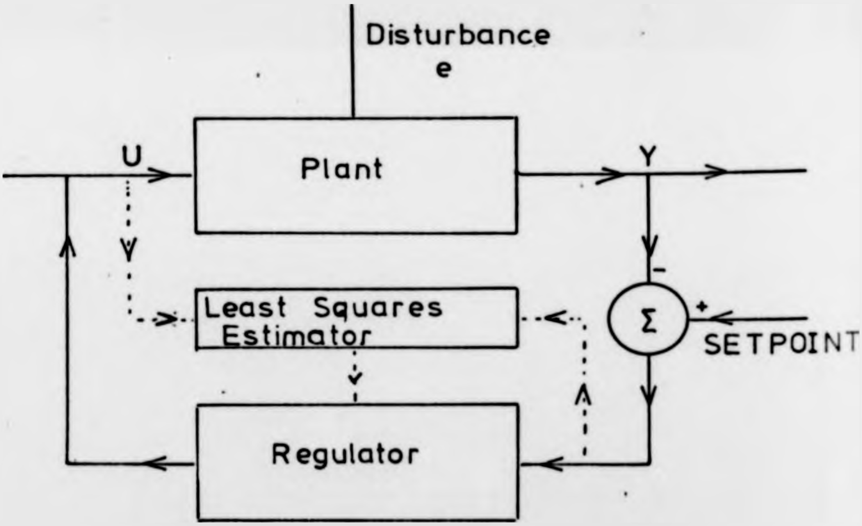


Figure 5.2

Introduction of setpoints as extra disturbances.

CHAPTER 5

5.1 Some Self tuning Controllers and Regulators

Self tuning regulators can be constructed by combining one of the regulators from Chapter 3 with the least squares estimator of Chapter 4. The recursive least squares estimator is used to obtain a model of the system, and regulator parameters are then calculated by assuming that the model of the system is correct (Figure 5.1). The use of a recursive least squares estimator means that the model of the system is continuously changing, and so the control law also changes at every sample period. Section 5.2 gives the asymptotic properties of these self tuning regulators, and section 5.3 describes their initial behaviour.

There are several ways of introducing setpoints for these regulators, turning them into self tuning controllers. The simplest way is to use the error in the output instead of the output for the self tuning regulator (Figure 5.2). This way treats any setpoint changes as disturbances on the system, and so setpoint changes will make the model found by the estimator less accurate. However, if setpoint changes are very infrequent, so that the main problem is just one of regulation, this approach can be used so long as provision is made in the estimator for D.C. offsets or bias. A second way of introducing setpoints is to leave the estimator to estimate the system parameters, and to change the control law to add in a function of the setpoint (Figure 5.3). This function is calculated from the model to give the correct D.C. gain. This method is satisfactory if there is only a white noise disturbance on the system. However, when there is coloured noise on the system the model parameters will be biased leading to poor setpoint following, and large errors with constant setpoints, since there is no direct feedback to compensate for errors due to non zero setpoints.

5.1.1 Using a minimum variance control law in a self tuning regulator

The minimum variance control law was given by

$$(1+F(z^{-1}))U(t) = G(z^{-1})\ Y(t) \tag{3.1}$$

where $(1+F(z^{-1})) = \ell z\ B'(z^{-1}) = \ell z\ P(z^{-1})B(z^{-1})$ (3.22)

and $G(z^{-1}) = -\ \ell z\ A'(z^{-1})$ (3.21)

with $(1+z^{-k}\ A'(z^{-1})) = (1+A(z^{-1}))\ P(z^{-1})$ (2.26)

if the system is described by

$$(1+A(z^{-1}))Y = z^{-k}\ B(z^{-1})\ U + e \tag{2.24}$$

A self tuning regulator which asymptotically approaches this regulator can be made by using a least squares estimator to estimate the coefficients of A and B in equation (2.24) above, calculating the polynomial $P(z^{-1})$ using equation (2.26) and using the resulting A and B polynomials to define the control law. This is very similar to the self tuning regulator described by Astrom in which he directly estimated the coefficients of the polynomials A' and B' by forming a model of the form (5.1).

$$(1+z^{-k}A'(z^{-1}))Y = z^{-1}B'(z^{-1})\,U + e \qquad (5.1)$$

and then using the same control law. The main difference is that he fixed the first estimated coefficient of the B' polynomial so that any lack of uniqueness due to the feedback was removed. In order to add setpoint changes by method indicated by Figure 5.3 the control law is changed to that given by equation 3.40.

$$(1+F(z^{-1}))\,U(t) = -G(z^{-1})\,Y(t) + R(z^{-1})\text{setpoint }(t) \qquad (3.40)$$

The transfer function is then given by equation 3.43.

$$Y = \frac{z^{-k}\,B(z^{-1})\,R(z^{-1})}{T'(z^{-1})}\text{ setpoint} \qquad (3.43)$$

substituting for $T'(z^{-1})$ using 3.23

giving $$Y = \frac{z^{-k}\,B(z^{-1})}{{}^{1}/b_1\, zB(z^{-1})}\,R(z^{-1})\text{ setpoint}$$

$$= b_1 z^{-k-1} R(z^{-1})\text{ setpoint}$$

Therefore the compensator polynomial $R(z^{-1})$ should just be $^{1}/b_1$ to give the correct setpoint following when the estimator gives unbiased estimates of the system parameters. The effects of non zero means can be overcome by changing the estimated model to:

$$(1+A(z^{-1}))\,Y(t) = z^{-k}\,B\,(z^{-1})\,U(t) + d + e(t) \qquad (5.2)$$
(c.f. 3.36)

and then making the control law

$$U = \frac{-A'(z^{-1})}{B'(z^{-1})}\,Y(t) - \frac{d}{B(1)} \qquad (5.3)$$
(c.f. 3.37)

or $$U = \frac{-A'(z^{-1})\,Y(t) - d}{B'(z^{-1})} \qquad (5.4)$$

The main constraints on the self tuning regulator made in this way is that the system must have all its zeros within the unit circle (see section 3.3). It was also shown in section 3.6 that this regulator can be conditionally stable if the system has many poles near the unit circle, or if there is a time delay.

5.1.2 Using the tailored control law suggested in 3.6.3 in a self tuning regulator

For this control the characteristic polynomial $T'(z^{-1})$ is $(1+F(z^{-1}))T^*$ and F and G are chosen so that

$$T'(z^{-1}) = (1+A)(1+F) + BG \tag{5.5}$$

$$\therefore \quad (1+F)T^* = (1+A)(1+F) + BG \tag{5.6}$$

$$\therefore \quad (1+F) = \ell z B \tag{5.7}$$

and $\ell z T^* = \ell z(1+A) + G$

$$\therefore \quad G = \ell z(T^*-1-A) \tag{5.8}$$

Hence the control law can be very easily calculated from the parameter estimates. When setpoints are added as in section 5.1.1 the transfer function is given by

$$Y = \frac{B(z^{-1})R(z^{-1})}{(1+F(z^{-1}))T^*} \quad \text{setpoint}$$

$$= b_1 z^{-1} \frac{R(z^{-1})}{T^*(z^{-1})} \quad \text{setpoint}$$

If the order of $R(z^{-1})$ is chosen as zero then the required gain for the setpoint $= r_o = \frac{T^*(1)}{b_1}$

The main constraints on the self tuning regulator made in this way is that the system must have all its zeros within the unit circle (see section 3.3). It was also shown in section 3.6 that this regulator can be conditionally stable if the system has many poles near the unit circle, or if there is a time delay.

5.1.2 Using the tailored control law suggested in 3.6.3 in a self tuning regulator

For this control the characteristic polynomial $T'(z^{-1})$ is $(1+F(z^{-1}))T^*$ and F and G are chosen so that

$$T'(z^{-1}) = (1+A)(1+F) + BG \tag{5.5}$$

$$\therefore \quad (1+F)T^* = (1+A)(1+F) + BG \tag{5.6}$$

$$\therefore \quad (1+F) = \ell z B \tag{5.7}$$

$$\text{and} \quad \ell z T^* = \ell z(1+A) + G$$

$$\therefore \quad G = \ell z(T^*-1-A) \tag{5.8}$$

Hence the control law can be very easily calculated from the parameter estimates. When setpoints are added as in section 5.1.1 the transfer function is given by

$$Y = \frac{B(z^{-1})R(z^{-1})}{(1+F(z^{-1}))T^*} \quad \text{setpoint}$$

$$= b_1 z^{-1} \frac{R(z^{-1})}{T^*(z^{-1})} \quad \text{setpoint}$$

If the order of $R(z^{-1})$ is chosen as zero then the required gain for the setpoint $= r_o = \dfrac{T^*(1)}{b_1}$

When the output has a non zero mean the model for the system becomes

$$(1 + A(z^{-1}))\ Y(t) = B(z^{-1})\ U(t) + d + e(t) \tag{5.9}$$

The required response is given by

$$T^*(z^{-1})\ Y(t) = e(t) \tag{5.10}$$

$\therefore$ The control for this response is obtained by combining the two equations (5.9 and 5.10) above

$$e(t) = (1 + A(z^{-1}))\ Y(t) - B(z^{-1})\ U(t) - d = T^*(z^{-1})\ Y(t)$$

$$\therefore \quad U(t) = \frac{(1+A(z^{-1}) - T^*)\ Y(t) - d}{B(z^{-1})}$$

5.1.3 Using the general pole shifting law

A self tuning regulator can be constructed by combining a least squares estimator with the general pole shifting law of section 3.4. For this the desired closed loop pole positions are chosen, and the denominator of closed loop transfer function constructed. The chosen control law is found by substituting the required $T'(z^{-1})$ in equation 3.7 and solving for the polynomials F and G. The current estimates of the system polynomials $A(z^{-1})$ and $B(z^{-1})$ are used in the calculation of the control law. There are several approaches which can be used to construct the required control law. One simple approach is to consider equation 3.7 as a set of linear simultaneous equations and solve these with the least squares estimator which has to be programmed to obtain the system estimates. However this approach is rather inefficient since the recursive least squares estimator has to be applied once for each of the simultaneous equations. One of the best ways of solving for F and G is to use one of the standard methods for solving sets of linear equations such as that proposed by

This approach decreases the computation being done, and improves the accuracy of the solution at the expense of using more core storage and using a more complicated program.

Another interesting approach is to notice that closed loop equation 3.4 implies that F and G should satisfy equation 5.12 below, which is the same as equation 3.7 with both sides multiplied by Y(t).

$$((1+A(z^{-1}))(1+F(z^{-1})) + z^{-k} B(z^{-1}) G(z^{-1})) Y(t) = T'(z^{-1})Y(t) \tag{5.12}$$

The Y(t) can be introduced since if this equation is satisfied for all the Y(t) which occur the closed loop equation 3.4 is the same as would be obtained by solving the equation 3.7. The control estimates can then be updated using the recursive least squares estimator just once at each sample time. To do this equation 5.12 is rewritten as

$$F(z^{-1})((1+A(z^{-1})) Y(t)) + G(z^{-1})(z^{-k} B(z^{-1}) Y(t))$$
$$= (T'(z^{-1}) - 1 - A(z^{-1})) Y(t) \tag{5.13}$$

or $$F(z^{-1}) X_1(t) + G(z^{-1}) X_2(t) = X_3(t) \tag{5.14}$$

The functions $X_1(t)$, $X_2(t)$ and $X_3(t)$ can all be calculated from the observed outputs Y(t), the required polynomial $T'(z^{-1})$ and the estimated polynomials $A(z^{-1})$ and $B(z^{-1})$. Equation 5.14 is of the correct form for the least squares estimator. This method has the advantage that it does not require a special part of the program to solve for G and F but instead can use the recursive least squares estimator used for the system estimates. A second advantage of this method is that a lower order regulator may be used as suggested in section 3.7. However this approach will fail if there are limits imposed on the allowed control signal, and the control signal frequently reaches these limits since then the $X_1(t)$, $X_2(t)$ and $X_3(t)$ do not vary sufficiently to calculate the correct control.

Another similar approach to calculating the control is to use a noise sequence instead of Y(t) in 5.12. This avoids the problem of saturation but retains the same simple calculation. The equation for the calculation of F and G then is:

$$((1+A(z^{-1}))(1+F(z^{-1})) + z^{-k} B(z^{-1})G(z^{-1}))\ e_1(t) = T'(z^{-1})\ e_1(t) \qquad (5.15)$$

or

$$F(z^{-1})((1+A(z^{-1}))e_1(t))+G(z^{-1})(z^{-k}B(z^{-1})e_1(t))=(T'(z^{-1})-1-A(z^{-1}))e_1(t) \qquad (5.16)$$

or

$$F(z^{-1})\ X^1(t) + G(z^{-1})\ X^2(t) = X^3(t) \qquad (5.17)$$

Non zero datum levels can be dealt with, for the general pole shifting law, in a similar way to that used in the previous section.

Suppose the model of the system is given by 5.9

$$(1+A(z^{-1}))\ Y(t) = B(z^{-1})\ U(t) + d + e(t) \qquad (5.9)$$

and the control law is altered to

$$U = \frac{-G(z^{-1})\ Y(t) + d'}{1 + F(z^{-1})}$$

Then in order to remove the effects of the offset d, d' should be chosen so that

$$\frac{B(z^{-1})\, d'}{1+F(z^{-1})} = d$$

since then the closed loop equation will be the same as if there was no offset.

$$\therefore \quad d' = \frac{\left(1 + F(1)\right) d}{B(1)}$$

It will be noticed that in the two previous sections the formula given above reduces to $d' = d$.

Setpoints can be added by the two methods suggested in the previous sections. It was shown in section 3.9 that the required gain

$$r_o = \frac{T'(1)}{B(1)}$$

In the examples in section 5.3 using the general pole shifting self tuning regulator the method of calculating the $F(z^{-1})$ and $G(z^{-1})$ polynomials was to consider equation 3.7 as a set of linear simultaneous equations and solve them using a standard method. The subroutine used for the solution of the equations was subroutine SIMQ from the Control System Centre's Library of Fortran subroutines.

5.2 Asymptotic properties of self tuning regulators

There are three possible types of behaviour self tuning regulators can have after they have been applied for a long time. They can settle to a constant stable controller, or they can give reasonable control but not settle to a constant controller, or they can produce an unstable closed loop system. Simulations of the self tuning regulators combining least squares estimation with pole shifting demonstrated all three of these types of behaviour.

The examples of instability occurred mainly with non linear systems; however, even with linear systems, there were some cases where a bad choice of initial condition led to rapid growth of the output. In particular when using the minimum variance control (section 5.1.1) if the initial estimate for the first parameter in the $B(z^{-1})$ polynomial is chosen as zero, or chosen such that it becomes exactly zero at some early stage the feedback immediately becomes infinite, and the system cannot recover. Similar sorts of numerical problems can be found for the general pole shifting law.

However, if these numerical problems are avoided, and the system being controlled is linear, the self tuning regulator using the general pole shifting law should give stable control, since if the parameter estimates settle to the correct values the control will be stable, and if the output becomes large the estimates will approach the correct values and so stabilize the system. (A similar argument was used by Astrom in describing the stabilizing properties of his self tuning regulator for minimum phase systems.)

Figure 5.4 a

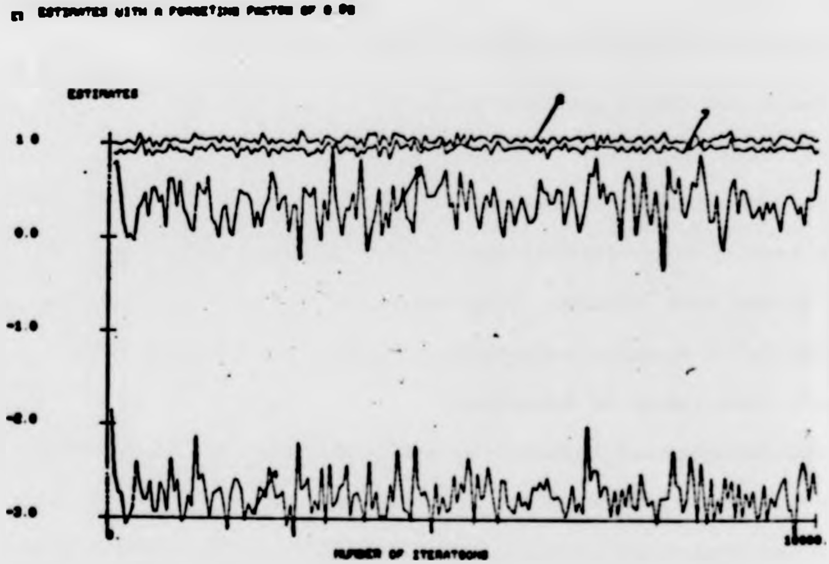
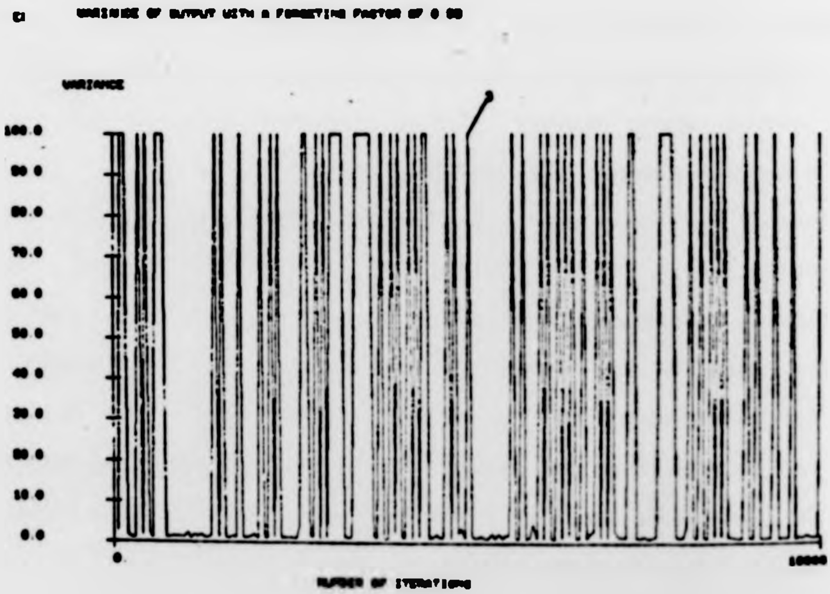


Figure 5.4 b



5.2.1 Convergence of parameter estimates

A self tuning regulator using the general pole shifting contol law is of the correct form to be analysed by a method proposed by L. Ljung and B. Wittenmark.[1] They associate a non linear differential equation with the behaviour of the estimates, and then assume that the estimates will settle if the equation is stable. However, since this method has a non linear differential equation to be solved, in any particular case it may be easier to simulate the system and regulator, and observe the behaviour of the estimates. Also the results obtained by using the non linear differential equation may be conservative in some cases since the solution of the non linear equation corresponds to the expected (in the statistical sense) trajectory of the parameters.

In their report[1] they show that a self tuning regulator which used the minimum variance control law (Chapter 3) can fail to converge. In particular the estimates can fail to converge if a weighted least squares estimator is used. They give an example (number 6.2 in their report) which should fail to converge. In this example the system is described by:

$$(1 - 1.6z^{-1} + 0.75z^{-2})\, Y(t) = (z^{-1} + z^{-2} + 0.9z^{-3})\, U(t) + (1 - 1.5z^{-1} + 0.75z^{-2})\, e(t) \qquad (5.18)$$

This is controlled by the regulator proposed by Astrom and Wittenmark.[2] Using a weighted least squares they obtained a picture similar to Fig. 5.4(a) which was also obtained using a weighted least squares. Fig. 5.4(b) shows the variance that would have

Figure 5.5a

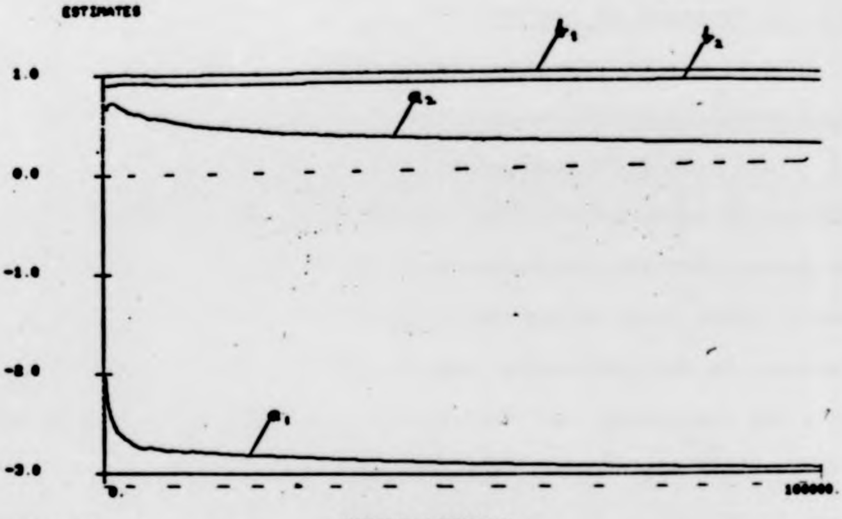


Figure 5.5 b

been obtained if the tuning of the control had been obtained if the tuning of the control had been stopped after any particular sample. It can be seen that the self tuning controller does sometimes choose a control which would be unstable. (These variances were obtained by a method proposed by Astrom,[3] using a Fortran program from his book, "Introduction to Stochastic Systems".) However, if an unweighted least squares estimator is used the estimates (Fig. 5.5a) do not oscillate in the same way. This implies that the theory is conservative in this case since the theory predicts instability with the unweighted least squares. The very slow convergence indicates that the estimate may be near to instability. The correct parameter estimates are $a_1 = -3.1$, $-a_2 = 0$, $b_1 = 1$, and $b_2 = 0.9$.

Figure 5.5b shows that the control is reasonably good while the estimates are still converging. The noise signal was chosen such that the minimum variance regulator would give an output variance of unity.

During the simulation of the self tuning regulators using an unweighted least squares estimator, and the pole shifting control law no cases were found of continually wandering control values when controlling linear systems. However, there may be systems for which the control does not settle down.

5.2.2 Non-Unique Parameter Estimates due to feedback

While the control law is continually changing there is no simple relationship between the input and output due to the feedback. However, when the control law settles down the input and output will be simply related by the regulator, and so the lack of uniqueness in the parameter estimates as described in Chapter 4.2 should appear. This would imply that the estimates may be prone to wandering. The uncertainty in the estimate does not lead to uncertainty in the control law when using the pole shifting control, since the variations in the estimates which can occur due to feedback do not in this case alter the control parameters.

Proof: that the uncertainty in parameter estimates does not lead to uncertainty in control.

Suppose the control is chosen by

$$(1+A(z^{-1}))(1+F(z^{-1})) + B(z^{-1})G(z^{-1}) = T(z^{-1}) \tag{5.19}$$

where $A(z^{-1})$ and $B(z^{-1})$ are the estimated parameters.

$$\text{Let } A^*(z^{-1}) = A(z^{-1}) + \ell z^{-1}G \tag{5.20}$$

$$\text{and } B^*(z^{-1}) = B(z^{-1}) - \ell z^{-1}(1+F) \tag{5.21}$$

With a fixed control law A^* and B^* would be equally good estimates of the system's parameters as A and B for any constant ℓ. These estimates would give a control chosen by:

$$(1+A^*(z^{-1}))(1+F^*(z^{-1})) + B^*(z^{-1})G^*(z^{-1}) = T(z^{-1}) \tag{5.22}$$

using 5.20 and 5.21 gives

$$(1+A(z^{-1}))(1+F^*(z^{-1})) + B(z^{-1})G^*(z^{-1}) + (1+F^*(z^{-1}))(\ell z^{-1}G(z^{-1}))$$

$$- G^*(z^{-1})\ \ell z^{-1}(1+F(z^{-1})) = T(z^{-1}) \tag{5.23}$$

$$\therefore \quad (1+A(z^{-1}))(1+F^*(z^{-1})) + B(z^{-1})\ G^*(z^{-1}) = T(z^{-1}) \tag{5.24}$$

But for a given $A(z^{-1})$ and $B(z^{-1})$ equation 5.24 gives a unique solution for F^* and G^*. Therefore $F^* = F$ and $G^* = G$. So all the possible estimates A^* and B^* give the same control law.

Figure 5.7a

Parameter Estimates

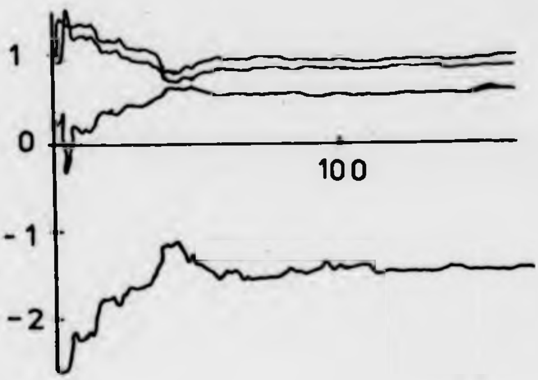


Figure 5.7b

Control Parameters

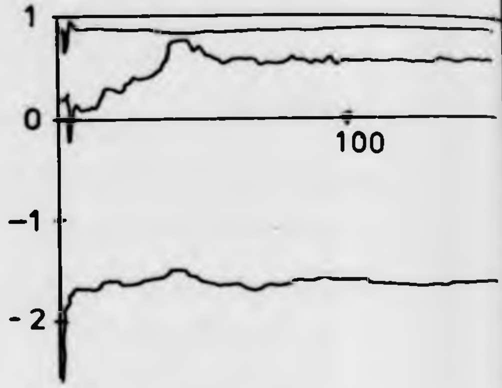
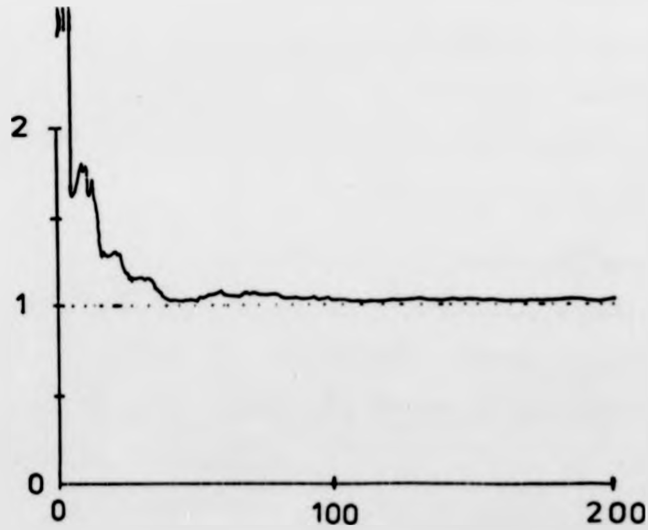


Figure 5.7c

Expected variance of control



Number of Iterations

Figure 5.6a

Parameter estimates with a weighted least squares with PHI=0.95

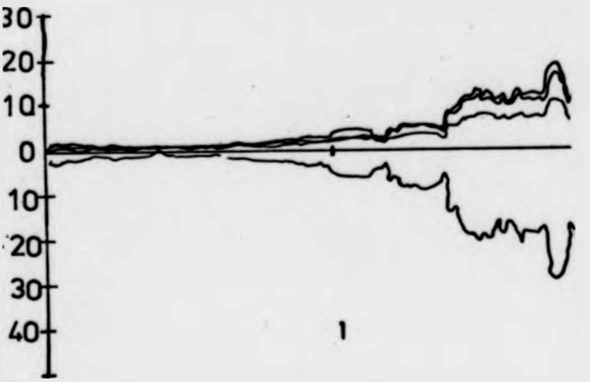


Figure 5.6b

Control parameters using a weighted least squares with PHI=0.95

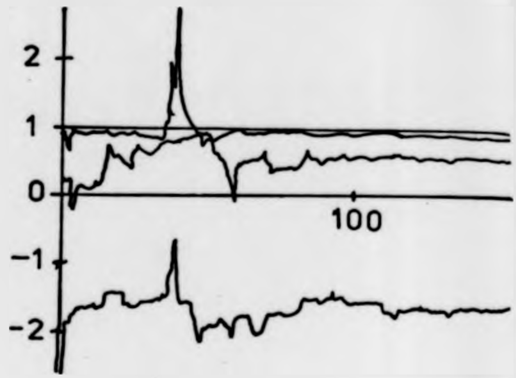


Figure 5.6c

Expected variance as a function of time using a weighted least squares PHI=0.95

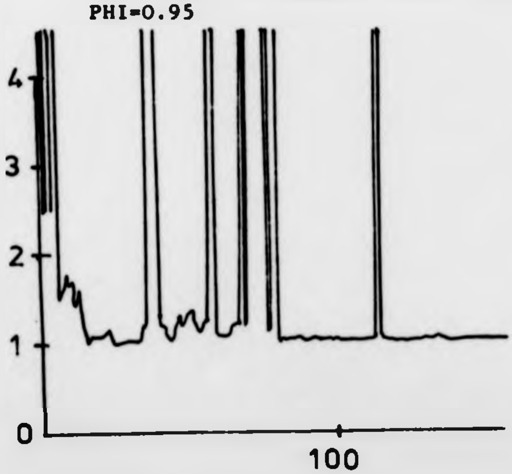
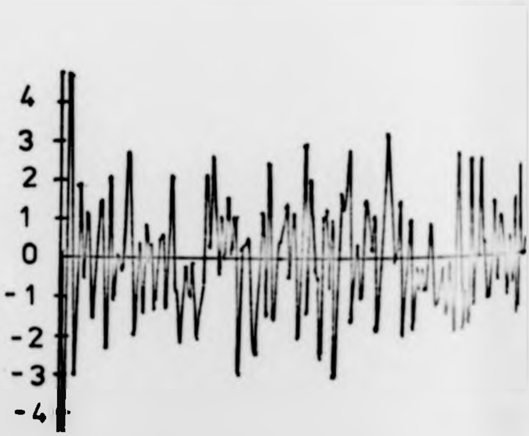


Figure 5.6d

Output of the system



This wandering of parameter estimates occurs most readily when a weighted least squares estimator is used. Figure 5.6(a) shows wandering estimates when a self tuning regulator using a weighted least squares estimator and the minimum variance control law was made to control a system described by:

$$(1 - 1.6z^{-1} + 0.75z^{-2})\ Y(t) = (z^{-1} + 0.9z^{-2})\ U(t) + e(t) \qquad (5.25)$$

The exponential weighting factor ϕ was equal to 0.95. Figure 5.6(b) shows that the control law remained fairly constant despite the large variations in the parameter estimates. Figures 5.6(c,d) show the corresponding expected variances and actual controlled output. The results with an unweighted least squares estimator are shown in Figures 5.7 (a,b,c). It can be seen that these are much more stable, but they still exhibit some wandering together in the initial stages of the run.

5.2.3 The Asymptotic effect of coloured noise *

When the disturbance on the system is coloured noise, the parameter estimates obtained by a least squares estimator will be biased. However, if this estimator is combined with the general pole shifting control law 5.28, to form a self tuning regulator, the final control will move the poles to the required places. Therefore the final control will be the same as would have been obtained using the proper system parameters (equation 5.26) together with the pole shifting law for systems with coloured noise (equation 5.29).

* Note that if the noise is coloured and there are set point variations this following theorem does not hold.

Theorem 5.1 *

If a system can be described by:

$$(1+A_s(z^{-1}))\ Y = B_s(z^{-1})\ U + C(z^{-1})\ e_s \qquad (5.26)$$

and a least squares estimator is used to fit coefficients of a model:

$$(1 + A(z^{-1}))\ Y = B(z^{-1})\ U + e \qquad (5.27)$$

with the order of A equal to the order of A_s and the order of B equal to the order of B_s. Then if the control at each stage is calculated using the latest estimates of A and B in an equation:

$$(1 + A(z^{-1}))(1 + F(z^{-1})) + B(z^{-1})\ G(z^{-1}) = T'(z^{-1}) \qquad (5.28)$$

with $\quad nt \leq na + nb - 1 - nc$

and the parameter estimates converge, the final control will be such that:

$$(1 + A_s(z^{-1}))(1 + F(z^{-1})) + B_s(z^{-1})\ G(z^{-1}) = C(z^{-1})\ T'(z^{-1}) \qquad (5.29)$$

The proof of this theorem is similar to that used by Astrom and Wittenmark for their self tuning regulators.[2,4] The result can simply be extended to cover the cases they dealt with.

Outline of proof

1) Use the properties of the estimator to say that the estimated disturbance 'e(t)' is not correlated with the previous 'na' outputs or 'nb' inputs.

2) Express Y(t) and U(t) in terms of 'e(t)', and then find another sequence 'w(t)' which is simply related to Y, U and e.

3) Show that $e(t+\tau)$ and $w(t)$ are uncorrelated for $na + nb - 1 \geq \tau \geq 1$.

* Note this proof implicitly covers the case with time delays since there are no restrictions on the zeros of B_s or B.

4) Use the Yule-Walker equations to show that $e(t+\tau)$ and $w(t)$ are uncorrelated for $\tau \geq 1$.

5) Show that $e(t+\tau)$ and $e(t)$ are uncorrelated for $\tau \geq 1$.

6) Deduce that $e(t) \rightarrow e_s(t)$, i.e. the residuals from the least squares estimator asymptotically are the same as the white noise driving the system.

The theorem is then obvious from the equation which relates $e(t)$ and $e_s(t)$.

<u>Proof</u>

1) The least squares estimator chooses the coefficients of the polynomials $A(z^{-1})$ and $B(z^{-1})$ so that the residuals '$e(t)$' are not correlated with any of the previous na outputs or nb inputs, since otherwise the variance of the residuals could be reduced.

Therefore $E(e(t+\tau) \,.\, Y(t)) = 0 \qquad 1 \leq \tau \leq na \qquad (5.30)$

$$E(e(t+\tau) \,.\, U(t)) = 0 \qquad 1 \leq \tau \leq nb \tag{5.31}$$

2) The control law is:

$$(1+F(z^{-1}))\, U(t) = -G(z^{-1})\, Y(t) \tag{5.32}$$

Combining this control with the model of the system (e.g. 5.27) gives

$$((1+A(z^{-1}))(1+F(z^{-1})) + B(z^{-1})G(z^{-1}))\, Y(t) = (1+F(z^{-1}))\, e(t) \tag{5.33}$$

Using 5.28 with 5.33 gives

$$(T'(z^{-1}))\, Y(t) = (1+F(z^{-1}))e(t) \tag{5.34}$$

Therefore

$$Y(t) = \frac{(1+F(z^{-1}))\, e(t)}{T'(z^{-1})} \tag{5.35}$$

Substituting for Y in 5.32 using 5.35 gives

$$(1+F(z^{-1}))\ U(t) = \frac{-G(z^{-1})(1+F(z^{-1}))e(t)}{T'(z^{-1})} \tag{5.36}$$

$$\therefore \quad U(t) = \frac{-G(z^{-1})\ e(t)}{T'(z^{-1})} \tag{5.37}$$

Let $w(t) = \dfrac{e(t)}{T'(z^{-1})}$ (5.38)

Then $Y(t) = (1+F(z^{-1}))\ w(t)$ (5.39)

and $U(t) = -G(z^{-1})\ w(t)$ (5.40)

3) Substituting for Y(t) in 5.30 using 5.39 gives:

$$E(e(t+\tau)\ .\ (1+F(z^{-1}))\ w(t)) = 0 \qquad 1 \leq \tau \leq na$$

$$\therefore \quad E(e(t+\tau)\ .\ w(t)+f_1 e(t+\tau).w(t-1)+ \ldots +f_{nf} e(t+\tau).w(t-nf))=0$$

$$\text{for } 1 \leq \tau \leq na \tag{5.41}$$

But expectation is a linear operator, therefore

$$E(e(t+\tau).w(t))+f_1 E(e(t+\tau).w(t-1))+ \ldots +f_{nf} E(e(t+\tau).w(t-nf))=0$$

$$\text{for } 1 \leq \tau \leq na \tag{5.42}$$

Similarly substituting for U(t) in 5.31 using 5.40 gives

$$g_o E(e(t+\tau).w(t))+g_1 E(e(t+\tau).w(t-1))+\ldots+g_{ng} E(e(t+\tau).w(t-ng))=0$$

$$\text{for } 1 \leq \tau \leq nb \tag{5.43}$$

The sets of equations 5.42 and 5.41 can be conveniently expressed as a matrix equation 5.44:

$$\begin{bmatrix} 1 & f_1 & f_2 & \cdots & f_{nf} & & & 0 \\ 0 & 1 & f_1 & f_2 & \cdots & f_{nf} & & \\ & & \ddots & \ddots & \ddots & & \ddots & \\ 0 & & & 1 & f_1 & f_2 & \cdots & f_{nf} \\ g_0 & g_1 & g_2 & \cdots & g_{ng} & & & 0 \\ 0 & g_0 & g_1 & \cdots & & g_{ng} & & \\ & & \ddots & \ddots & \ddots & & \ddots & \\ 0 & & & g_0 & g_1 & g_2 & \cdots & g_{ng} \end{bmatrix} \begin{bmatrix} E[e(t+1).w(t)] \\ E[e(t+2).w(t)] \\ \vdots \\ E[e(t+nf).w(t)] \\ E[e(t+1+nf).w(t)] \\ \vdots \\ \\ E[e(t+s).w(t)] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

(5.44)

Where $s = nf + na = nb + ng = na + nb - 1$ (5.45)

If G and (1+F) have no common factors the only solution of equation 5.44 is that

$$E(e(t+\tau).w(t)) = 0 \qquad 1 \leq \tau \leq s \tag{5.46}$$

<u>4) To extend to $\tau > s$</u>

Substituting for $e(t+\tau)$ in the expectation of $e(t+\tau).w(t)$ using 5.38 gives

$$E(e(t+\tau).w(t)) = E(T'(z^{-1})\ w(t).w(t)) \tag{5.47}$$

Let $E(w(t+\tau).w(t)) = p(\tau)$ (5.48)

Then $E(e(t+\tau).w(t)) = T'(z^{-1})\ p(\tau)$ (5.49)

Equation 5.46 gives that

$$T'(z^{-1})\ p(\tau) = 0 \qquad 1 \leq \tau \leq s \tag{5.50}$$

Since the sequence $e_s(t)$ is known to be a white noise sequence the sequence $w(t)$ will now be expressed in terms of e_s. Substituting for Y and U using 5.39 and 5.40 in equation 5.26 (which governs the system) gives:

$$(1+A_s(z^{-1}))(1+F(z^{-1}))w(t)+B_s(z^{-1})G(z^{-1})w(t) = C(z^{-1})\ e_s(t) \tag{5.51}$$

So $w(t)$ is a mixed autoregressive moving average process; using the notation used by Box and Jenkins[5] it is an ARMA(s,nc) process. In section 3.4.2 of reference 5 it is shown that for such a process the autocorrelations $p(t)$ are related by:

$$p(\tau) = \phi_1 p(\tau-1)+\ldots+\phi_s p(\tau-s) \quad \text{for} \quad \tau \geq nc+1 \tag{5.52}$$

(This is equation 3.4.3 in ref. 5)

$$\text{Let} \quad \phi(z^{-1}) = z^{-1}\phi_1 + z^{-2}\phi_2+\ldots+z^{-s}\phi_s \tag{5.53}$$

Equation 5.52 becomes

$$p(\tau) = \phi(z^{-1})\ p(\tau) \quad \text{for} \quad \tau \geq nc + 1 \tag{5.54}$$

Substituting in 5.49 gives

$$E(e(t+\tau).w(t)) = T'(z^{-1})\ \phi(z^{-1})\ p(\tau) \tag{5.55}$$

$$\tau \geq nc + nt + 1$$

The condition $\tau \geq nc + nt + 1$ is required since the multiplying by $T'(z^{-1})$ introduces delayed versions of $p(\tau)$ and the expansion 5.54 has to be valid for all of them. Equation 5.55 can now be used to extend the range of τ for which equation 5.49 holds if

$nc + nt \leq s$ since 5.55 gives

$$E(e(t+s+1).w(t)) = T'(z^{-1})\phi(z^{-1})p(s+1) \qquad (5.56)$$

$$\therefore \quad E(e(t+s+1).w(t)) = \phi(z^{-1})T'(z^{-1})p(s+1) \qquad (5.57)$$

But $\phi(z^{-1})$ only has powers of z^{-1} between 1 and s

$$\therefore \quad E(e(t+s+1).w(t)) = \phi(z^{-1}).0 = 0 \qquad (5.58)$$

(by using the equation 5.50)

Rewriting gives $T'(z^{-1})p(s+1) = 0$

The range of valid τ can now be extended by repeating this argument. Hence

$$E(e(t+\tau).w(t)) = 0 \qquad 1 \leq \tau \qquad (5.59)$$

5) Autocorrelation of e(t)

The autocorrelation coefficients of e(t) are given by

$$E(e(t+\tau).e(t)) = E(e(t+\tau).T'(z^{-1})w(t)) \qquad (5.60)$$

$$\therefore \quad E(e(t+\tau).e(t)) = 0 \qquad \tau \geq 1 \qquad (5.61)$$

(using 5.59)

6) Therefore e(t) is a white noise sequence but it is a function of $e_s(t)$ since substituting for w(t) in 5.51 gives:

$$\frac{(1+A_s(z^{-1}))(1+F(z^{-1}))e(t)}{T'(z^{-1})} + \frac{B_s(z^{-1})G(z^{-1})e(t)}{T'(z^{-1})} = C(z^{-1})\, e_s(t) \qquad (5.62)$$

Therefore $e(t) = e_s(t)$. This can be seen by rewriting Equation 5.62 as:

$$e_s(t) = \frac{((1+A_s(z^{-1}))(1+F(z^{-1})) + B_s(z^{-1})G(z^{-1}))\; e(t)}{C(z^{-1})T'(z^{-1})} \qquad (5.63)$$

But $C(z^{-1})T(z^{-1})$ has all its zeros inside the unit circle. Therefore

$$e_s(t) = k_o e(t) + k_1 e(t-\Delta t)+\ldots+k_n e(t-n\Delta t)+\ldots \tag{5.64}$$

where $\quad k_n = E(e_s(t).e(t-n\Delta t)) \qquad (5.65)$

since e(t) is a white noise sequence.

Similarly

$$e(t) = \frac{C(z^{-1})T(z^{-1})e_s(t)}{((1+A_s(z^{-1}))(1+F(z^{-1}))+B_s(z^{-1})G(z^{-1}))}$$

and e(t) is bounded

$$\therefore \quad e(t) = k_o' e_s(t) +\ldots+k_n' e_s(t-n\Delta t)+\ldots \tag{5.66}$$

substituting for $e(t-n\Delta t)$ in 5.65 using 5.66 gives

$$k_n = E(e_s(t).[k_o' e_s(t-n\Delta t)+k_1' e_s(t-(n-1)\Delta t)\ldots])$$

$$= k_o' E(e_s(t).e_s(t-n\Delta t))+k_1' E(e_s(t).e_s(t-(n-1)\Delta t))\ldots \tag{5.67}$$

But e_s is a white noise sequence $\therefore$

$$k_n = 0 \text{ if } n > 0$$

and $\quad k_o = k_o' = 1$ *

$$\therefore \quad e_s(t) = e(t)$$

Therefore 5.62 becomes

$$((1+A_s(z^{-1}))(1+F(z^{-1})) + B_s(z^{-1})G(z^{-1}))e_s(t) = C(z^{-1})T(z^{-1})e_s(t) \tag{5.68}$$

* k_o can be seen to be 1 by considering the first term in the expansion.

But $e_s(t)$ is a white noise sequence so the only solution to 5.68 is

$$(1+A_s(z^{-1}))(1+F(z^{-1}))+B_s(z^{-1})G(z^{-1}) = C(z^{-1})T(z^{-1}) \qquad (5.69)$$

Extension of Theorem 5.1 to some cases where $T(z^{-1})$ is chosen to be a function of the parameter estimates.

In the two self tuning regulators presented by Astrom and Wittenmark,[2,4] the control chosen is equivalent to choosing $T(z^{-1})$ as a function of some of the parameter estimates. The proof of Theorem 5.1 can be extended to such self tuning regulators if the parameter estimates used in calculating $T(z^{-1})$ are unbiased. There are two interesting cases when some of the parameter estimates are unbiased. These are when one of the factors of $T(z^{-1})$ is chosen to be equal to a factor of the denominator polynomial $1+A(z^{-1})$ or the numerator $B(z^{-1})$. In these cases equation 5.28 implies that this common factor is also a factor of $G(z^{-1})$ or $1+F(z^{-1})$ respectively. Equation 5.69 then implies the common factor is a factor of $1+A_s(z^{-1})$ or $B_s(z^{-1})$ respectively, and so the estimate of such a common factor is unbiased despite any coloured noise. This extends the theorem 5.1 to the self tuning regulator using the minimum variance control law[2] since this chooses $T(z^{-1})$ to be the product of the factors of $B(z^{-1})$, and so the estimates of the zeros of the system are unbiased. In reference 4 Astrom and Wittenmark prove that the asymptotic result of a self tuning regulator using the constrained minimum variance control law will be the proper constrained minimum variance control in two cases. The first is when $B_0(x) = 1$ which means that all the zeros of the system are inside the unit circle. The polynomial $T(z^{-1})$ is then equal to the product of the factors of $B_s(z^{-1})$, which is the same as above. The second case is when the $C(z^{-1}) = 1$, when they say that the parameter estimates will be unbiased.

5.3 The initial properties of self tuning regulators

The initial properties of a self tuning regulator composed of a least squares estimator, and a linear feedback law depend on the initial estimates of the parameters and inverse covariance matrix. The behaviour of the system can also be altered by imposing limits on the control signal during these initial steps. In this section several digital and analogue simulations have been used to illustrate how the initial estimates affect the initial behaviour. Examples have also been constructed to illustrate how the settling time for these estimates is related to the type of disturbance on the system.

5.3.1 Initial Parameter Estimates

Usually something will be known about the dynamic properties of a system before a self tuning regulator is applied to it. This knowledge should be used when choosing the initial estimates for the parameters to ensure that estimates have the same order of magnitude as the system parameters. Also care should be taken not to start with estimates which immediately upset the control algorithm being used. For example with the minimum variance control law (section 5.1.1) the initial estimate for b_1 (the first term in the numerator) should not be zero since this will lead to an infinite initial control signal. With the general pole shifting control law (section 5.1.3) the initial estimates of the numerator and denominator of the system's z-transfer function should not have a common factor since it was shown in section 3.6.4 that this gives a badly defined control law.

5.3.1.1 Estimates of denominator of the z-transfer function

It was shown in section 2 that if the system's Laplace transfer function has a pole which has a time constant several times longer than the sample period the corresponding pole in the z-transfer function is near to 1. Hence if a system has 'm' poles a reasonable initial guess for the denominator of the z-transfer function would be $(1+A(z^{-1}))=(1-z^{-1})^m$.

5.3.1.2 Estimates of the numerator of the z-transfer function

The values for the numerator are much more difficult to estimate, since they depend on the system's gain and the time constants as well as the sample period. However, the required order of magnitude can be estimated in several ways. When an estimate of the system's Laplace transfer function is available, an estimate of the system's 'z' transfer function can be obtained by converting the estimate of the Laplace transfer function (Chapter 2, section 3). When a step response has been obtained for the system, the order of magnitude of the coefficients can be calculated from the output at a time equal to the time delay plus one sample period after the step was applied. For a system with no time delay the output for a unit step input after one sample period is equal to the coefficient of z^{-1} in the numerator of the 'z' transfer function. The other coefficients would usually have the same order of magnitude as this first one.

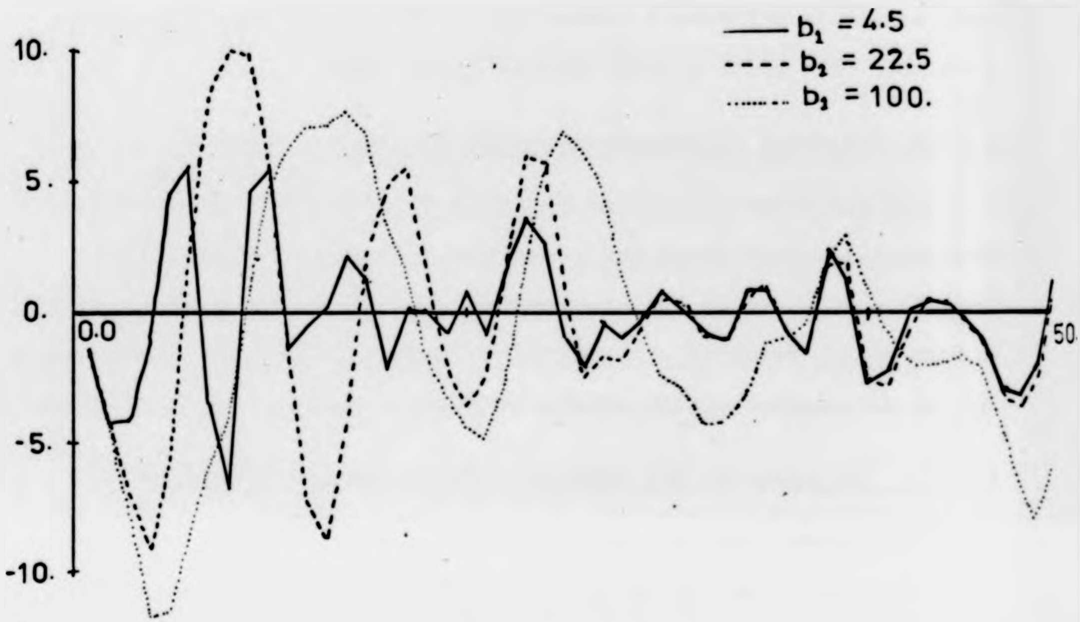


Figure 5.8c

Regulated system output using three different initial estimates for the B polynomial. The system is described by equation 5.70. The initial estimates are

$$B(z^{-1}) = 4.5\ z^{-1} + 4.5\ z^{-2}$$

$$B(z^{-1}) = 22.5\ z^{-1} + 22.5\ z^{-2}$$

$$B(z^{-1}) = 100\ z^{-1} + 100\ z^{-2}$$

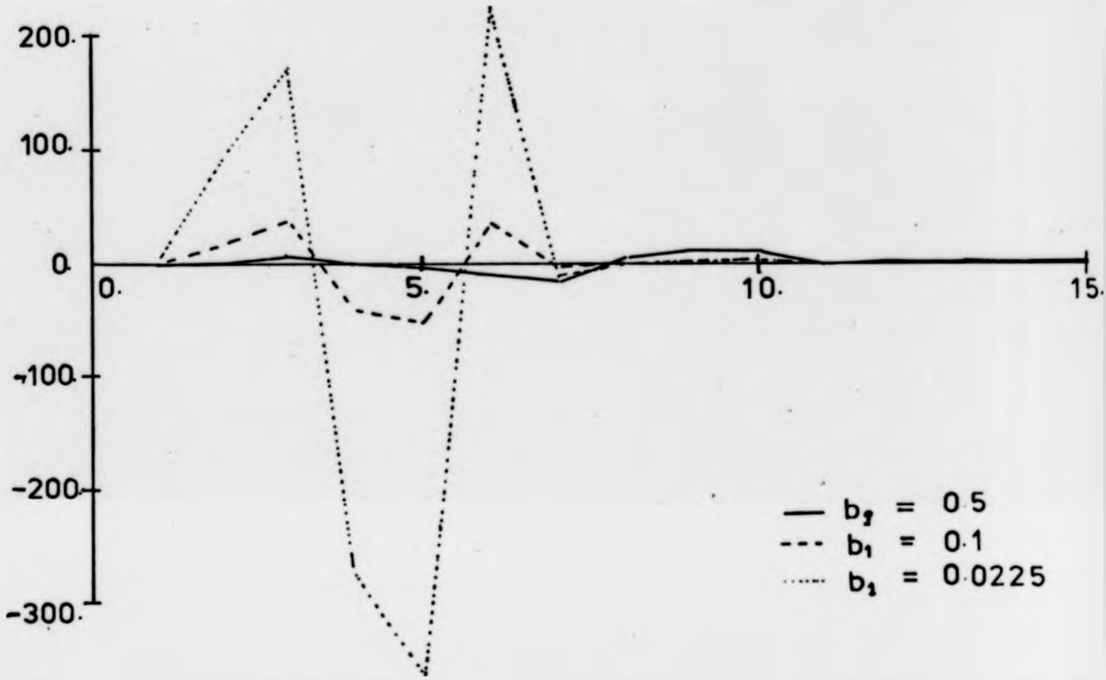


Figure 5.8 b

Regulated system output using three different initial estimates for the B polynomial. The system is described by equation 5.70. The initial estimates are

$$B(z^{-1}) = 0.5\ z^{-1} + 0.5\ z^{-2}$$

$$B(z^{-1}) = 0.1\ z^{-1} + 0.1\ z^{-2}$$

$$B(z^{-1}) = 0.0225\ z^{-1} + 0.0225\ z^{-2}$$

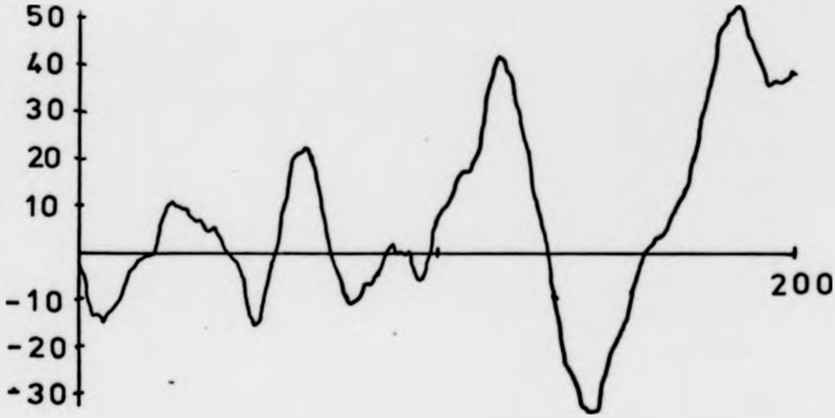


Figure 5.8 a

Uncontrolled output of a system described by

$$(1 - 1.8\, z^{-1} + 0.81\, z^{-2})\, Y(t) = (z^{-1} + 2z^{-2})\, U(t) + e_s(t)$$

This system with this noise sequence $e_s(t)$ was used for examples 5.1, 5.2, 5.3 and 5.6.

Fortunately the choice of the initial estimate of the numerator parameters does not seem to be very critical as long as the initial estimates are too large rather than too small. Large initial estimates mean that the system's gain has been overestimated, and so result in lower control gains initially. Smaller controller gains result in smaller control signals, and so less initial disturbance to the system.

Example 5.1*

To illustrate the effect of different choices of estimates for the numerator parameters.

For this example the system being controlled was simulated using an equation

$$(1-1.8z^{-1}+0.81z^{-2})Y(t) = (z^{-1}+2z^{-2})U(t)+e_s(t) \qquad (5.70)$$

Where $e_s(t)$ was a psuedo random white noise sequence. The output with no control applied is shown in Figure 5.8a. This system was then controlled by the self tuning regulator using the general pole shifting law with $T(z^{-1}) = 1$. Figure 5.8b shows the controlled outputs when the initial numerator estimates were too small. It can be seen that the disturbance to the system becomes progressively larger as the initial parameter estimates are decreased. The initial estimates for the three curves shown were

$$(1-2z^{-1}+z^{-2})Y(t) = (0.0225z^{-1}+0.0225z^{-2})U(t) + e(t)$$

$$(1-2z^{-1}+z^{-2})Y(t) = (0.1z^{-1}+0.1z^{-2})U(t) + e(t)$$

and $$(1-2z^{-1}+z^{-2})Y(t) = (0.5z^{-1}+0.5z^{-2})U(t) + e(t)$$

Figure 5.8c shows the controlled outputs when the initial estimates were too large. It can be seen that in this case the large initial

* In the examples in this section unless stated otherwise the chosen $T'(z^{-1})$ polynomial was 1, an unweighted least squares estimator was used and the initial estimate of the inverse covariance matrix was 100 I.

estimates did not produce excessive disturbances on the system. Comparison with the uncontrolled output (Fig. 5.8a) show that all these regulators considerably reduced the noise on the system. However the large estimates can be seen to decrease the rate of convergence. The initial estimates used were:

$$(1-2z^{-1}+z^{-2})Y(t) = (100.z^{-1}+100.z^{-2})U(t) + e(t)$$

$$(1-2z^{-1}+z^{-2})Y(t) = (22.5z^{-1}+22.5z^{-2})U(t) + e(t)$$

and $$(1-2z^{-1}+z^{-2})Y(t) = (4.5z^{-1}+4.5z^{-1})U(t) + e(t)$$

5.1.2 Initial choice of Inverse covariance matrix

In theory the recursive least squares estimators will asymptotically give the same results as an ordinary least squares estimator if the initial estimate for the inverse covariance matrix has large values on the diagonal and all other elements zero. Usually large values on the diagonal gave good results but occasionally very large values produced difficulties. These difficulties may arise because very large values in the estimated inverse covariance matrix imply that there is virtually nothing known about the system, and so the parameter estimates are initially very free. Hence the estimates may wander close to one of the situations mentioned in section 5.3.1 where the calculation of the control law is badly conditioned. When the initial values are small the initial rate of adaptation is slow since the small values imply that a lot of information is known about the system.

Example 5.2

To illustrate the effects of choosing very large values on the diagonal of the inverse covariance matrix.

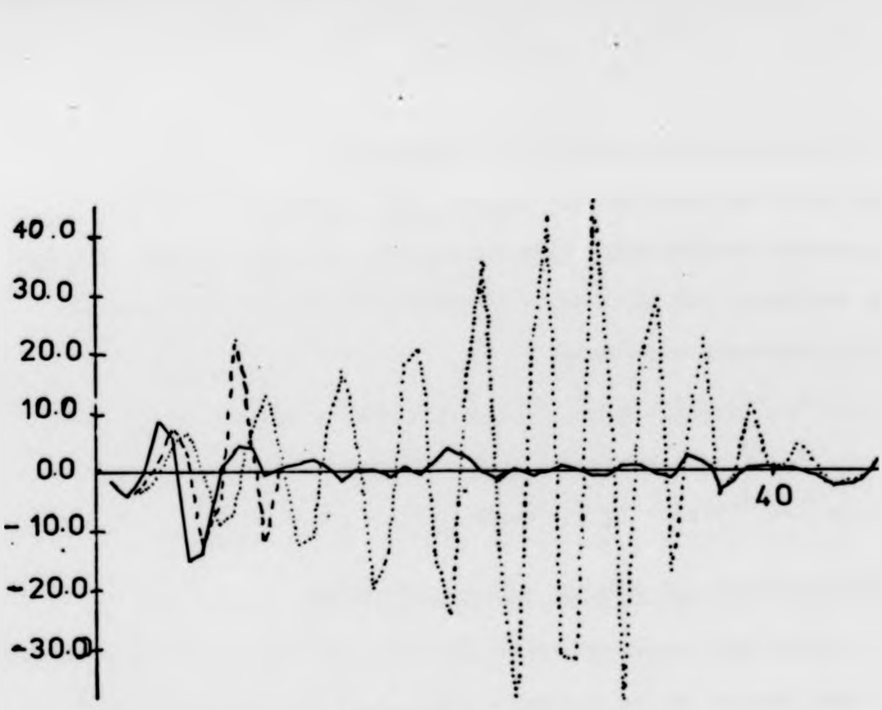


Figure 5.10a

Regulated outputs for example 5.3 for various values k I for the initial estimate of the inverse covariance matrix.

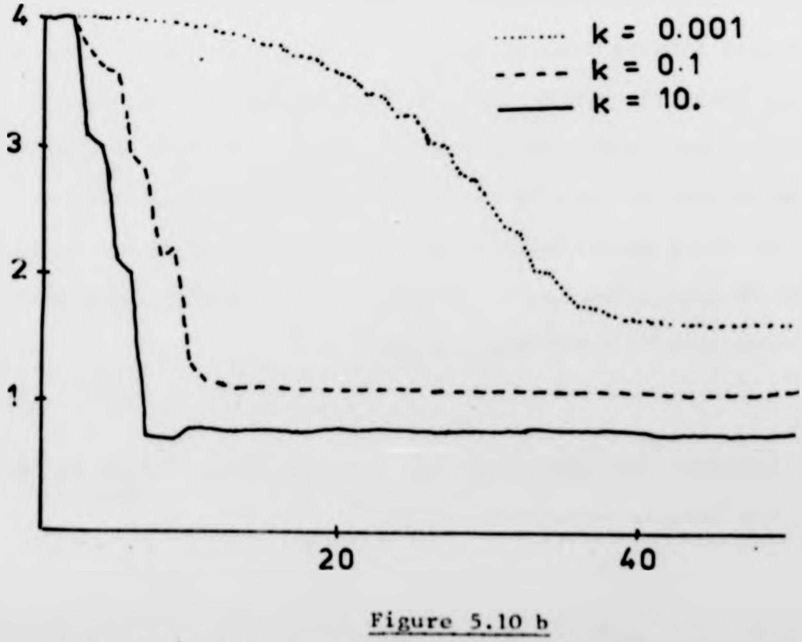


Figure 5.10 b

Estimates of the parameter b_1 for three different initial inverse covariance matrices in example 5.3.

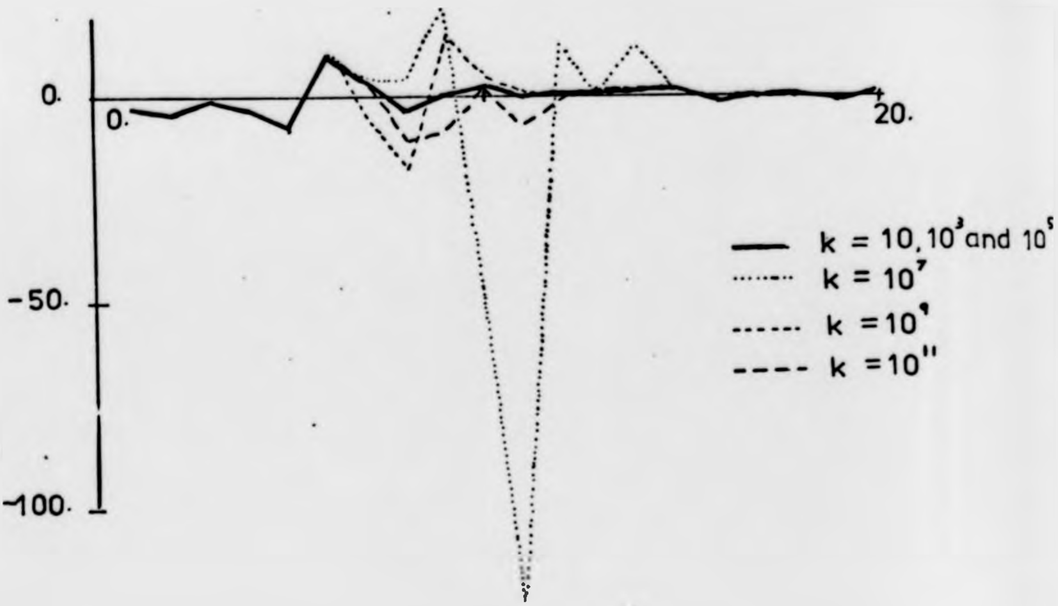


Figure 5.9a

Regulated outputs for example 5.2 for various values k I for the initial estimate of the inverse covariance matrix.

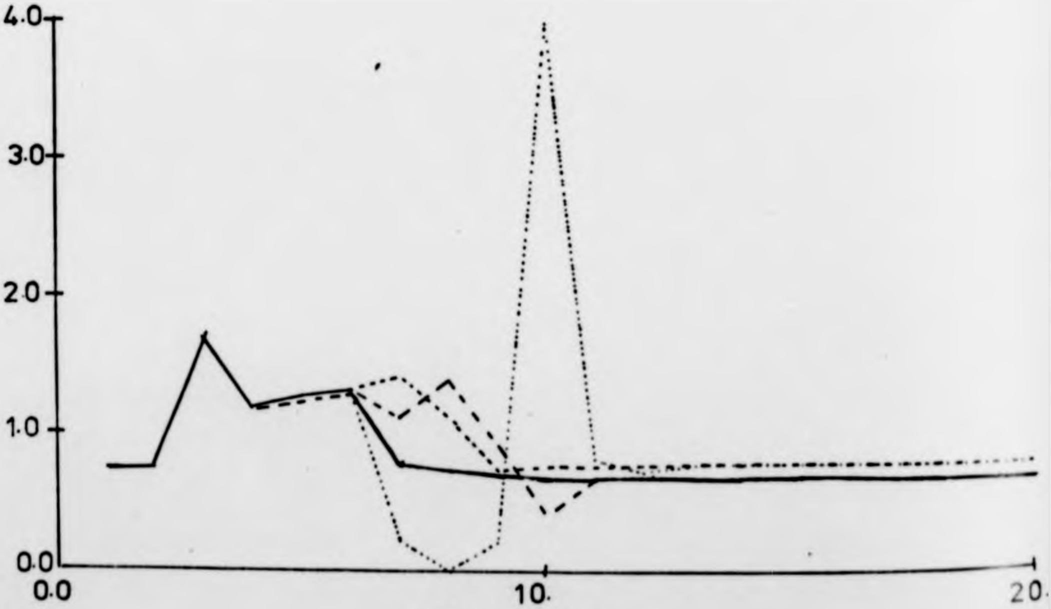


Figure 5.9 b

Estimates for the control parameter g_1

For this example the system being controlled was the same as in example 5.1, and the initial parameter estimates were chosen to be the same as the system's parameter values. The system was then simulated and controlled for initial values in the inverse covariance matrix of 10, 10^3, 10^5, 10^7, 10^9 and 10^{11}. Figure 5.9a shows the system output for each of the rows. The output was virtually identical for the three curves using 10, 10^3 and 10^5 initially. The output with an initial value of 10^7 was very disturbed by about the tenth sample. The curves for initial values of 10^9 and 10^{11} have a similar magnitude of disturbance to the first three. Figure 5.9b shows the value of the control parameter g_1 for each of the simulations. It can be seen that this control parameter was a large peak for the system using an initial value of 10^7 indicating that the calculation of the control law happened to become badly conditioned at about the tenth step.

<u>Example 5.3</u>

To illustrate slow convergence of parameters when using small initial values in the inverse covariance matrix.

For this example the system being controlled was the same as in the previous two examples.

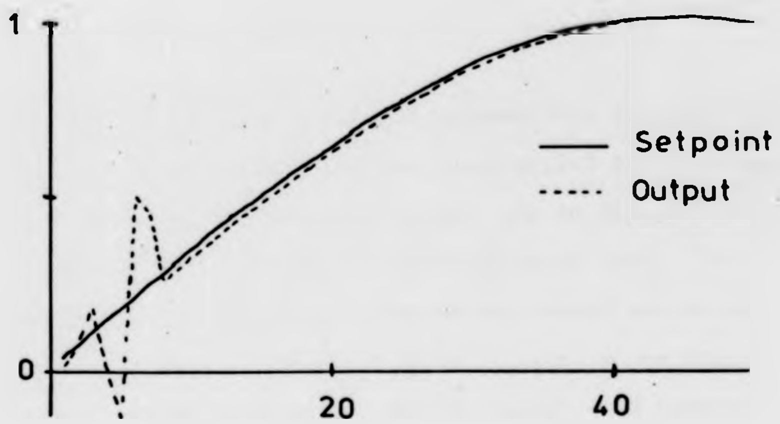
The system was described by equation 5.70, and the initial estimate of the system was

$$(1-2z^{-1}+z^{-2})Y = (4z^{-1}+0.0z^{-2})U + e$$

Figure 5.10b shows how the estimate of the coefficient b_1 changed with time for three different initial choices of the inverse covariance matrix. It can be seen that the convergence rates decreased as the initial choice of the inverse covariance matrix was decreased. Figure 5.10a shows the corresponding system outputs. As could be expected the run with slowly converging estimates disturbed the system output considerably.

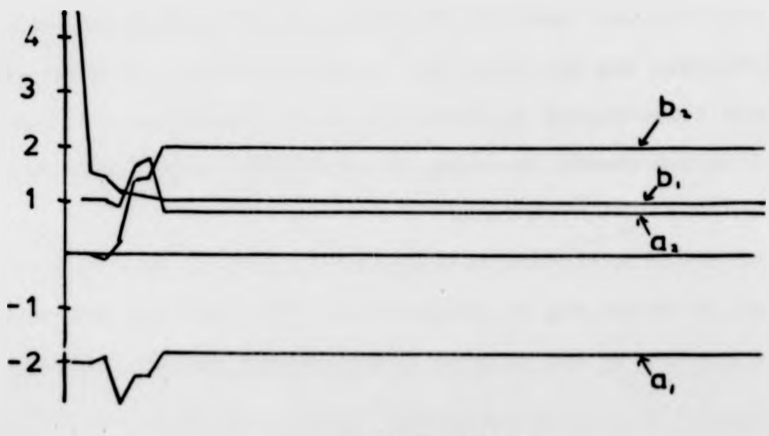
From the above two examples it can be seen that if the output is in the range +5 to −5 fairly good results are obtained if the initial values on the diagonal of the inverse covariance matrix were in the range $10^1 - 10^5$. This range of suitable initial values depends on the magnitude of the inputs and outputs. For example, if the noise driving a system was multiplied by 10 the inputs and output variances would be increased by a factor of 100, and so the initial inverse covariance matrix would have to be divided by 100 to give the same initial behaviour for the parameter estimates. Therefore, if the input and output are increased by a factor of 10 the suitable range of initial values is decreased by a factor of 100. When the inputs and outputs have different orders of magnitude it may be best to include constant multiplying factors to give them similar magnitudes in order to have similar initial rates of adaptation for all the parameters. If multiplying factors are not used, the initial choice of diagonal values should be made large enough to allow all of the parameters to adapt, i.e. the diagonal values should be chosen by considering the signal which is smallest.

Two of the main factors influencing the rate of convergence of the parameter estimates can be altered or decided when the self tuning regulator is applied to the system. These factors are the initial choice



<u>Figure 5.11 a</u>

Control output following a sinusoidal setpoint with no noise, on the system. Example 5.4.



<u>Figure 5.11 b</u>

Parameter estimates for example 5.4.

of inverse covariance matrix mentioned in the previous section and the introduction of limits on the control signals. However, the major factor influencing the convergence rates is the nature of any random disturbance on the system, and this cannot usually be altered. When there is no random disturbance on the system, the convergence of the estimates is very rapid. White noise on the system slows the convergence slightly, but still gives rapid convergence if there are any large initial disturbances to act as test signals. Coloured noise slows the convergence considerably since then the final estimates have to be biased in a special way which depends on the noise. With coloured noise large initial disturbances do not help the convergence very much because these disturbances tend to give unbiased estimates instead of the required biased estimates.

Example 5.4

To illustrate rapid convergence when no noise is present.

The system being controlled obeyed an equation

$$(1-1.8z^{-1}+.81z^{-2})Y(t) = (z^{-1}+2z^{-2})U(t)$$

This was controlled using the general pole shifting self tuning regulator with $T(z^{-1}) = 1$. Figure 5.11a shows the output using a sinusoidal setpoint. The parameter estimates are shown in Figure 5.11. It can be seen that the parameters have reached their correct values after 7 samples. When no noise is present the estimates usually take just two or three more samples than there are parameters to reach the correct values.

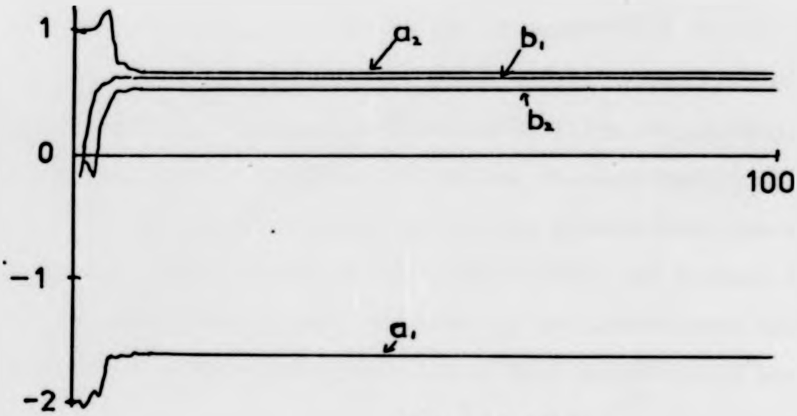


Figure 5.13 a

Parameter estimates for example 5.5 where a noise free analogue computer is being controlled by a self tuning regulator.

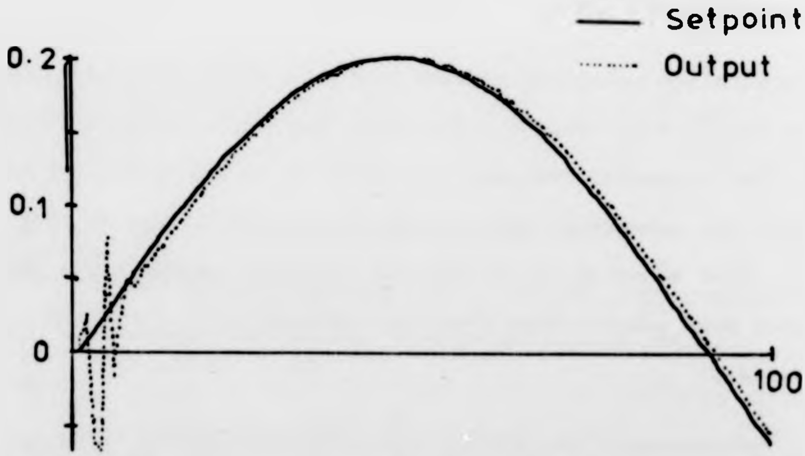


Figure 5.13 b

Setpoint and Analogue output for example 5.5.

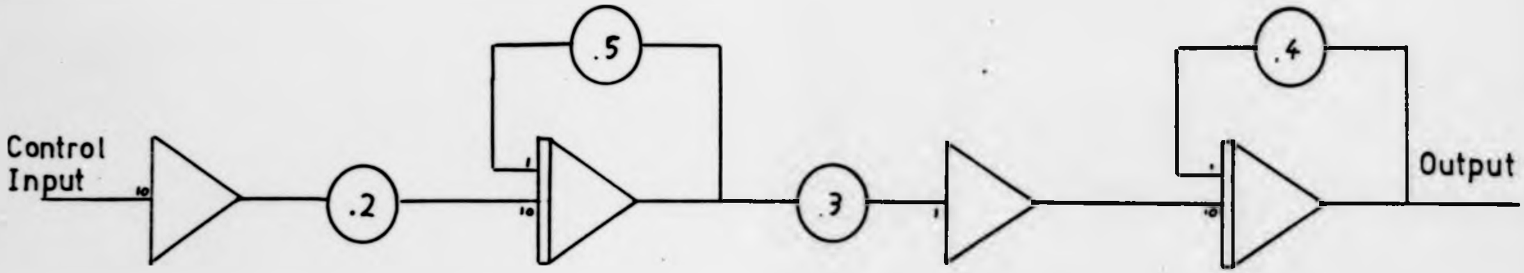


Figure 5.12

Analogue patching diagram for example 5.5.

Example 5.5

An analogue example showing rapid convergence with little noise present.

For this an analogue simulation was patched on one of the Ci175 analogue computers in the Control System Centre, U.M.I.S.T. This simulation was then controlled by a PDP-10 digital computer using a self tuning regulator with a minimum variance control law. The simulation was not run continuously since the PDP-10 has a multiaccess time sharing system and the hybrid subroutines do not allow the use of disk storage while working at a high enough priority to ensure an immediate response at the sample times. The simulation was carried out by setting the analogue computer into compute mode for 500 milliseconds, and then putting it into hold mode while the digital computer sampled the output and calculated the next control input. When the digital computer had sent the new control input, the analogue computer was again put into compute mode for 500 milliseconds.

The patching diagram for the analogue simulation for this example is shown in Figure 5.12. This corresponds with a system having a Laplace transfer function given in equation 5.71.

$$G(s) = \frac{30}{(1+2s)(1+2.5s)} = \frac{30}{1+4.5s+5s^2} \tag{5.71}$$

The z-transfer function is

$$G(z) = \frac{0.6465(z^{-1}+.8607z^{-2})}{(1-0.81873z^{-1})(1-0.7788z^{-1})}$$

Figure 5.13a shows the convergence of the parameters when no deliberate disturbances were added, and the set point was varying. Figure 5.13b shows the output of the system, and the setpoint which is being followed.

Figure 5.14 A rapid convergence of estimates with a large initial disturbance

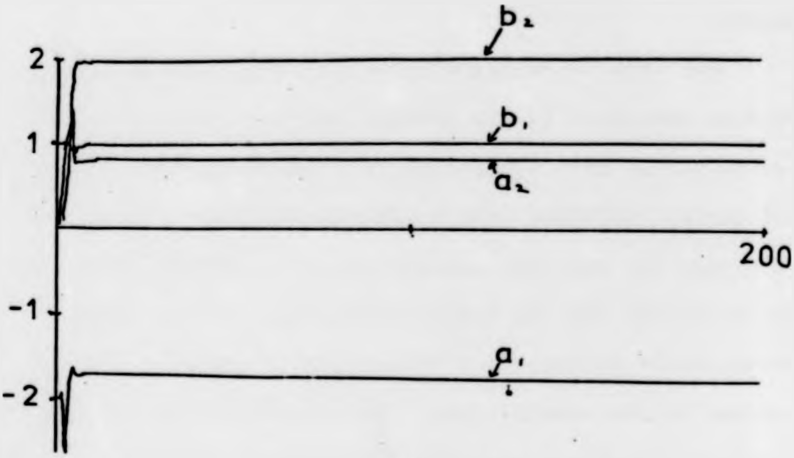


Figure 5.14 a

Rapid convergence of estimates with large initial disturbances in example 5.6.

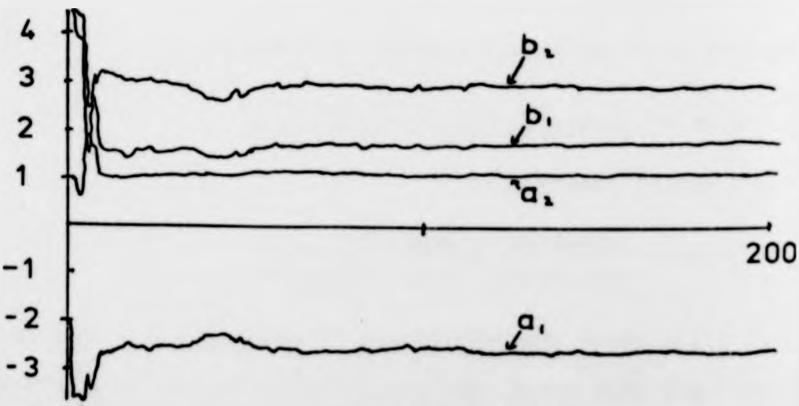


Figure 5.14 b

Slower convergence of parameter estimates with no large initial disturbance in example 5.6.

The initial estimates of the numerator polynomial (Figure 5.13a) started with the wrong sign, which is equivalent to having the wrong sign on the gain of the system.

The final estimated transfer function was

$$G_E(z^{-1}) = \frac{0.111863z^{-1} + 0.0931618z^{-2}}{1 - 1.66137z^{-1}+0.702652z^{-2}}$$

This corresponds with a Laplace transfer function

$$G_e(s) = \frac{5}{1+3.5s+5s^2}$$

The two examples just given illustrate how the parameter estimates converge very rapidly if the system is linear, and there are no unknown random disturbances present. When there is just white noise the convergence is still fairly rapid, particularly if there is a large initial disturbance to the system due to incorrect control. The convergence speed is improved by a large initial disturbance because this acts as a test signal.

Example 5.6

To show rapid convergence when a large initial disturbance occurs.

This rapid convergence can be seen in the parameter estimates of the runs used in example 5.1. Figure 5.8b showed two simulation runs which had an excessive initial disturbance. The parameter estimates for the run with starting estimates given by:

$$(1-2z^{-1}+z^{-2})Y(t) = (0.1z^{-1}+0.1z^{-2})U(t) + e(t)$$

are shown in Figure 5.14a. It can be seen that the estimates settle very rapidly. Figure 5.14b shows the parameter estimates for the run starting with:

$$(1-2z^{-1}+z^{-2})Y(t) = (4.5z^{-1}+4.5z^{-2})U(t) + e(t)$$

The output for this run is shown in 5.8b and does not have an excessive initial disturbance. It can be seen that the settling time for the estimates is longer.

The convergence of parameters in the case with white noise is fairly rapid since any initial incorrect control acts as a test signal, and this test signal helps the parameters to adjust to those which describe the system. This rapidly gives good control since, when the noise is white, the estimator gives unbiased estimates of system parameters which lead to the control law which is required asymptotically by the self tuning regulator. However, when the noise is not white the final system estimates have to be biased in the correct manner to give the required control. Therefore, in the case with coloured noise, any initial poor control will not drive the system estimates to the required values. Similarly, setpoint variations will also tend to drive the estimates from the required values. An extreme example of the resulting slow convergence was given by the example given in section 5.2.1 (Figure 5.5).

Example 5.7

To illustrate slower convergence due to coloured noise.

For this example the same system as in example 5.1 was controlled in the presence of three different disturbances. The system being controlled was described by equation 5.72

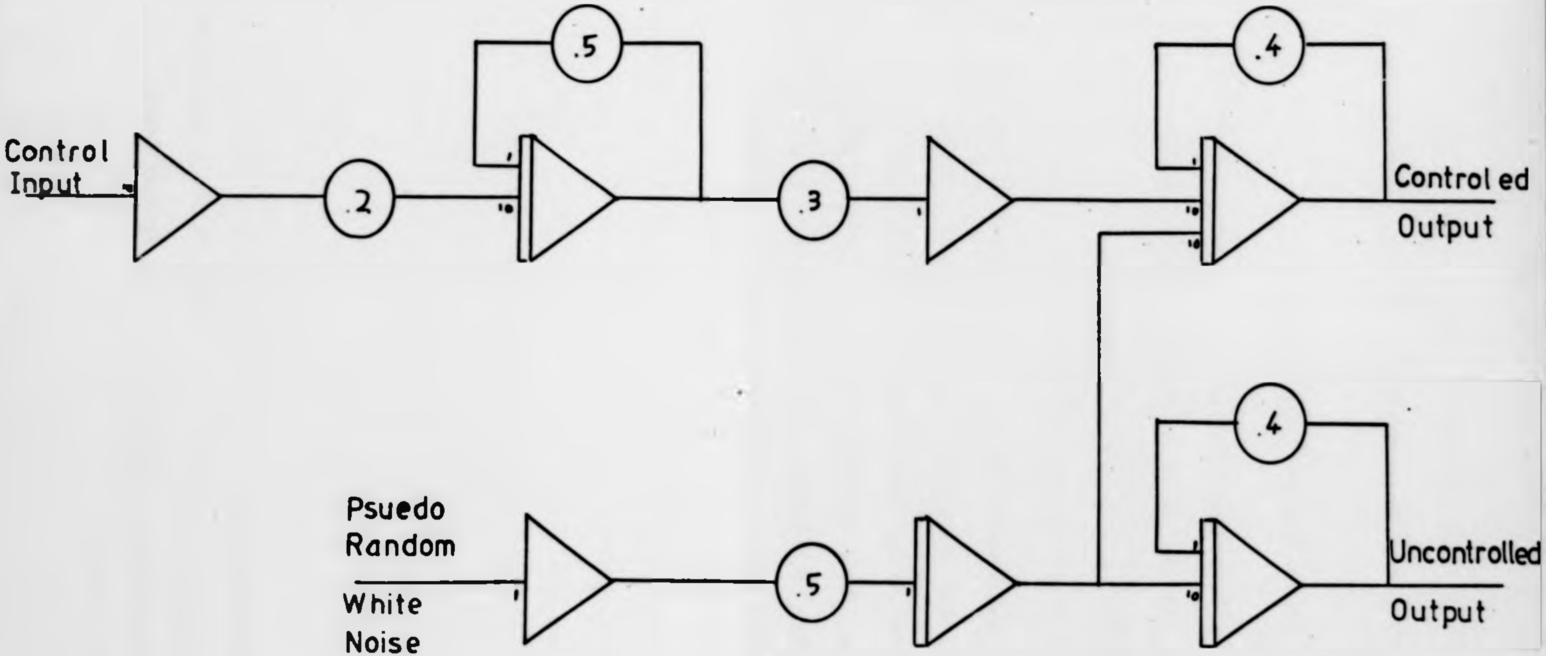


Figure 5.16

Analogue patching diagram for example 5.8 where a coloured noise is added to the system.

Figure 5.15c

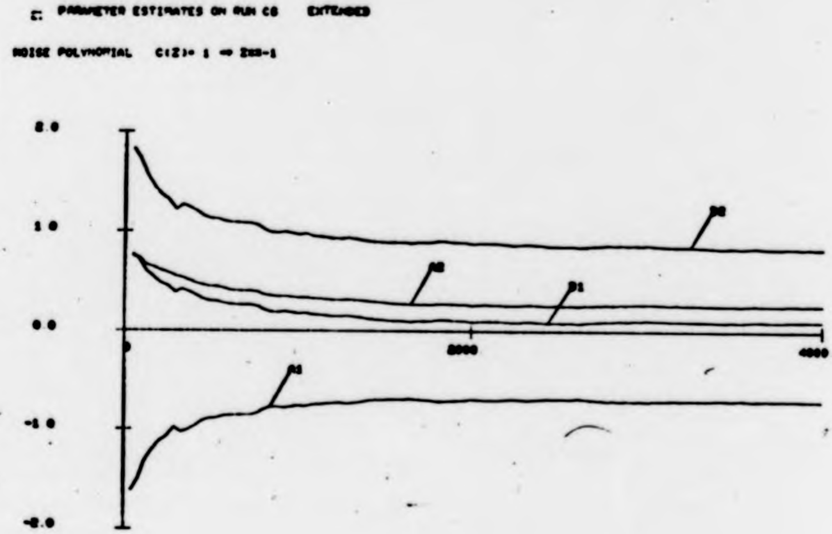


Figure 5.15 d

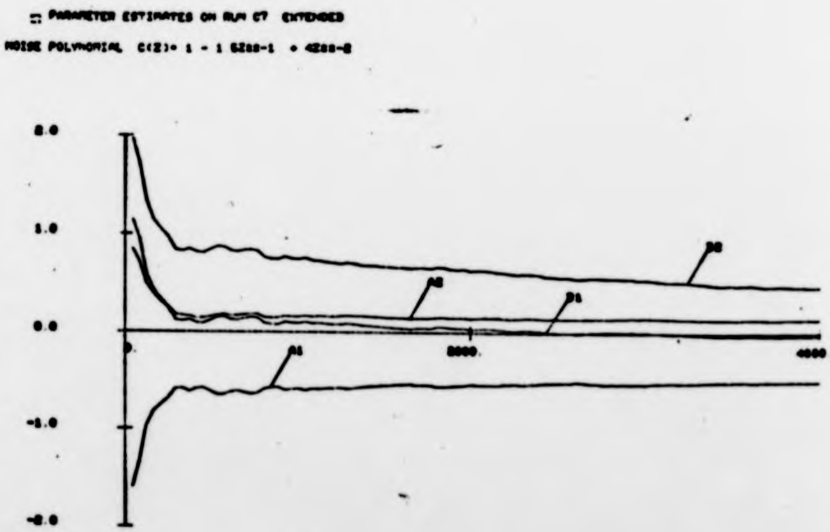


Figure 5.15 a

PARAMETER ESTIMATES ON RUN C4

THIS SIMULATION USED C(Z)=1 (WHITE NOISE)

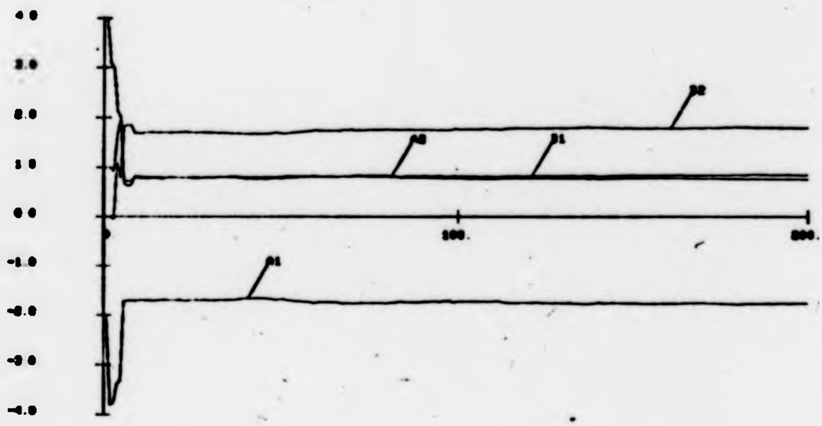
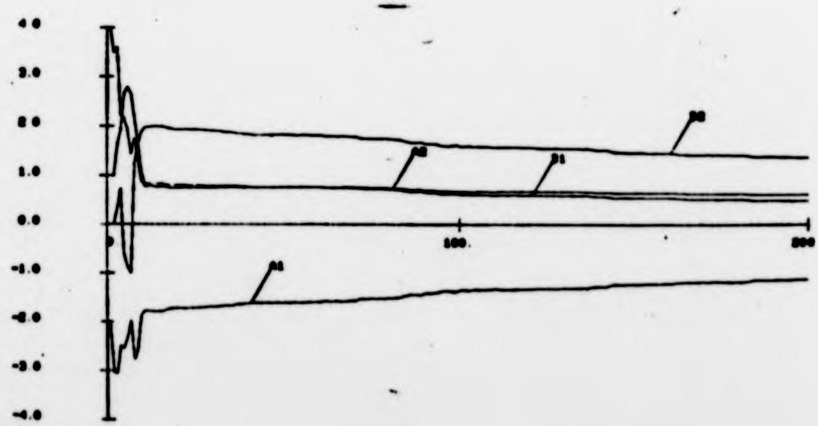


Figure 5.15 b

PARAMETER ESTIMATES ON RUN C5

THE NOISE POLYNOMIAL FOR THIS RUN WAS C(Z)= 1.- 2ZZ-1



$$(1-1.8z^{-1}+.81z^{-2})Y(t) = (z^{-1}+2z^{-1})U(t) + C(z^{-1})e_s(t) \qquad (5.72)$$

using an initial estimate $(1-2z^{-1}+z^{-2})Y(t)=(4z^{-1}+0.0z^{-2})U(t) + e(t)$.

Figure 5.15a shows the parameter estimates with $C(z^{-1}) = 1$, i.e. white noise. Here the estimates settle rapidly. Figure 5.15b shows the parameter estimates with $C(z^{-1}) = 1-z^{-1}$. It can be seen that the estimates this time have not settled within the first 200 samples. Figure 5.15c shows the estimates with this noise for a longer period. It can be seen that the estimates are fairly settled after about 2000 samples. Figure 5.15d shows the parameter estimates with $C(z^{-1}) = 1-1.5z^{-1}+0.4z^{-2}$. Here the estimates have not settled after 4000 samples, so it can be seen that with this system colouring of the disturbance can considerably increase the settling time for the estimates.

Example 5.8

For this example a pseudo random coloured noise signal was added to the analogue simulation used in example 5.5.

The patching diagram for the analogue simulation is shown in Figure 5.16. A pseudo random noise sequence was used in order to make the experiment repeatable. The uncontrolled system output was also simulated on the analogue computer in order to have a direct measure of how well the regulator was managing. It was found that the results from this experiment were not exactly repeatable, possibly due to small variations in the offset voltage on the D.A.C. being used for the pseudo random noise. These small variations in offset would be considerably magnified by the integrator which was being used to colour the noise.

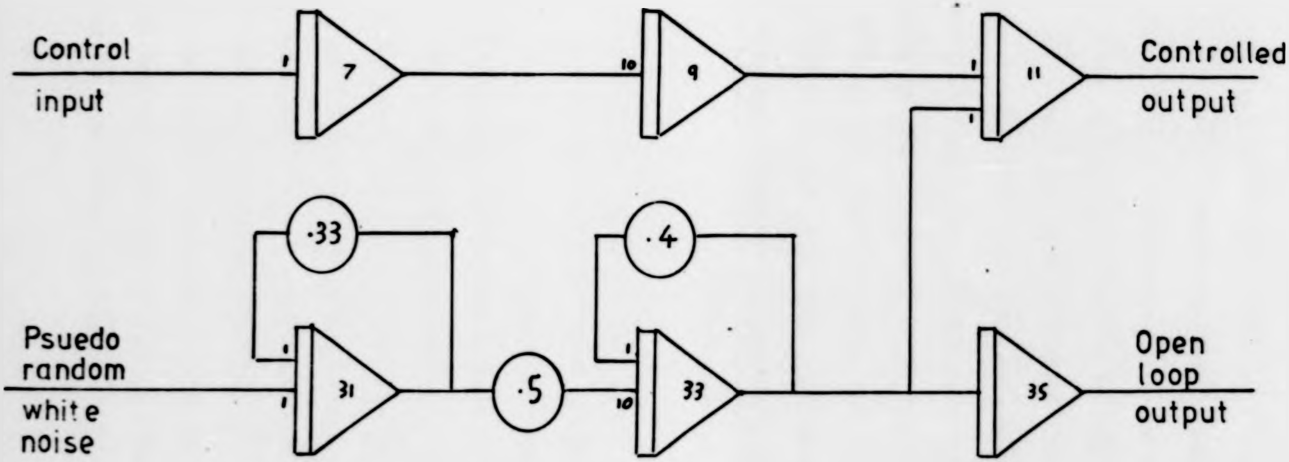


Figure 5.18

Analogue patching diagram for example 5.9. This was controlled using the general pole shifting self tuning regulator.

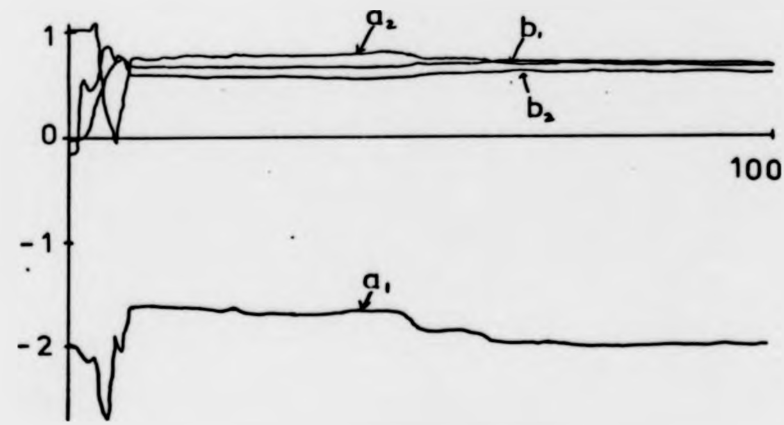


Figure 5.17a

Parameter estimates for example 5.8 with a minimum variance self tuning regulator controlling a disturbed analogue simulation.

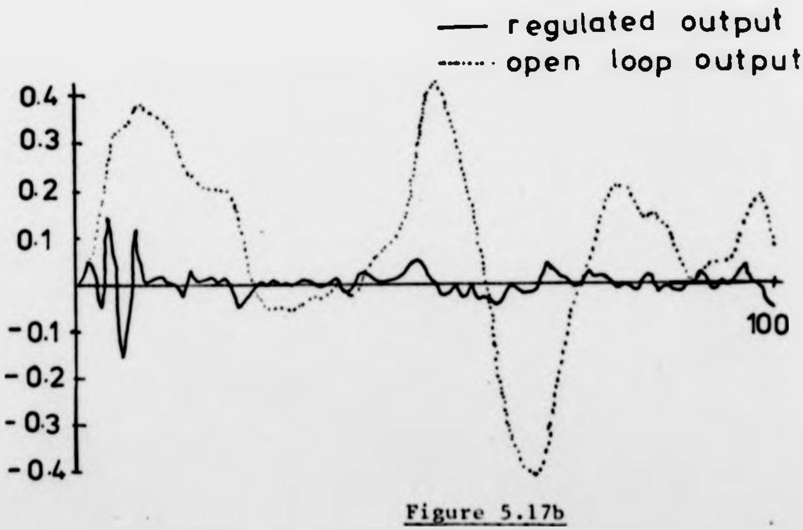


Figure 5.17b

Regulated and open loop outputs for example 5.8.

The Laplace transfer function for the system was given by 5.73.

$$Y = \frac{30\,U}{(1+2s)(1+2.5s)} + \frac{k\,e}{s(1+2.5s)} \qquad (5.73)$$

The corresponding z-transfer function is

$$Y = \frac{.6465(z^{-1}+0.8607z^{-2})}{(1-0.8187z^{-1})(1-0.7788z^{-2})}\;U. + \frac{k\;(0.04682z^{-1}+0.0438z^{-2})}{(1-z^{-1})(1-0.8187z^{-1})}\,e$$

This system was controlled using a minimum variance self tuning regulator using an initial estimate

$$(1-2z^{-1}+z^{-2})\;Y(t) = (-0.2z^{-1}-0.1z^{-2})\;U(t) + e(t)$$

Figure 5.17a shows the parameter estimates are much slower settling than the estimates in example 5.5 (Figure 5.13a & b). The corresponding controlled and uncontrolled outputs are shown in Figure 5.17b.

Example 5.9

A hybrid example with coloured noise using the general pole shifting self tuning regulator.

The analogue patching diagram for this example is shown in Figure 5.18. This corresponds to a system with a Laplace transfer function:

$$Y = \frac{10}{s^3}\;U + \frac{37.5\,e}{s(1+3s)(1+2.5s)}$$

The corresponding z-transfer function is

$$Y = \frac{1.3(z^{-1}+4z^{-2}+z^{-3})}{1-3z^{-1}+3z^{-2}-z^{-3}}\,U + \frac{35 * 0.002537(1+0.2443z^{-1})(1+3.407z^{-1})}{(1-z^{-1})(1-.8465z^{-1})(1-.8187z^{-1})}\,e$$

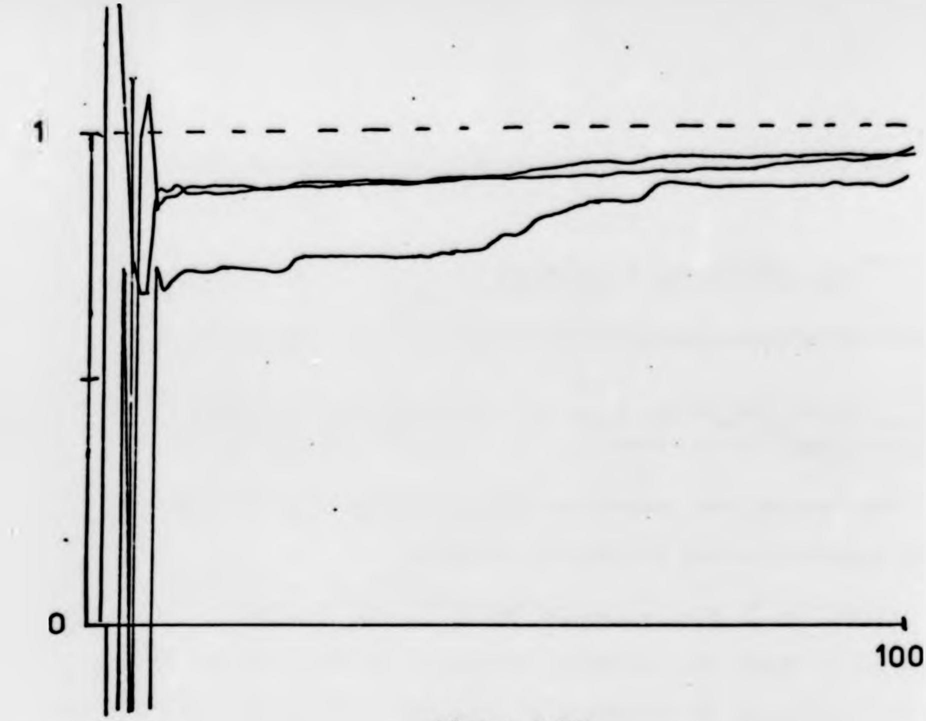


Figure 5.20a

Convergence of the ratios of the elements of an inverse covariance matrix in example 5.10.

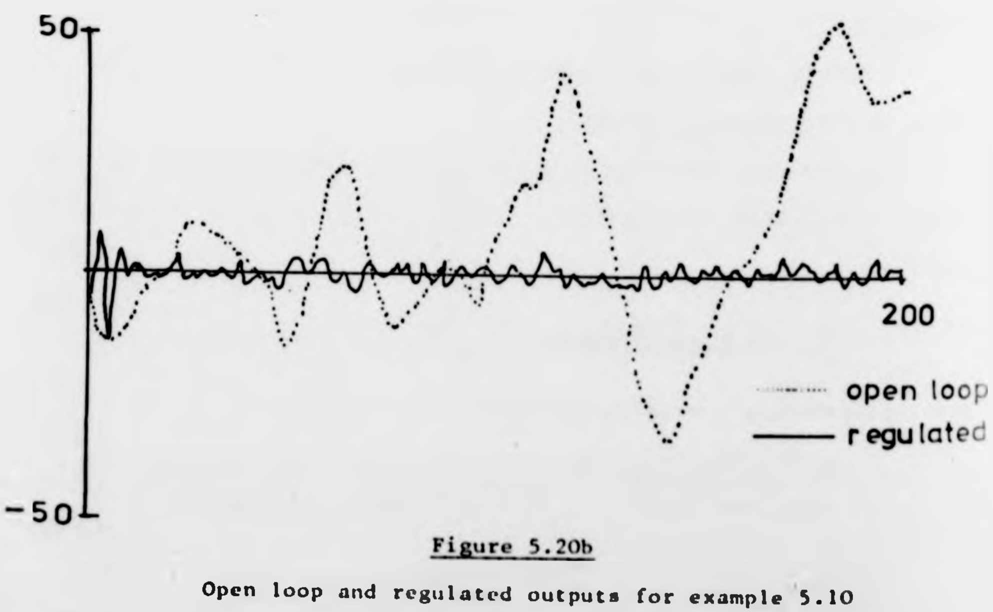


Figure 5.20b

Open loop and regulated outputs for example 5.10

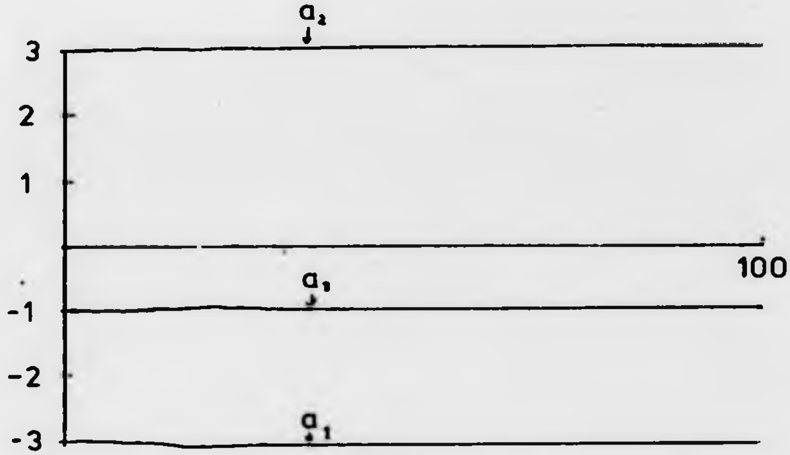


Figure 5.19c

Parameter estimates for example 5.9. The initial estimates are correct since in all the samples the starting estimates for the A polynomial have been obtained by assuming that the system is a collection of pure integrators, and in this case it is.

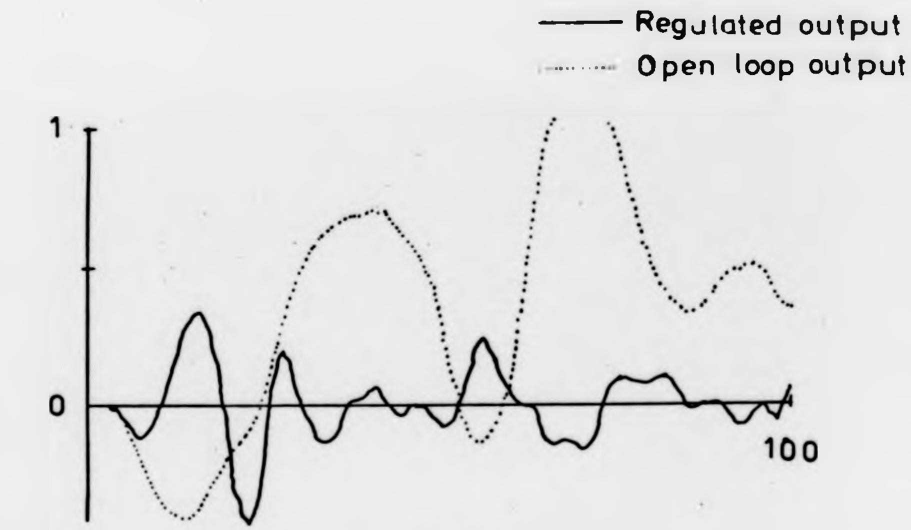


Figure 19a

Regulated and open loop outputs for example 5.9 with an analogue simulation with coloured noise being controlled by the pole shifting regulator.

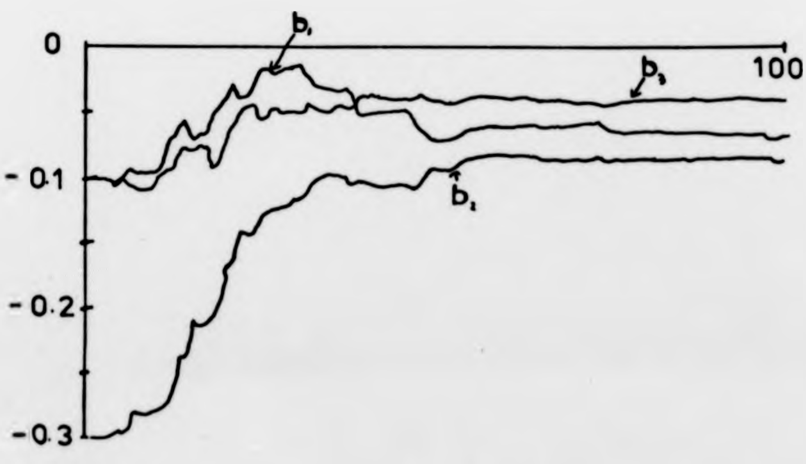


Figure 19b

Parameter estimates for example 5.9.

Notice that this has a zero outside the unit circle, and so the minimum variance controller section 5.11 would be unstable.

Figure 5.19a shows the controlled and uncontrolled outputs with the particular noise sequence used.

Figures 5.19b and c show the behaviour of the parameter estimates.

5.3.4 Convergence of inverse covariance matrix

The elements of the estimated inverse covariance used by the least squares estimation algorithm seem to settle slightly slower than the corresponding parameter estimates. However, they rapidly get to the correct order of magnitude so probably very little would be gained by trying to get the elements in the correct sort of ratios before starting the simulation.

Example 5.10

To illustrate the convergence of the inverse covariance matrix.

Since this matrix continually decreases it is only the ratios of the elements which converge.

In order to display the convergence of these ratios the ratios have been divided by the corresponding ratio reached after 300 samples, so that all the curves should approach 1 as time increases.

$$\text{The displayed curve at } (t) = \frac{C_{i,j}(t)/C_{i,1}(t)}{C_{i,j}(300)/C_{i,1}(300)}$$

Figure 5.20a shows a selection of these ratios for the first system used in example 5.7. The convergence of the parameters is shown in Figure 5.15a. Figure 5.20b shows the corresponding time response together with the uncontrolled time response.

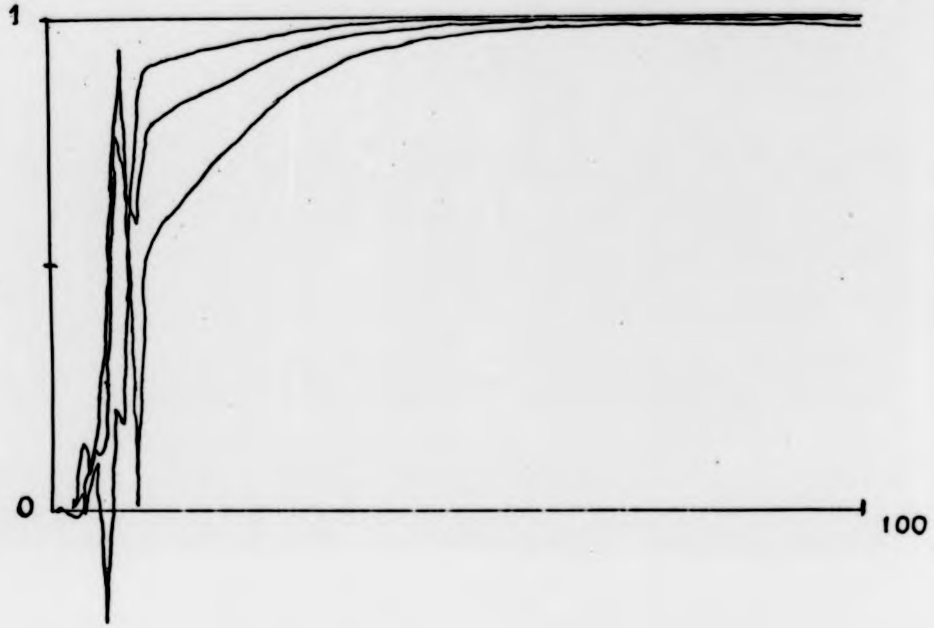


Figure 5.21

Convergence of inverse covariance matrix for example 5.4.

Figure 5.21 shows a selection of these ratios for the example 5.4 which had very rapid settling for the parameter estimates (Figure 5.11b).

References for Chapter 5

1 Ljung, L. and Wittenmark, B. "Asymptotic Properties of Self Tuning Regulators", Report 7404 : Division of Automatic Control, Lund Institute of Technology, February, 1974

2 Astrom, K.J. and Wittenmark, B., "On Self Tuning Regulators", Automatica, Vol. 9, 1973, pp. 185-199

3 Astrom, K.J. "Introduction to Stochastic Control", Academic Press, 1970

4. Astrom, K.J. and Wittenmark, B. "Analysis of a self tuning regulator for non minimum phase systems", I.F.A.C. Symposium on Stochastic Control, Budapest, 1974

5. Box, G.E.P. and Jenkins, G.M. "Time Series Analysis Forecasting and Control", Holden-Day, 1970, Section 3.4

CHAPTER 6 COMPUTATION

6.1 Comparison of two methods using a minimum variance control

In section 5.1 two methods of implementing a minimum variance self tuning regulator in the presence of time delays were suggested. These two methods are compared in this section.

The first method of implementing a minimum variance regulator has been called method 1 in this chapter, and the associated computer programs.

For this method the model of the system used is

$$(1 + z^{-k}A'(z^{-1}))Y(t) = z^{-k} B'(z^{-1}) U(t) + e(t) \qquad (6.1)$$

The control is then calculated as

$$U = \frac{A'(z^{-1}) Y(t)}{B'(z^{-1})} \qquad (6.2)$$

In the second method called method 2 in this chapter, the model used for the system is:

$$(1 + A(z^{-1})) Y(t) = z^{-k} B(z^{-1}) U(t) + e(t) \qquad (6.3)$$

The polynomials A' and B' are then calculated using the relationships

$$(1 + z^{-k} A'(z^{-1})) = (1 + A(z^{-1})) P(z^{-1}) \qquad (6.4)$$

and $$B'(z^{-1}) = P(z^{-1}) B(z^{-1})$$

The control is then calculated using equation 6.2.

In the first method the order of the numerator polynomial estimated by the least squares estimator is k greater than that in the second method, so it would be expected that the second method should converge more rapidly since it has less degrees of freedom.

Table 6.1 Showing parameter estimates after 100 samples in example 6.1 using two different estimators.

	True Values	Estimator using $P(z^{-1})$ Method = 2	Direct Estimator Method = 1
a_1'	-3.281	-3.428	-3.592
a_2'	+2.362	2.500	2.675
b_1'	1.999	2.043	2.712
b_2'	4.598	4.500	5.039
b_3'	6.658	6.499	6.901
b_4'	8.259	8.100	7.834
b_5'	2.915	2.352	1.596

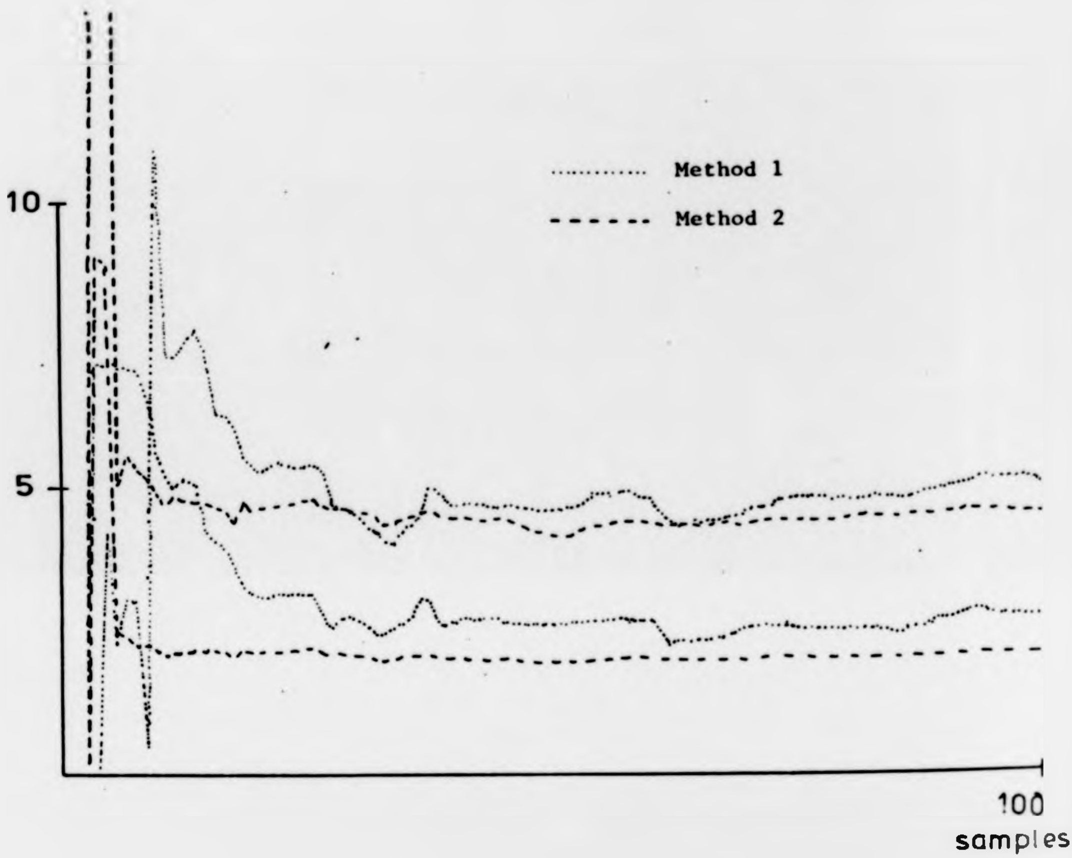


Figure 6.2

Open loop estimates of the first two terms ofthe $B'(z^{-1})$ polynomials with a disturbance on the system.

Method 1 estimates $B'(z^{-1})$ directly.

Method 2 estimates $A(z^{-1})$ and $B(z^{-1})$ and calculates $B'(z^{-1})$

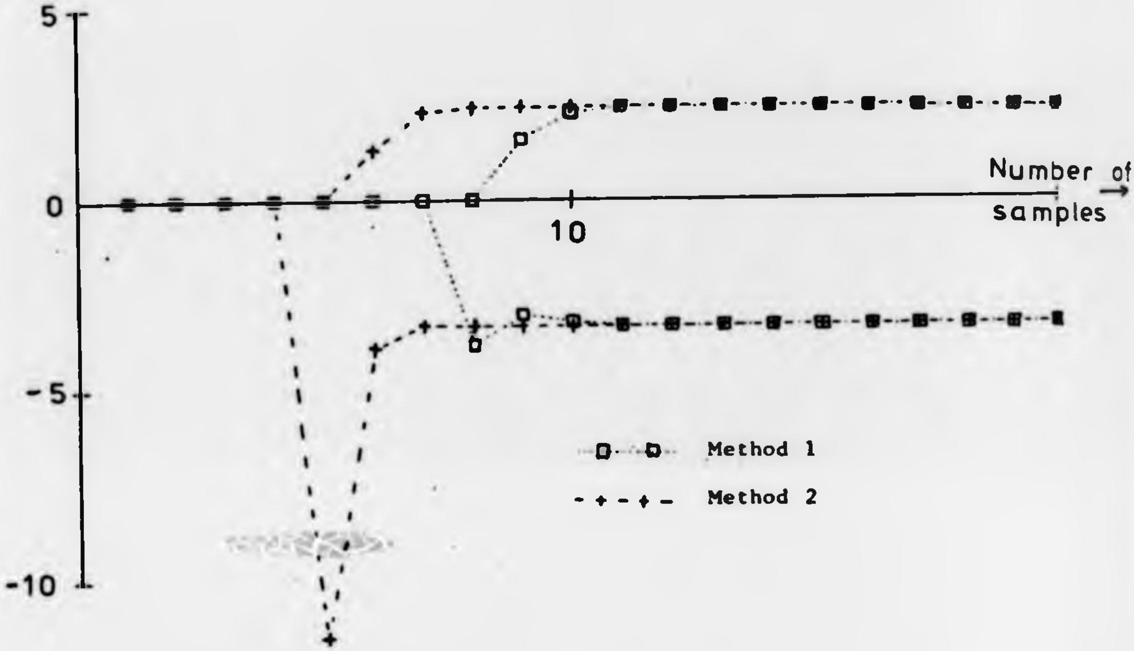


Figure 6.1

Open Loop Estimates of the $A'(z)$ polynomials using the two methods of calculation.

Method 1 estimates $A'(z^{-1})$ directly

Method 2 estimates $A(z^{-1})$ and then calculates $A'(z^{-1})$

Example 6.1

In order to illustrate the more rapid convergence with the smaller order estimator the two estimators were used on the same open loop disturbed system. An open loop system was used to eliminate effects of poor feedback control on the system estimates.

A system described by

$$(1-1.8z^{-1}+.81z^{-2})\ Y(t) = (2z^{-4}+z^{-5})\ U(t) + \ell e(t) \tag{6.6}$$

was simulated with U(t) being generated by a white noise generator. An alternative model of the system is

$$(1 - 3.2805z^{-4}+2.362z^{-5})\ Y(t)$$

$$= (2.0z^{-4}+4.6z^{-5}+6.66z^{-6}+8.262z^{-7}+2.196z^{-8})\ U(t) + \ell e(t) \tag{6.7}$$

Figure 6.1 shows the estimates of the A polynomials obtained by the two methods when there was no unknown disturbance e(t). It can be seen that both converge to within machine accuracy within about 14 steps, with the estimates from the lower order estimator taking slightly less time.

Figure 6.2 shows the estimates of the first two terms of the B polynomials when a random disturbance e(t) was added. It can be seen here that estimates with the lower order estimator settle more rapidly and more quickly approach the correct values. Table 6.1 gives the parameter estimates after 100 samples for the two methods.

In a self tuning regulator this slower convergence of parameters could lead to slower convergence of the control law or more disturbance to the system. In the examples tried it usually resulted in more disturbance to the system. The disturbance to the system due to inaccurate estimates was probably magnified by the chosen controllers frequently being unstable (see Chapter 2). The situation is also probably aggravated by the way that the first method has several extra parameters, and so effetively the $P(z^{-1})$ polynomial which has the unstable roots is estimated twice, once in the estimation of A' and once in the estimation of B', and so different unstable polynomials are used in the calculation of the numerator and denominator polynomials of the control law.

Example 6.2

To demonstrate the unsettling effect of using the higher order estimator.

In order to compare the two self tuning regulators they were both used to control the same system with the same disturbances, and with equivalent initial estimates of the system. The system being controlled was similar to that in example 6.1 and was described by equation 6.8.

$$(1-1.8z^{-1}+0.81z^{-1})\ Y(t) = (1.0z^{-4}+0.5z^{-5})\ U(t) + e(t) \qquad (6.8)$$

An identical model would be:

$$(1-1.8z^{-4}+0.81z^{-5})\ Y(t)$$
$$= (1.0z^{-4}+2.3z^{-5}+3.33z^{-6}+4.131z^{-7}+1.098z^{-8})\ U(t) + e(t) \qquad (6.9)$$

The initial estimate for the first method was:

$$(1-5z^{-4}+4z^{-5})Y(t) = (z^{-4}+2z^{-5}+3z^{-6}+4z^{-7}+0.0z^{-8})U(t) + e'(t) \qquad (6.10)$$

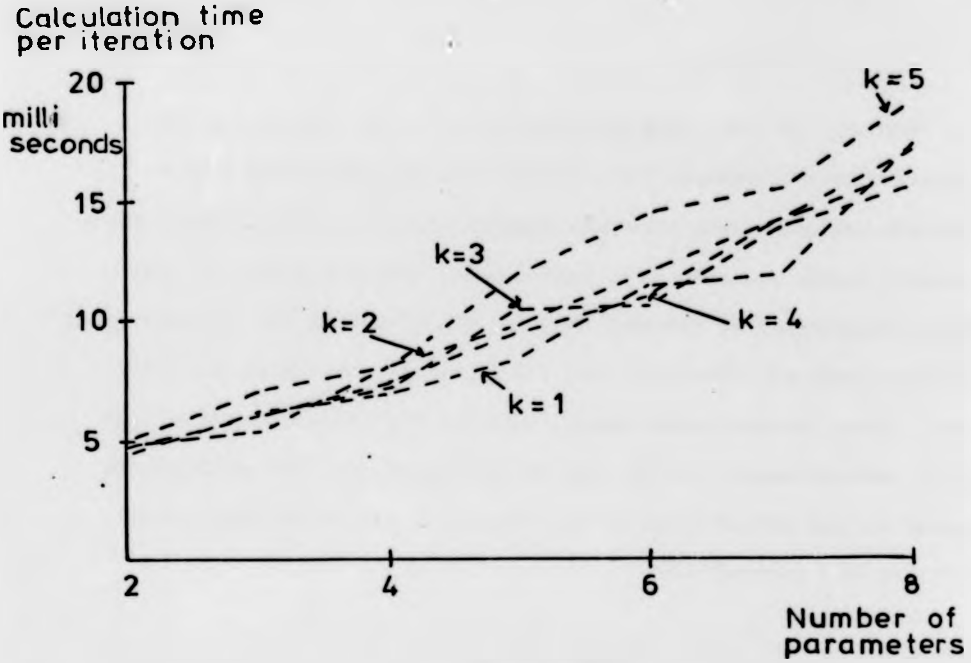


Figure 6.4

C.P.U. time per iteration for Method 2 for various values of delay k.

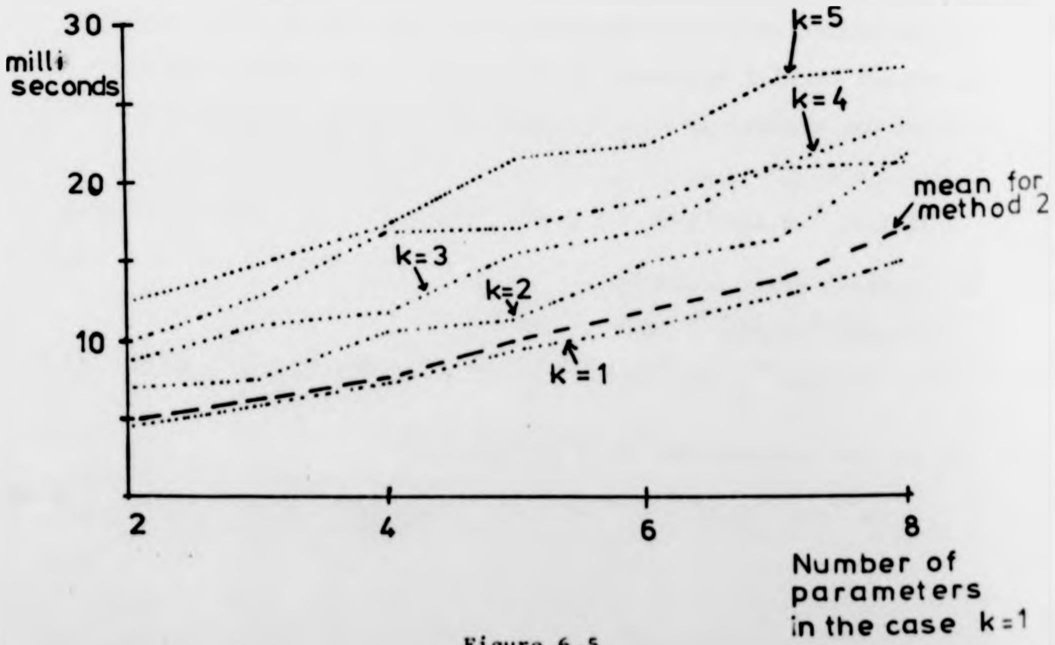
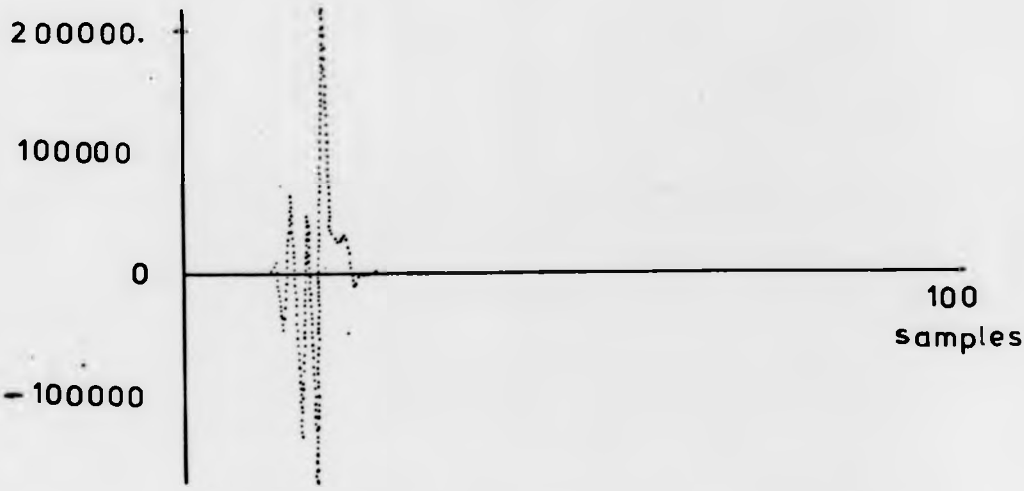


Figure 6.5

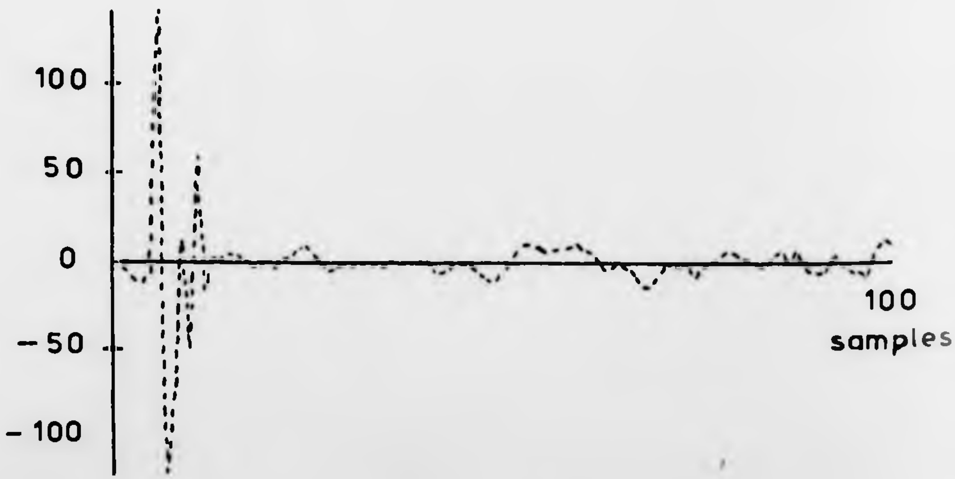
C.P.U. time per iteration for Method 1 for various values of delay.

Note different scales have been used for Figures 6.4 and 6.5



<u>Figure 6.3a</u>

Regulated System Output using Method 1 which estimates $A'(z^{-1})$ and $B'(z^{-1})$ directly.



<u>Figure 6.3b</u>

Regulated System Output using Method 2 which estimates $A(z^{-1})$ and $B(z^{-1})$ and then calculates $A'(z^{-1}$ and $B'(z^{-1})$.

Note different scales have been used for Figures 6.3a and 6.3b.

If the system had obeyed 6.10 it would also have obeyed 6.11

$$(1-2z^{-1}+z^{-2})Y(t) = (z^{-4}+0.0z^{-5})U(t) + e'(t) \qquad (6.11)$$

Therefore the starting estimates used for the second method were those in equation 6.11. The resulting output using the first method is shown in Figure 6.3a, and the output using the second method is shown in Figure 6.3b.

It can be seen that in this example the system was considerably disturbed by the first self tuning regulator. Note that in order to be able to see both curves different scaling has been used in the y direction.

Another advantage of using the second method is that the amount of computation required is less than that for the first method. Figure 6.4 shows the C.P.U. time used at each sample time as a function of the number of parameters being estimated, for time delays of 1, 2, 3, 4 and 5 sample intervals using the second method. It can be seen that the computation time was fairly independent of the time delay on the system. Some of the variations which occurred could be accounted for by variations in the loading of the time sharing computer being used since it will be noticed that the curves for different time delays keep crossing each other, while the amount of work done slightly increases with increases in the time delay. Figure 6.5 shows the C.P.U. time used at each sample time using the first method in which the order of the estimator is increased to compensate for the time delay. Figure 6.5 also shows the mean of the times for the second method. It can be seen that with the first method the calculation time increases with the time delay, and generally takes longer than the second method. The only disadvantage of the second method seems to be that it requires a few extra lines of computer program to implement it.

6.2 Comparison of general pole shifting self tuning regulators

In section 5.1 the following four different methods were proposed for solving the equation 6.12 for the control polynomials F and G

$$((1+A(z^{-1}))(1+F(z^{-1})) + B(z^{-1})\ G(z^{-1}))\ Y(t) = T'(z^{-1})\ Y(t) \qquad (6.12)$$

One method, which will be called method 3, was to rearrange 6.12 to become 6.13.

$$F(z^{-1})\ [(1+A(z^{-1}))Y(t)] + G(z^{-1})[B(z^{-1})Y(t)] = (T'(z^{-1})-1-A(z^{-1}))Y(t) \qquad (6.13)$$

The coefficients of F and G can then be recursively estimated using a least squares estimator, preferably an exponentially weighted estimator to remove the effects of poor initial estimates of $A(z^{-1})$ and $B(z^{-1})$. It was remarked that this method can have difficulties if there are not sufficient variations in the system output Y(t) to ensure that the control law converges rapidly.

To overcome this difficulty of insufficient variations in the system output it was proposed that a white noise sequence should be used in equation 6.13 instead of the output sequence Y(t). This modified method has been called method 4 in this chapter and the associated computer programs. It would be expected that this modification should make the regulator converge more rapidly and more consistently without significantly increasing the amount of computation required. However because the control estimates are just updated at each sample time using the current system estimates, the control estimates are bound to converge more slowly than the estimates of the system parameters.

Setpoint

System output

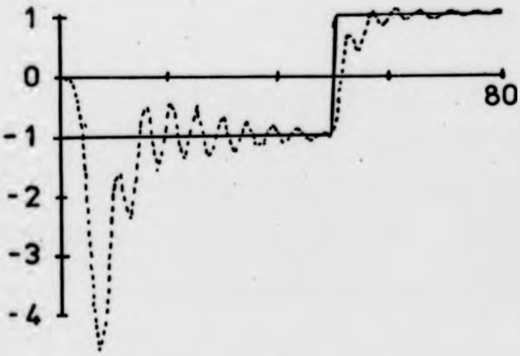


Figure 6.6a

Method 3

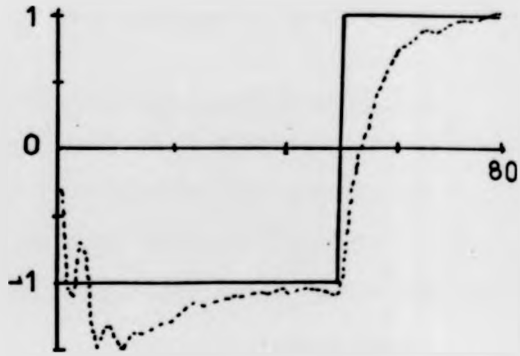


Figure 6.6b

Method 4

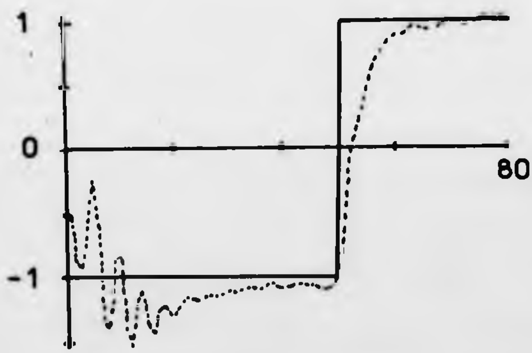


Figure 6.6c

Method 5

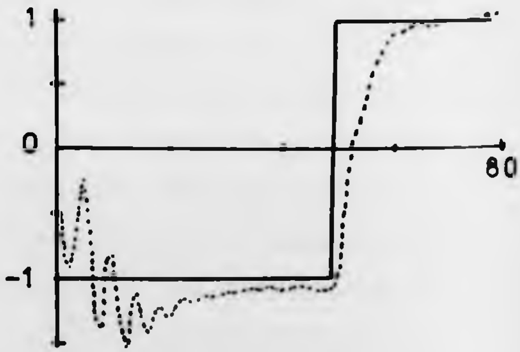


Figure 6.6d

Method 100

Figure 6.6

Initial setpoint following using the four different versions of the general pole shifting regulator.

One method of ensuring that the control converges as quickly as the system estimates is to consider equation 6.12 as a set of linear simultaneous equations, and solve them completely at each sample time. The control estimates will then correspond directly with the system estimates. The set of equations was solved by two different approaches. The first was to solve them by repeated calls to a recursive least squares estimator, this was called method 5. The second approach was to use a subroutine specially written to solve a set of simultaneous linear equations, this self tuning regulator was called method 100. The second of these approaches requires less computation, at the expense of a more complicated computer program.

Example 6.3

To illustrate the variation of the initial responses of the different ways of implementing the general pole shifting self tuning regulator.

For this example a system described by

$$(1-1.8z^{-1}+0.81z^{-2})\ Y(t) = (z^{-1}+2z^{-2})U(t) + e(t)$$

was controlled by each of the four self tuning regulators, in each case trying to follow a unit square wave input. The input to the system was limited to be between +0.3 and −0.3, so as to decrease the initial transients. The initial estimate used in each case was:

$$(1-2z^{-1}+z^{-2})Y(t) = (4z^{-1}+0.0z^{-2})U(t) + e\ (t)$$

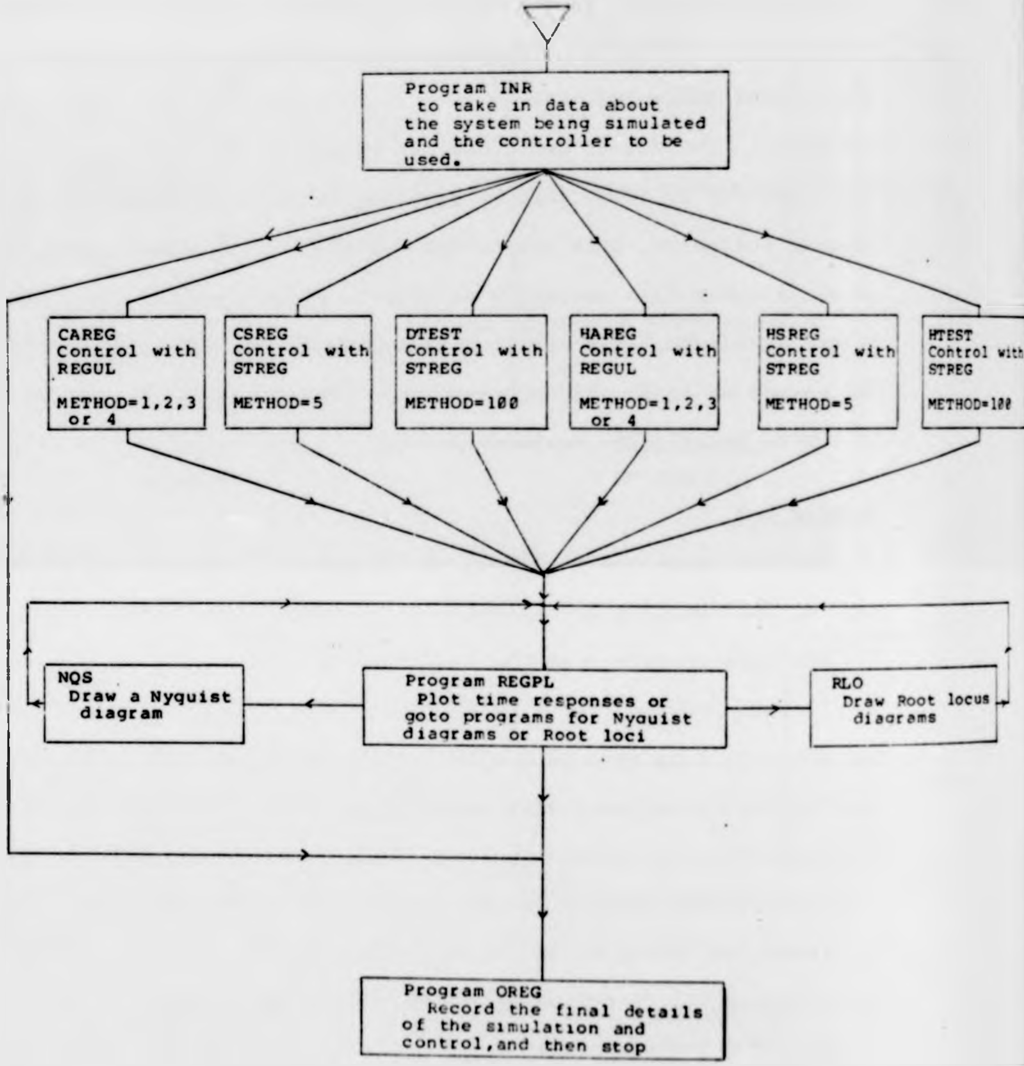
The required $T(z^{-1})$ polynomial in each case was

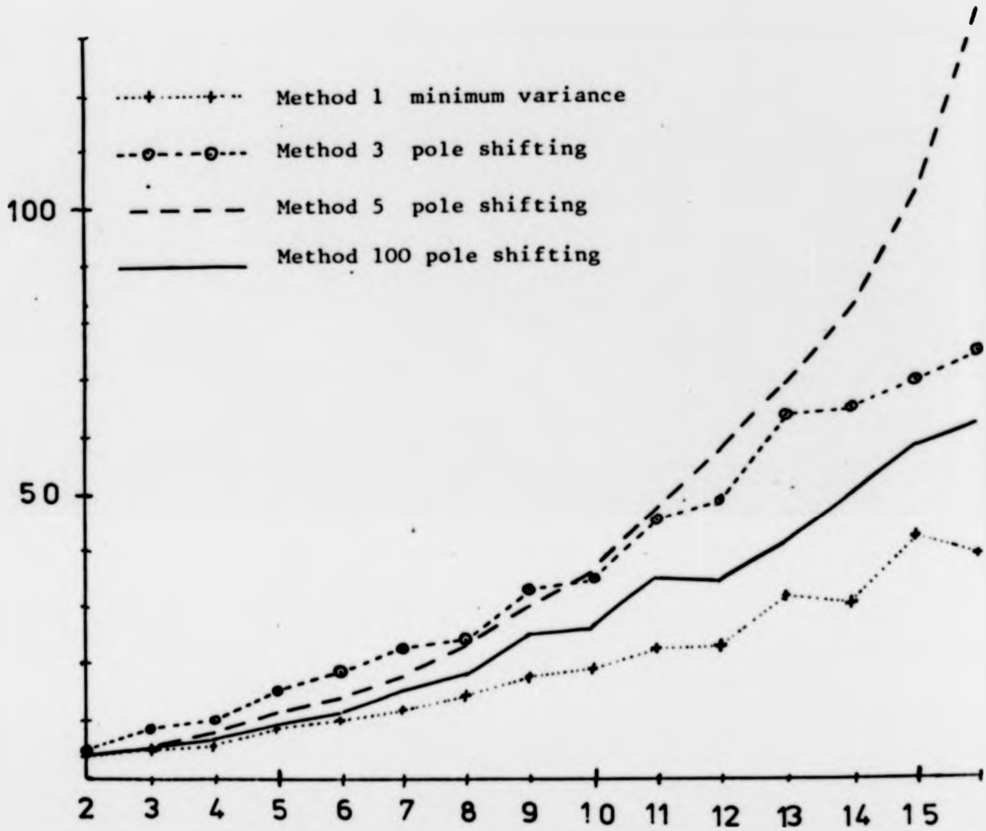
$$T(z^{-1}) = 1 - 0.606531z^{-1}.$$

Figures 6.6a, b, c and d show the resulting system outputs together with the setpoint. As would be expected the outputs with methods 5 and 100, which solve the complete set of linear equations at each sample time, were both very similar. The output using method 4 was similar to the outputs with methods 5 and 100, so the parameter estimates were following fairly closely. However, the output using method 3 looks different, indicating that in the initial steps the controller

Figure 6.9

General layout of the programs





<u>Figure 6.8</u>

C.P.U. time per iteration as a function of the number of parameters in the model, for each of the self tuning regulators. Note Method 2 will take the same time as Method 1 and Method 4 will take the same time as Method 3.

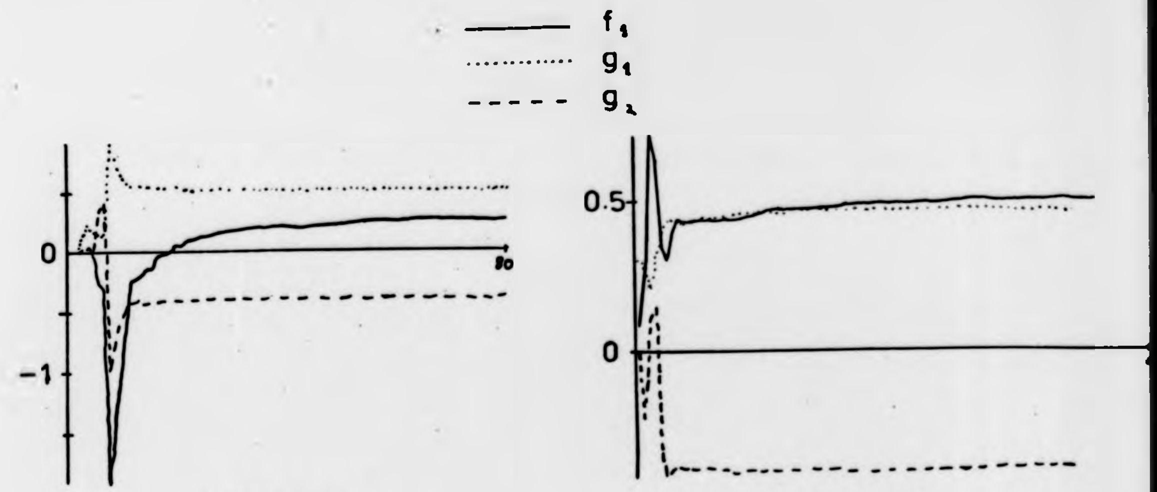


Figure 6.7a

Method 3

Figure 6.7b

Method 4

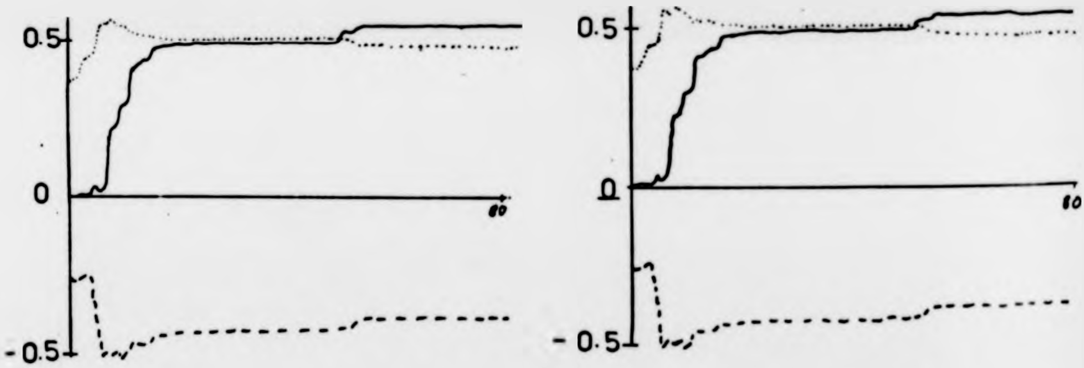


Figure 6.7c

Method 5

Figure 6.7d

Method 100

Figure 6.7

Parameter estimates using the four different versions of the general pole shifting regulator. The outputs which produced these estimates are shown in Figure 6.6.

estimates tend to be further from the proper values obtained by methods 5 and 100. Figure 6.7 a, b, c and d show the controller parameters in these four examples. It can be seen that the estimates with method 4 do follow the proper estimates more closely than those with method 3.

Figure 6.8 shows the computation time required for each of the methods 1, 3, 5 and 100 as a function of the number of parameters. The computation for method 2 is the same as for 1 and the time for method 4 is the same as that for 3. It can be seen that the fastest general pole shifting method was 100 which used a special program for solving the simultaneous equations. The methods 3, 4 and 5 took very similar C.P.U. times until there were more than 12 parameters being estimated, from then on the method 5 took much longer.

6.3 General Layout of Programs

Figure 6.9 is a flow chart of the basic set of programs used to produce the results in this and the previous chapter.

In order to minimize the amount of core storage being used at any time the program is split into several smaller programs, and only one of these runs at a time. When a particular program finishes it then uses a facility available on the Control System Centre PDP-10 to start the next program running. All communication between the separate programs is conducted through data files. The input program INR writes two files BEGIN.DAT and CREG.DAT, which are identical and contain all the required data for the simulation. This input program

Table 6.2 List of flow charts

Figure No.	Description	
	Name	Use
6.11	INR	Accept simulation data interactively
6.12	SETUP	Reading subroutine for INR
6.13	REC	Writing subroutine for INR
6.14	CAREG CSREG DTEST	Digital Simulation Program.
6.15	HAREG HSREG HTEST	Hybrid Simulation Program.
6.16	REGUL	Self tuning regulator for methods 1,2,3,4 used by CAREG & HAREG
6.17	STREG	Self tuning regulator for methods 5 & 100 used by CSREG, DTEST, HSREG and HTEST.
6.18	WRCLLS RECLLS	Least squares estimator used by the self tuning regulators.
6.19	SIMUL	Simulation subroutine used by the digital simulation programs.
6.20	RNOISE	White noise generator used for the simulations.
6.21	SINP	Subroutine to generate setpoint changes
6.22	DECR INCR	Subroutines to move data along arrays: used by many of the programs
6.23	PRODUC	Function to take an inner product of two vectors: used extensively
6.24	REGPL	Program to plot time responses
6.25	SUBPL	Subroutine called by REGPL to do the actual plotting.
6.26	READI or READS	Subroutine used by SUBPL
6.27	NQS	Nyquist Diagram Program
6.28	NQST	Nyquist Diagram subroutine called by NQS
6.29	RLO	Root locus program
6.30	RLOC	Root locus subroutine called by RLO
6.31	OREG	stopping program.

also writes a formated data file TEMP.DAT which contains the data for the simulation in a more readable form. The input program then starts the required simulation program running; this could be CAREG, CSREG, DTEST, HAREG, HSREG or HTEST. The simulation continues until the specified end time, or until the simulation is aborted. At specified intervals during the simulation time the system estimates, the control values, the input values, and the output values are recorded in a file TH1.DAT. On completion of the simulation the simulation program stores the final controller values, system estimates etc. in file CREG.DAT and then starts the program REGPL running. Program REGPL reads the file CREG.DAT, and then allows the choice of plotting the time responses or estimates or drawing Nyquist or root locus diagrams or stopping. To draw Nyquist diagrams the program REGPL is stopped and the program NQS is run. On completion, program NQS stops and runs the program REGPL. Similarly root locus diagrams are plotted by program RLO. These programs NQS and RLO read the required data from CREG.DAT. If the program REGPL receives a command to stop it runs the program OREG which reads CREG.DAT, copies TEMP.DAT into CR.DAT, adds a formated version of the final estimates to CR.DAT, renames the file BEGIN.DAT to a specified name given as a run identifier, and then renames the file CREG.DAT and stops. The file BEGIN.DAT or its renamed version are used to allow a run to be easily repeated by specifying them as the input data file on entry to the input program INR. Figures 6.11 - 6.31 give the flow charts for the various parts of the program. Table 6.2 gives a list of the flow charts with a short description of the function of each part. Flow charts have not been included for the following programs:

Subroutine RUNTIM(I): This is a subroutine which finds the current total C.P.U. time used in milliseconds and returns it as I.

Subroutine RPROG('NAME') This subroutine stops the current program and runs the named program.

Subroutines LOOK and Intest, which handle files and check for Teletype input.

There are also several programs out of the Hybrid computer library HYDMUL.REL, and from the Graphics package, GRAPHS.REL. for which no flow charts have been included.

FIGURE 4.11
PROGRAM INPUT

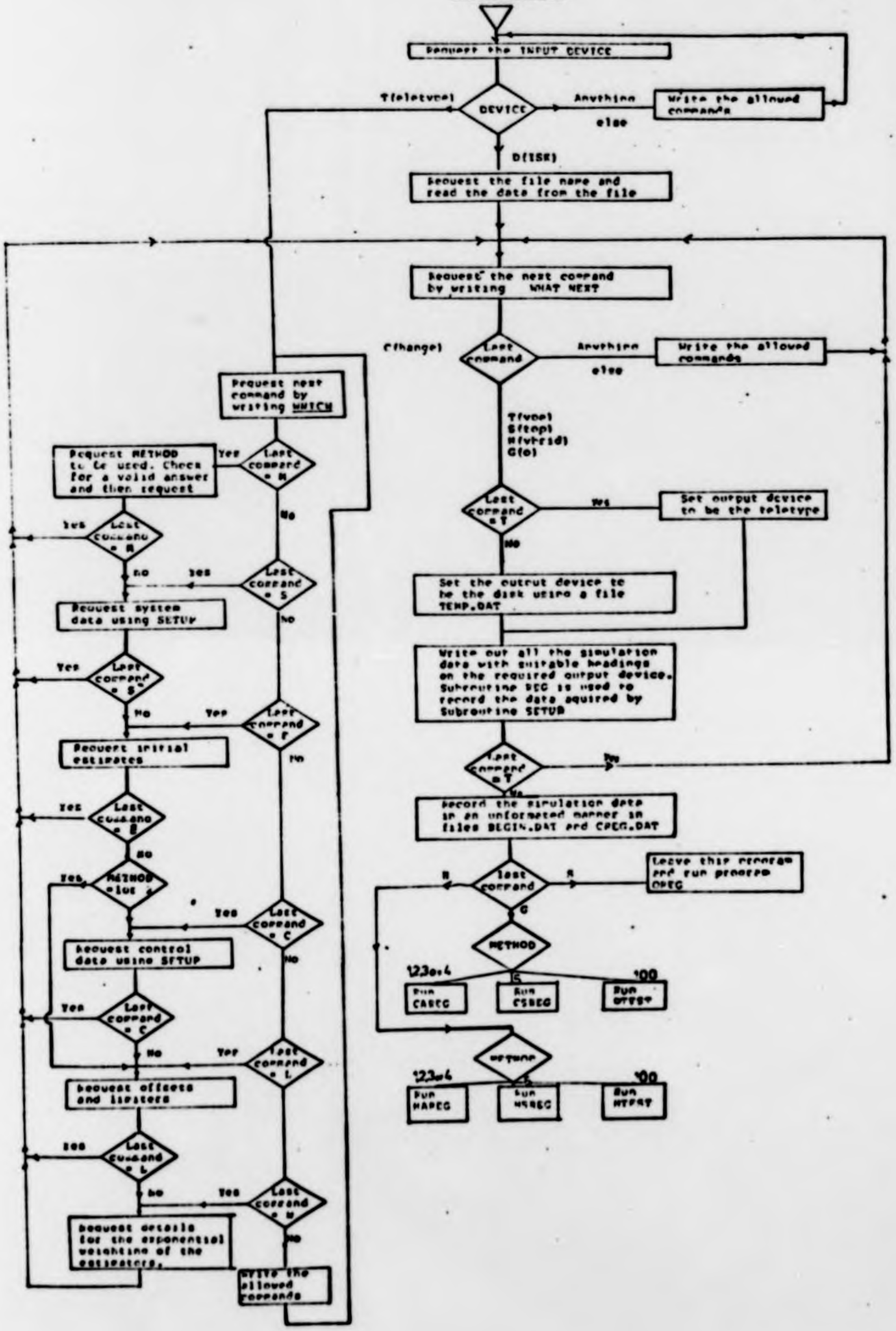


Figure 6.12

Subroutine SETUP

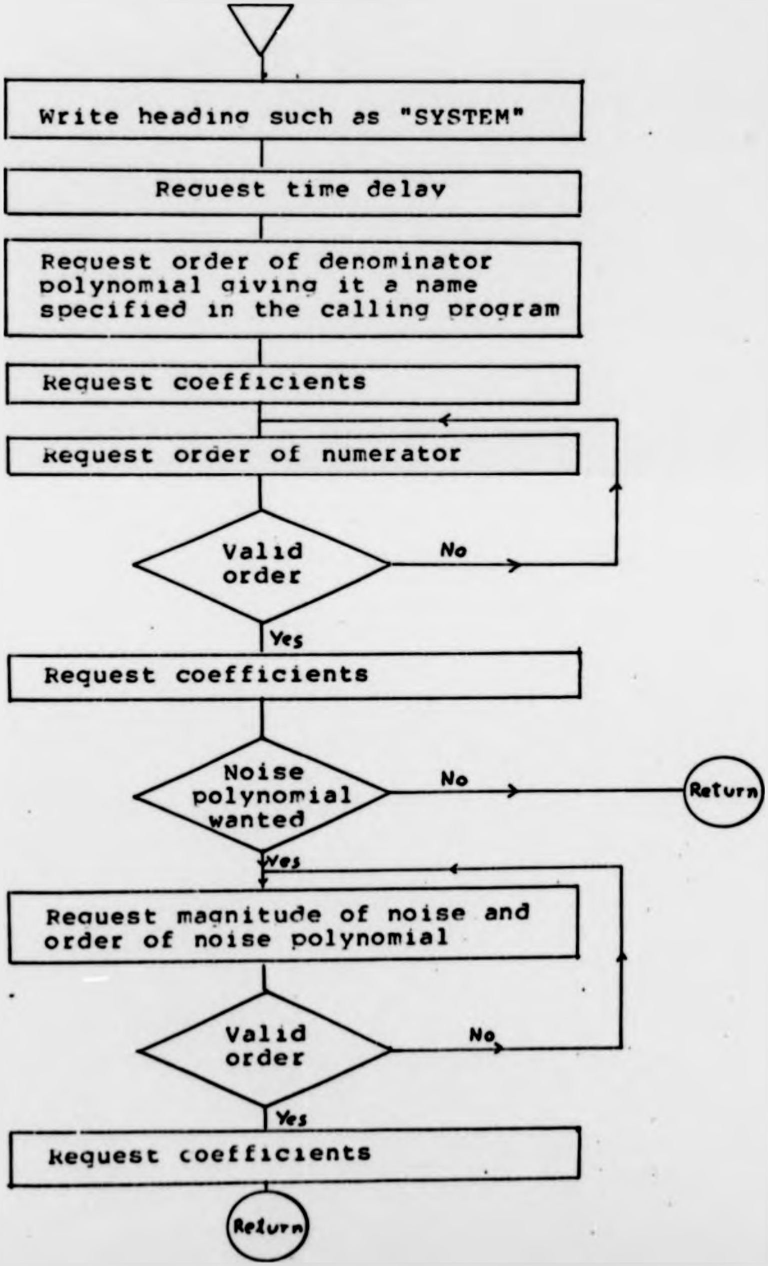


Figure 6.14

Programs CAREG,CSREG and DTEST

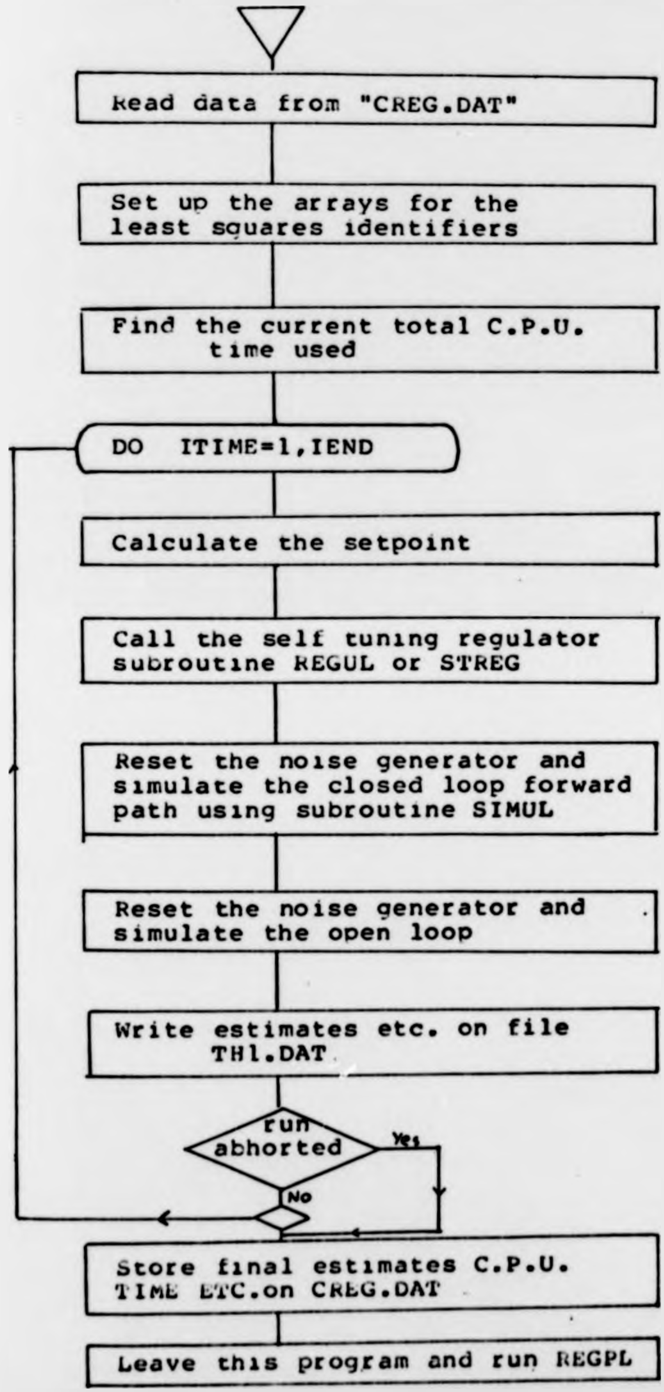


Figure 6.15

PROGRAMS HAREG,HSREG and HTEST

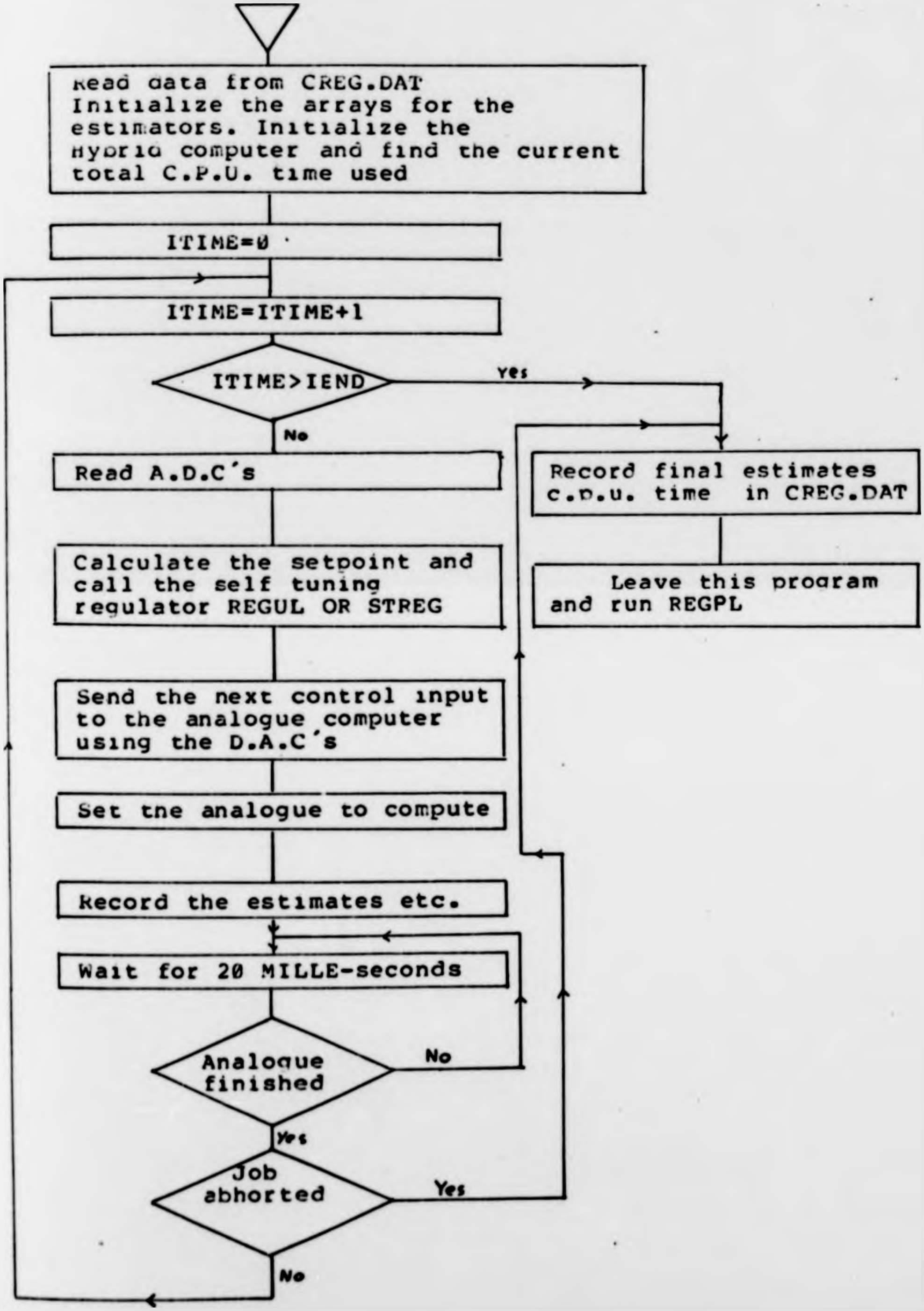


Figure 6.16

REGUL

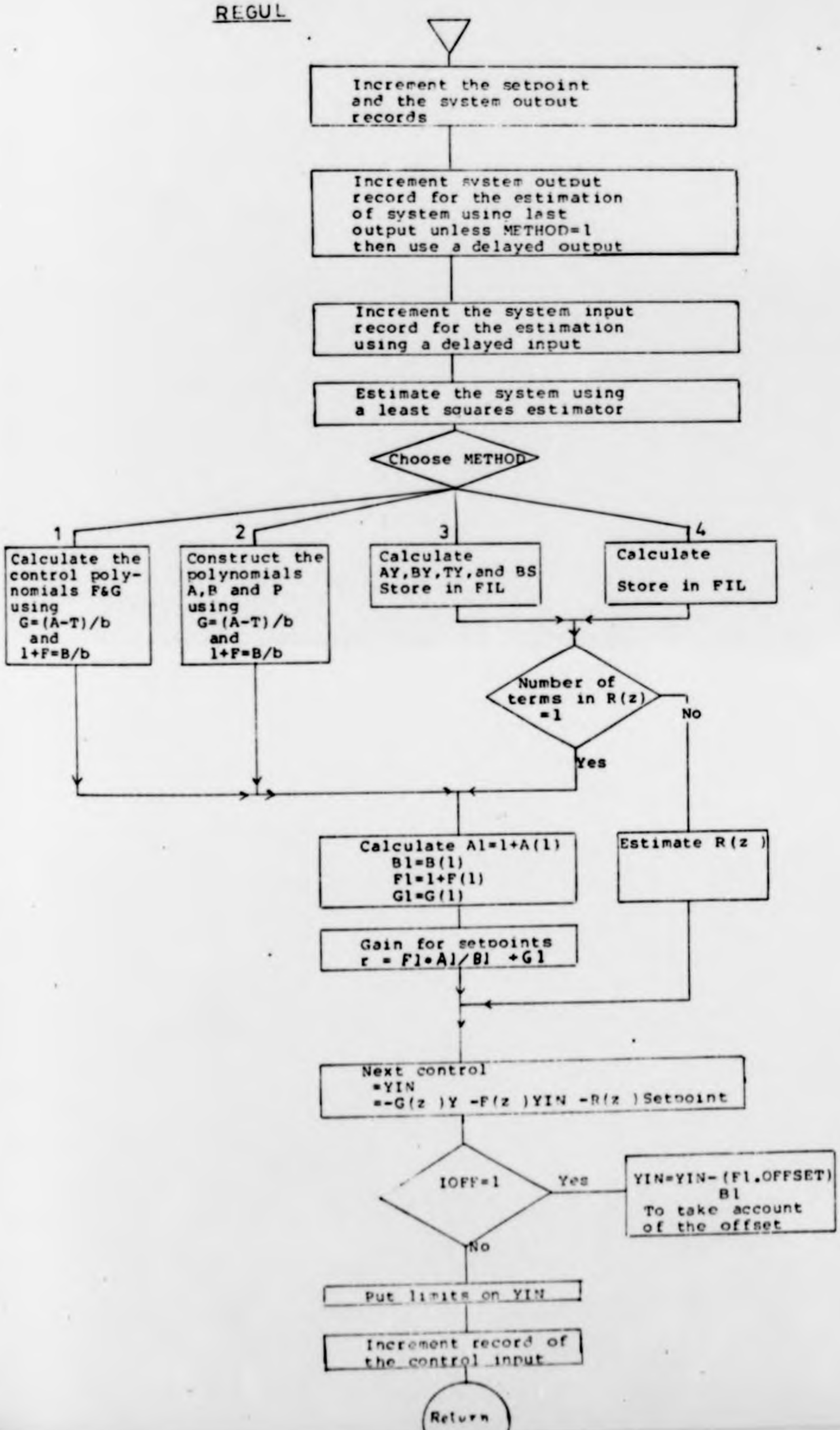


Figure 6.17

Self tuning regulator STREG used by CSREG,DTEST,HSREG,UTEST

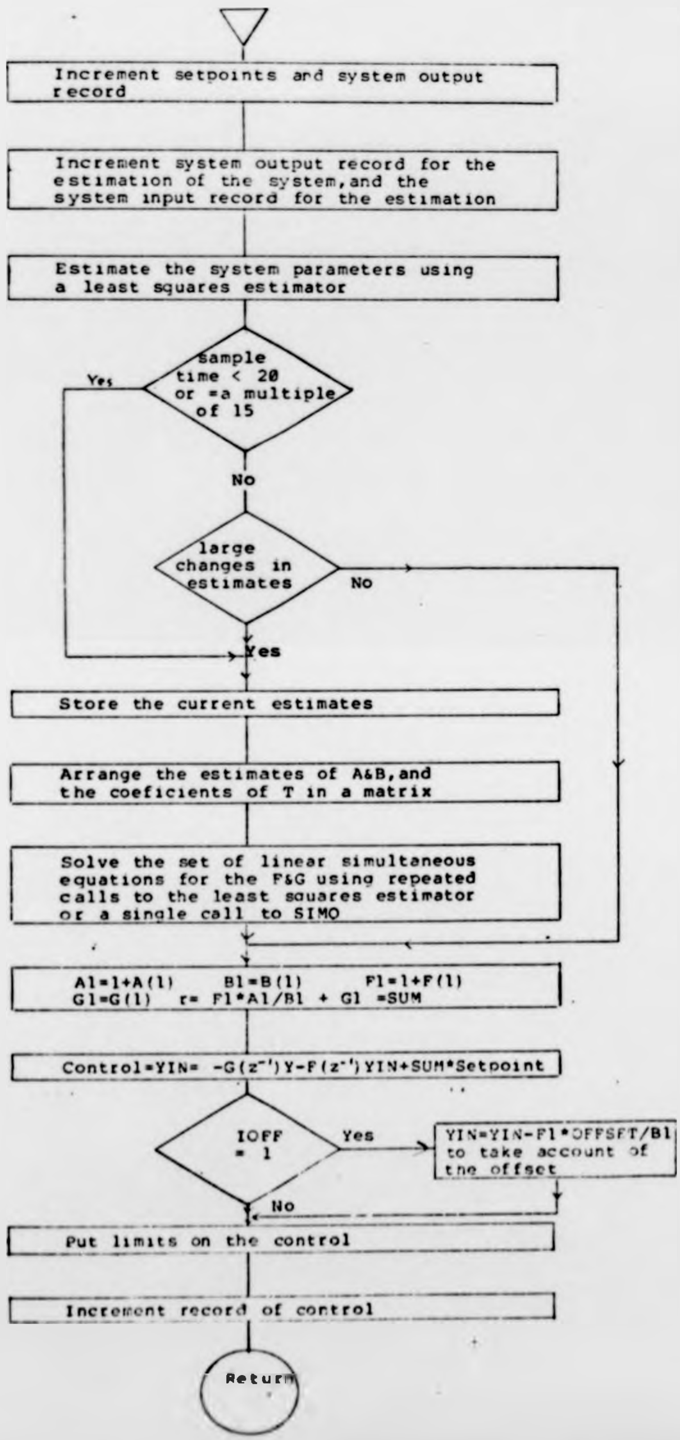
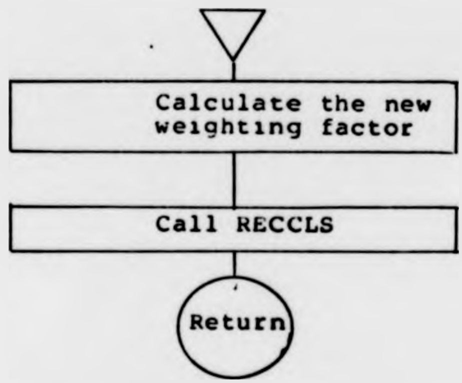


Figure 6.18

Least Squares Subroutines RECLLS and WRCLLS

WRCLLS



RECLLS

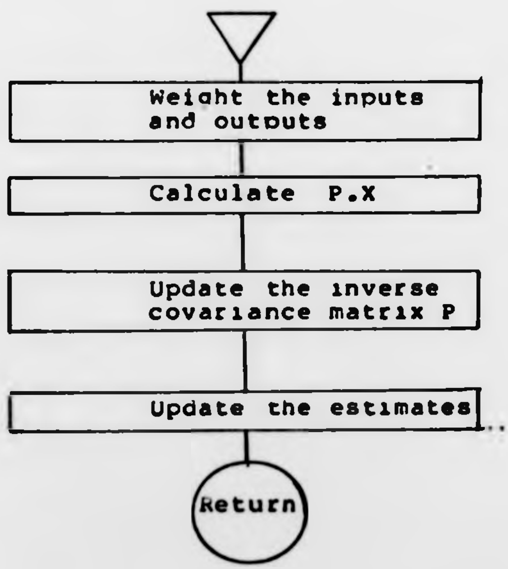


Figure 6.19

SIMUL

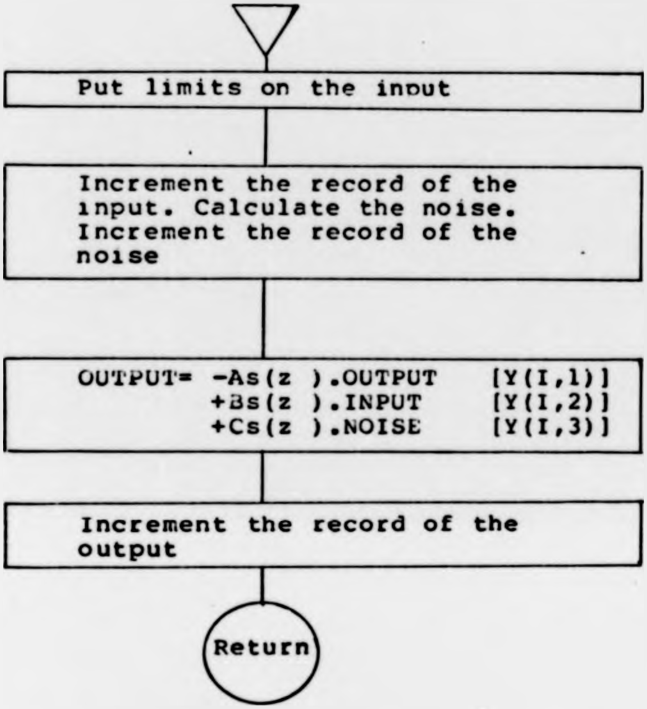


Figure 6.20

RNOISE To generate white noise

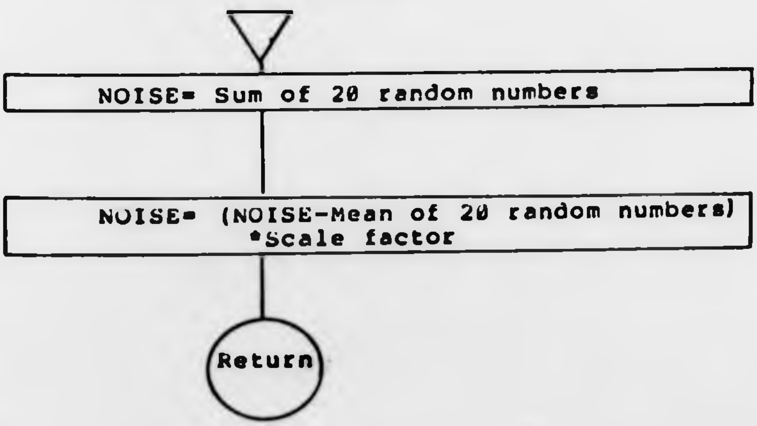


Figure 6.21

SINP to generate setpoints

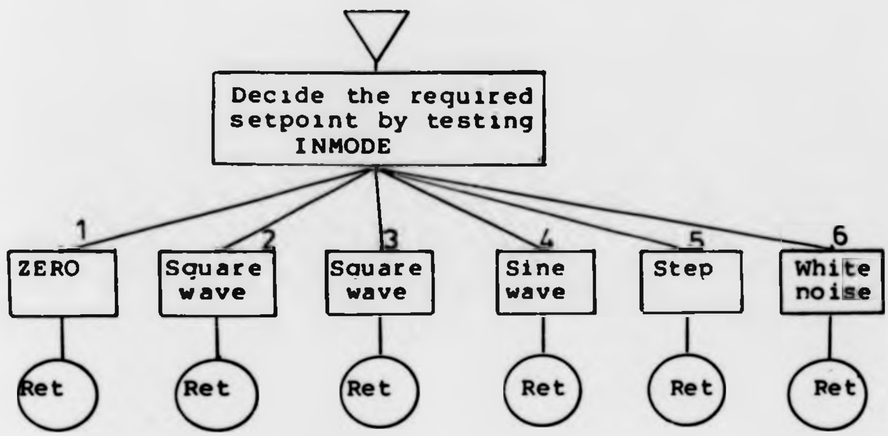


Figure 6.22

DECR and INCR To move data in an array

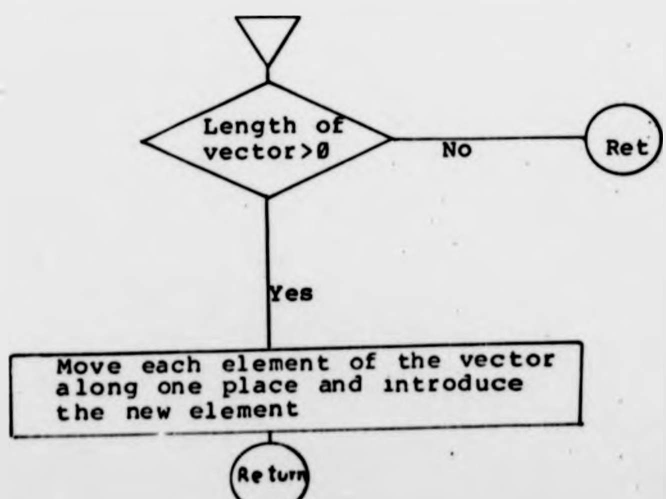


Figure 6.23

PRODUC   To form a scalar product of two vectors

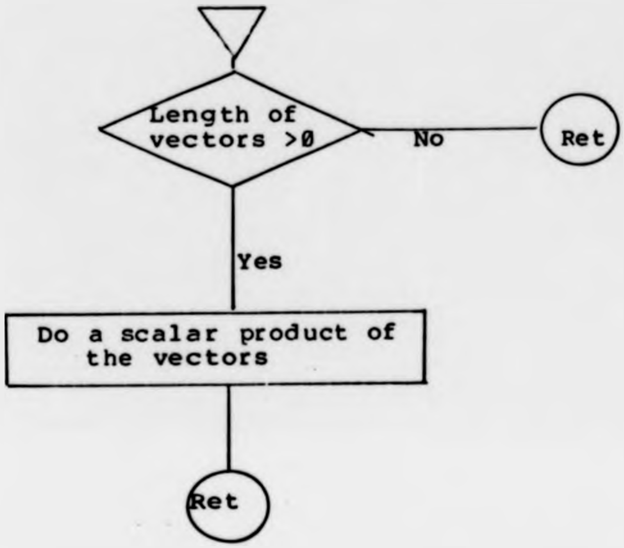


Figure 6.24

Program REGPL

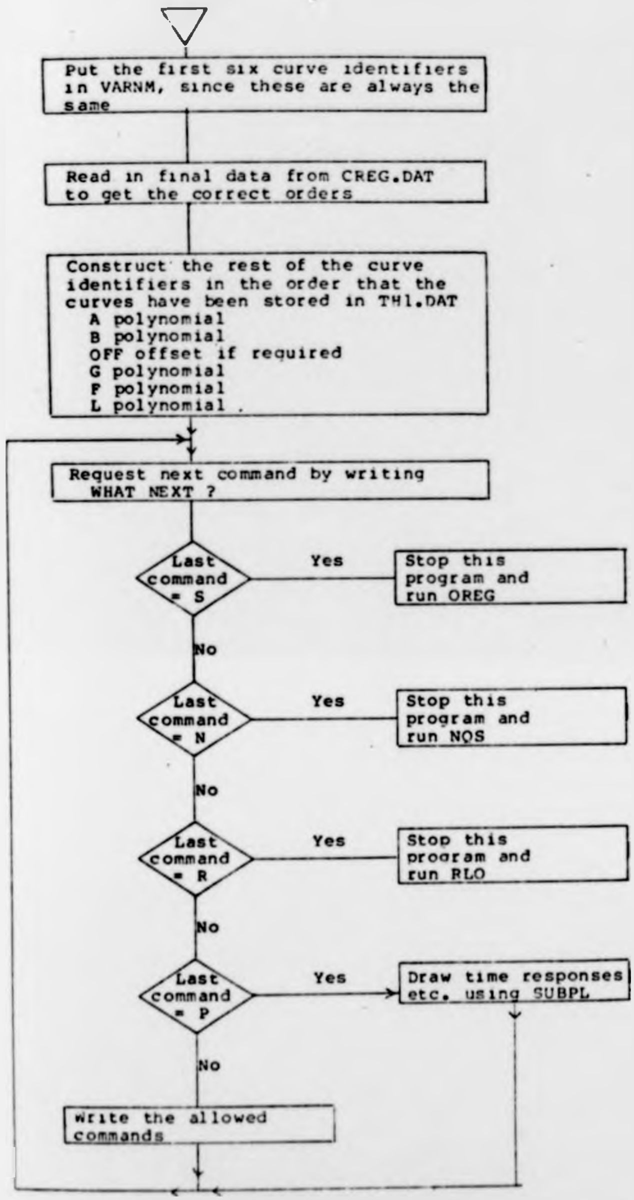


Figure 6.25

Subroutine SUBPL

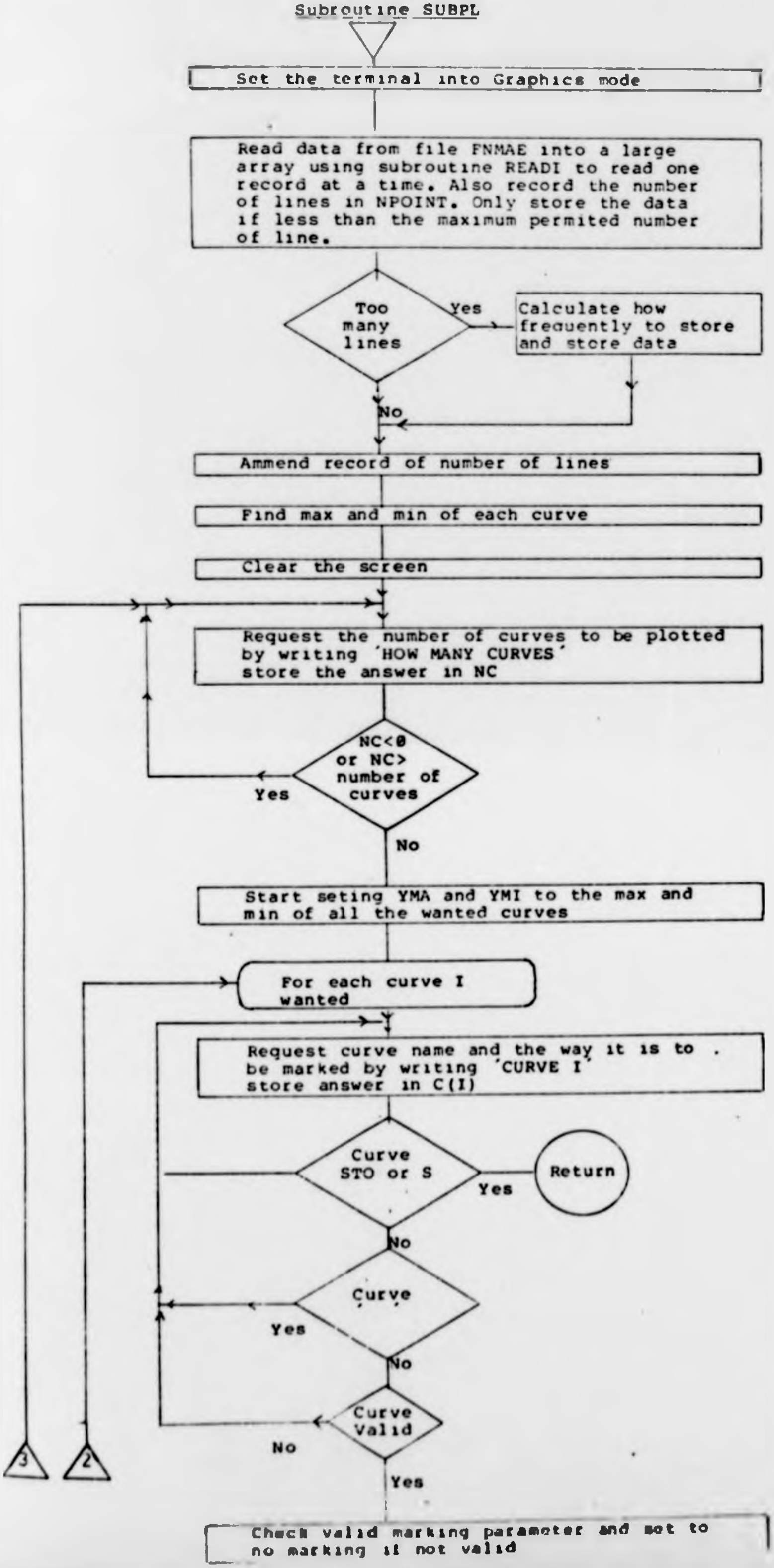


Figure 6.25 continued

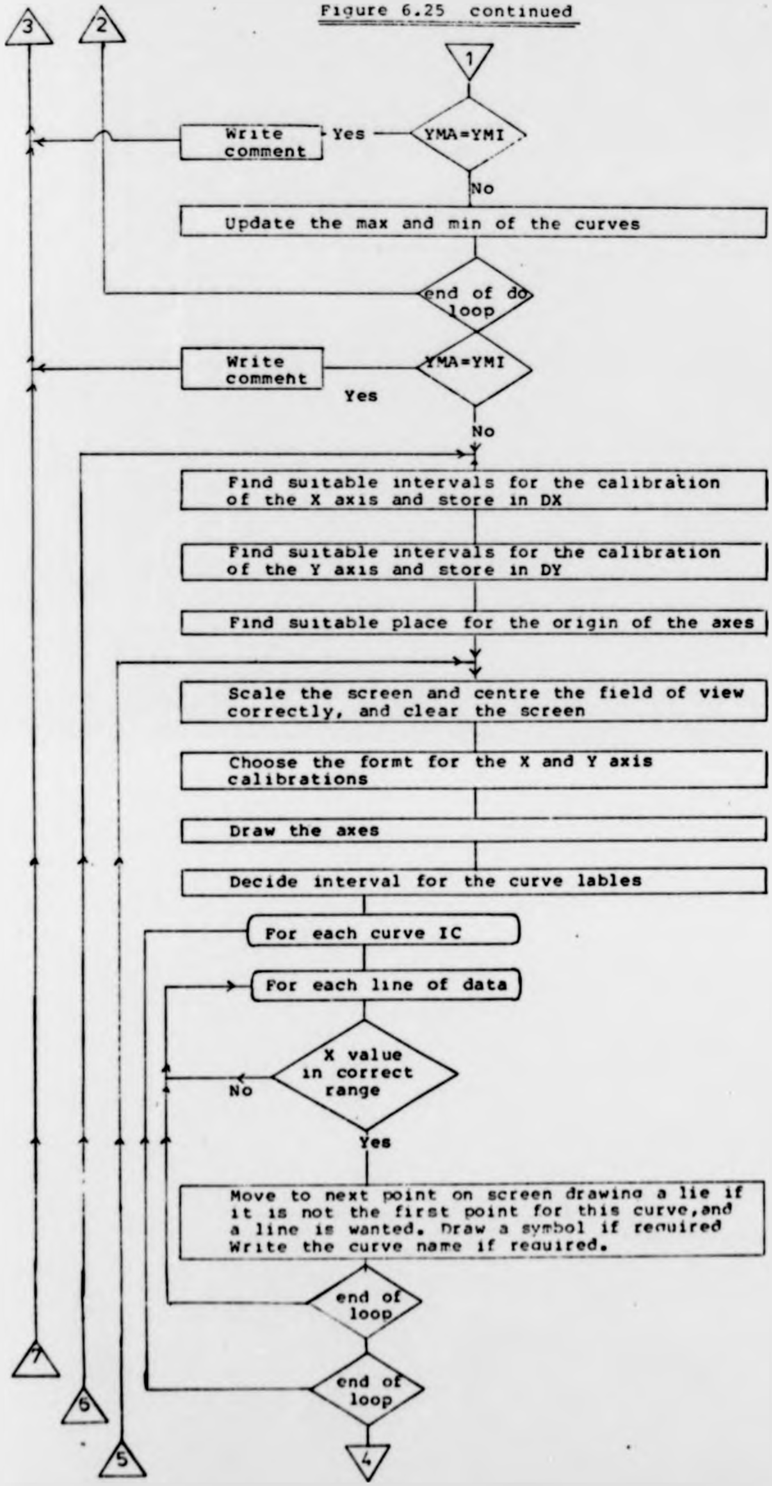


Figure 6.25 continued

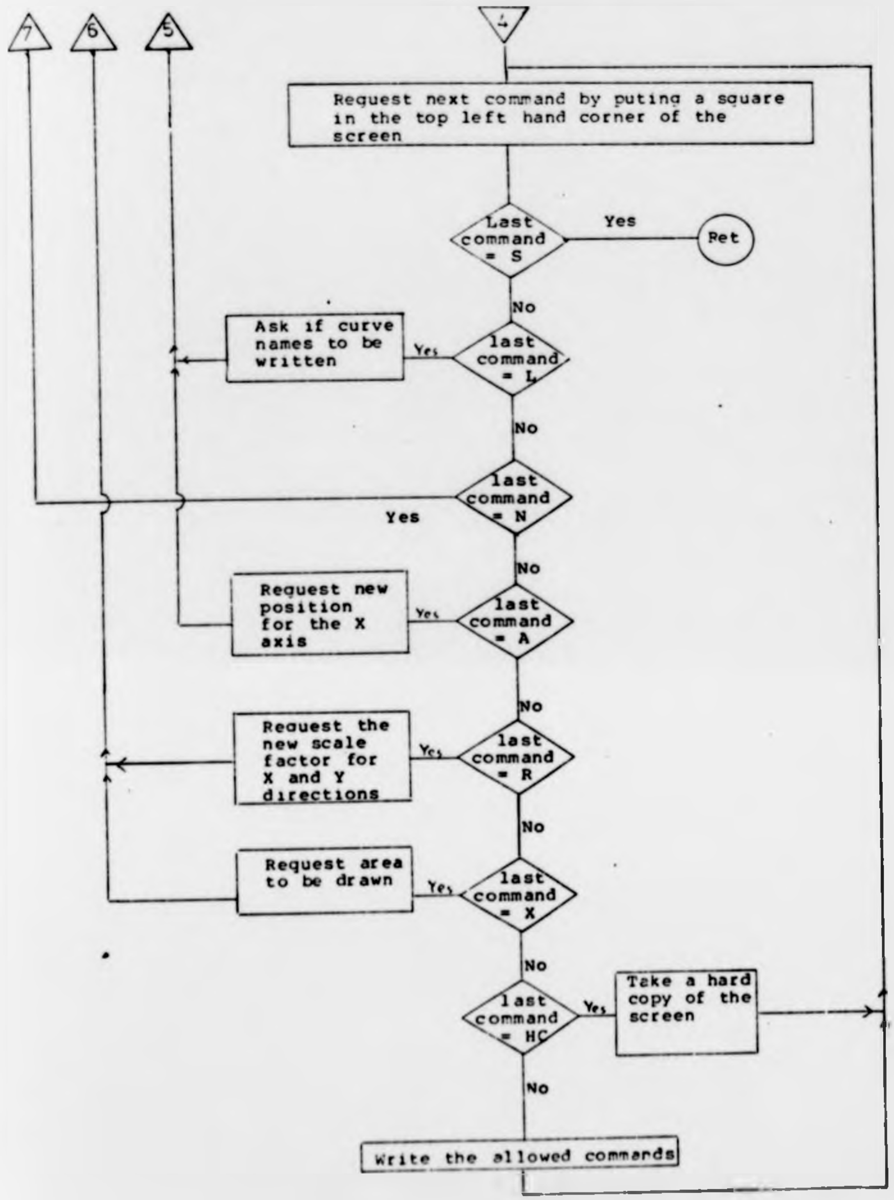


Figure 6.26

READS AND READI

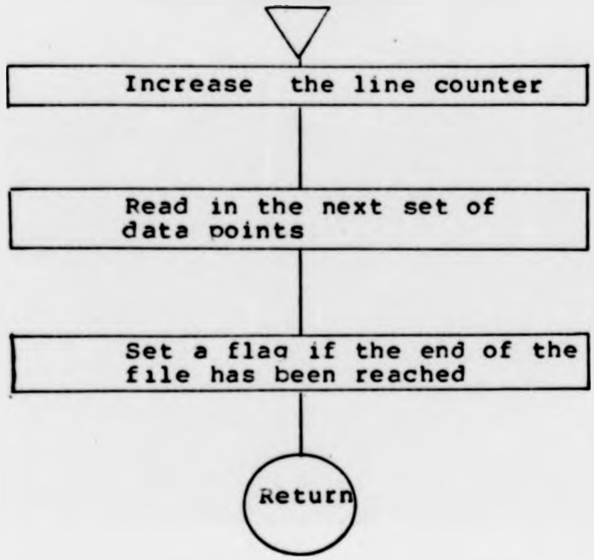


Figure 6.27

PROGRAM NQS

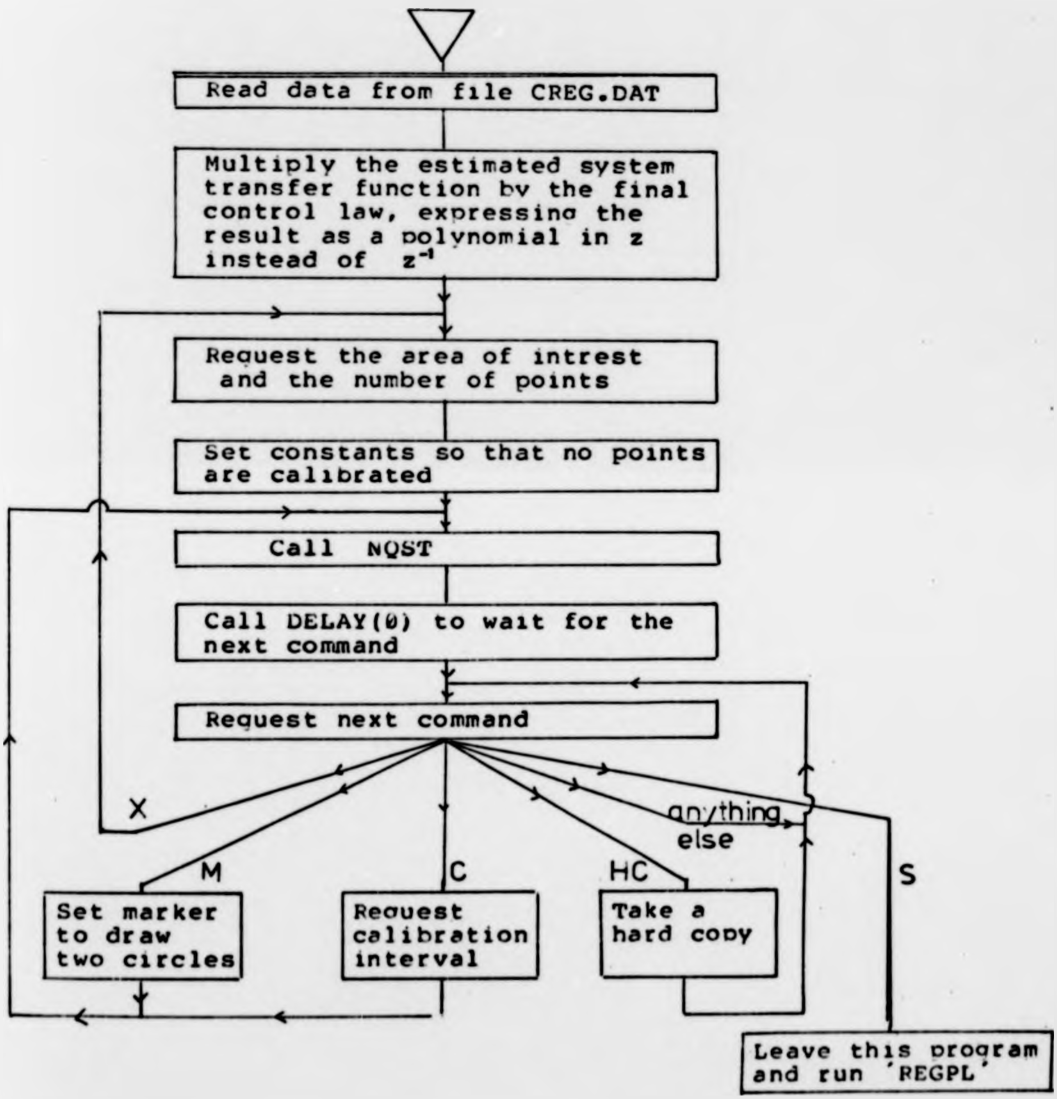


Figure 6.28

Subroutine NOST

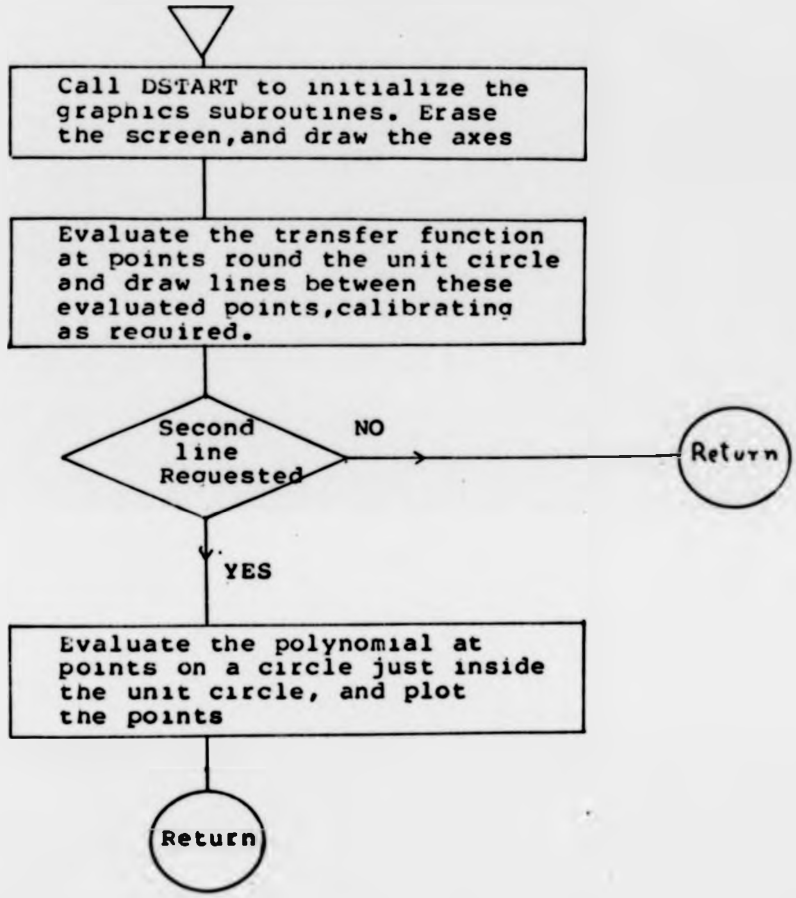


Figure 6.29

PROGRAM PLO

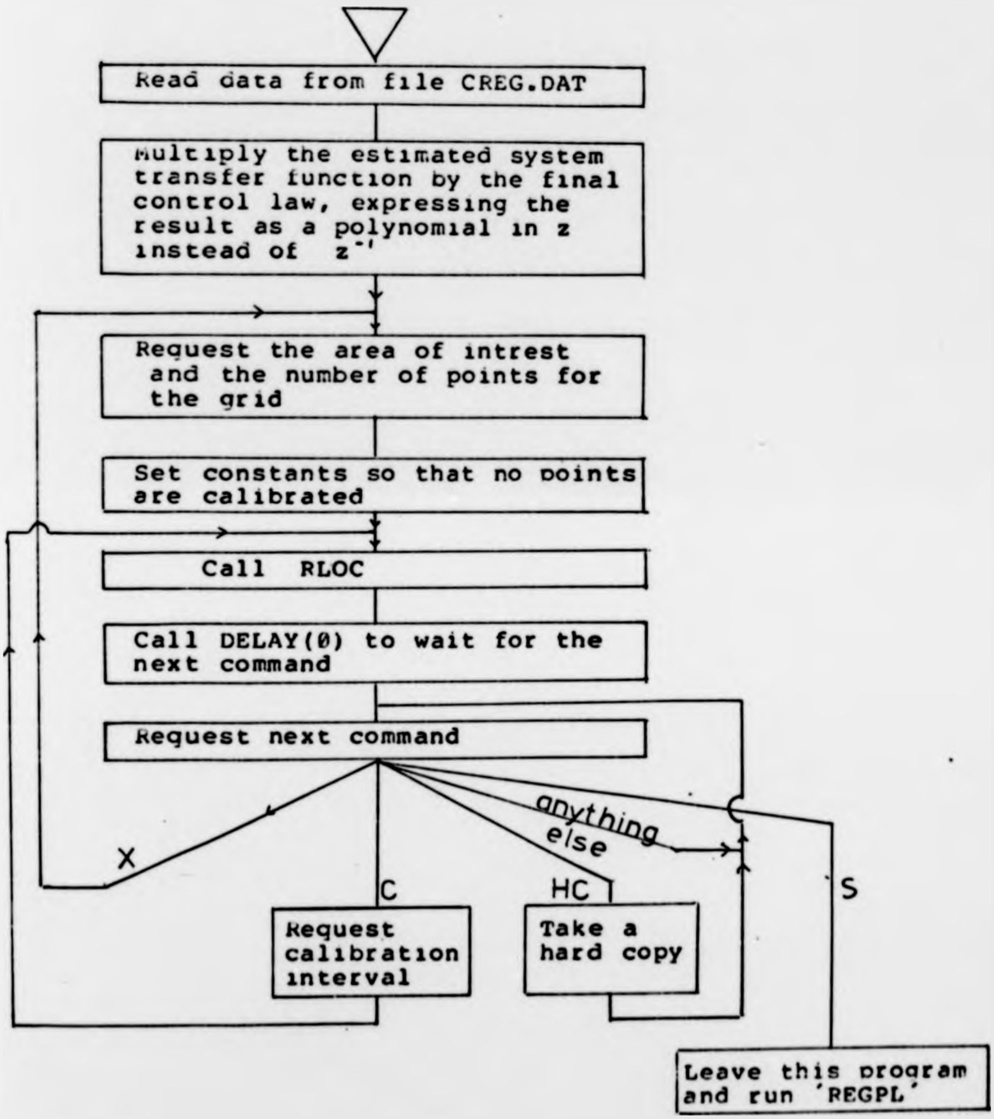


Figure 6.30

Subroutine RLOC

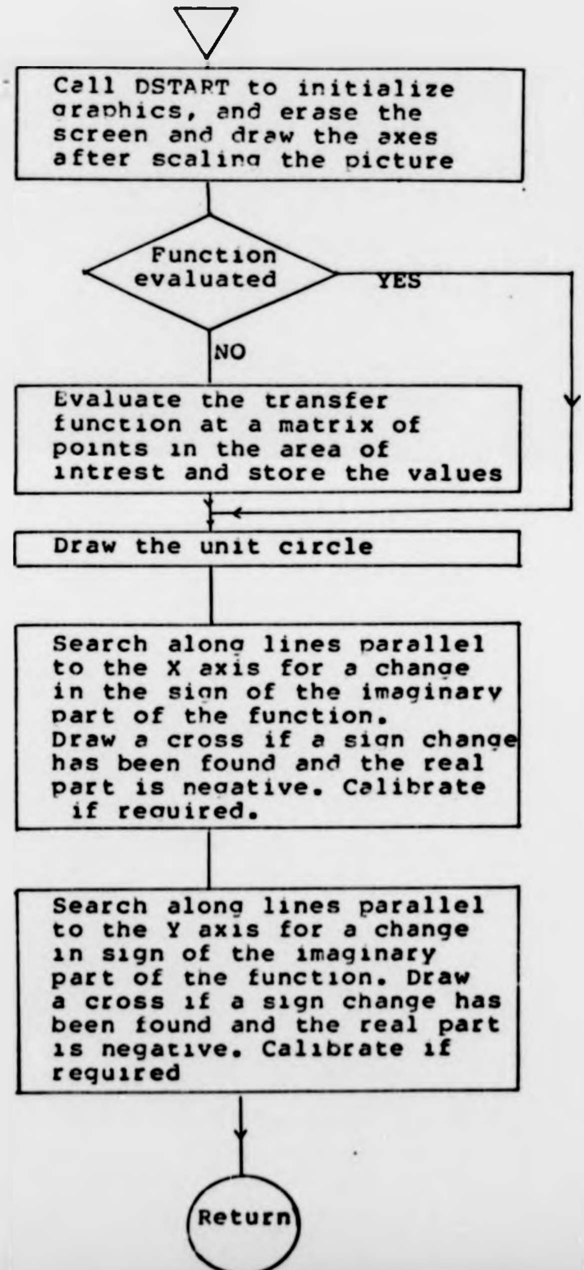
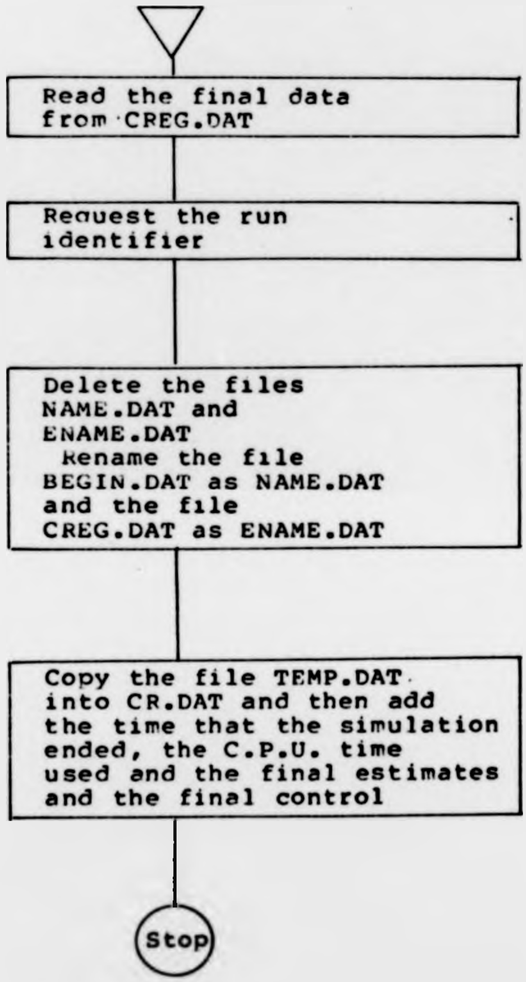


Figure 6.31

Program OREG



CHAPTER 7

Multivariable Self Tuning Regulators

There are several problems in extending the theory of the previous chapters to multivariable systems. Two of the main problems arise from the difficulty of choosing (i) a suitable form for the system model and, (ii) an appropriate regulator structure. A third problem arises when extending the regulator theory of Chapter 3 which relied on the commutivity of the z^{-1} polynomials. In this chapter a possible approach has been suggested. It may be that the theorems for the coloured noise situation can be extended to the multivariable case, but this has not been attempted in this brief study.

7.1 System model

Assuming that there are the same number of inputs as outputs, a suitable model structure for identification with a least squares estimator would be:

$$(I+A(z^{-1}))\,\underset{\sim}{Y}(t) = B(z^{-1})\,\underset{\sim}{U}(t) + \underset{\sim}{e}(t) \tag{7.1}$$

Where $A(z^{-1})$ is a matrix of polynomials in z^{-1} with no constant terms, $B(z^{-1})$ is a matrix of polynomials in z^{-1} with no constant terms, I is a unit matrix, and $\underset{\sim}{Y}(t)$, $\underset{\sim}{U}(t)$ and $\underset{\sim}{e}(t)$ are vectors of outputs, inputs and disturbances.

This form of model has more parameters than required but is convenient.

7.2 Regulator

A regulator structure which would allow easy calculation of the control signal would be

$$(I+F(z^{-1}))\,\underset{\sim}{U}(t) = -G(z^{-1})\,\underset{\sim}{Y}(t) \qquad (7.2)$$

Where $F(z^{-1})$ is a matrix of polynomials in z^{-1} with no constant terms and $G(z^{-1})$ is matrix of polynomials in z^{-1} with constant terms allowed.

7.3 Closed Loop equation

Combining equations 7.1 and 7.2 gives a closed loop equation

$$(I+A)\underset{\sim}{Y}(t) + B(z^{-1})(I+F(z^{-1}))^{-1}\, G(z^{-1})\, \underset{\sim}{Y}(t) = \underset{\sim}{e}(t) \qquad (7.3)$$

7.4 Equivalent of the minimum variance control law of section 3.3.1

The minimum variance control law of section 3.3.1 was obtained by setting $\underset{\sim}{Y}(t) = \underset{\sim}{e}(t)$. This implies the control matrices F and G satisfy

$$(I+A) + B(z^{-1})(I+F(z^{-1}))^{-1}\, G(z^{-1}) = I \qquad (7.4)$$

$$\therefore \quad A + B(z^{-1})(I+F(z^{-1}))^{-1}\, G(z^{-1}) = 0 \qquad (7.5)$$

$B(z^{-1})$ can be expressed as a sum of matrices

$$B(z^{-1}) = z^{-1}\, B_1 + z^{-2}\, B_2 + \ldots + z^{-h}\, B_n \qquad (7.6)$$

where each of the B_i is a matrix of constants.

If the matrix B_1 is non singular, F can be chosen such that:

$$(I+F(z^{-1})) = z\, B_1^{-1}\, B(z^{-1}) \qquad (7.7)$$

F can be chosen this way since equation 7.6 implies that

$$zB_1^{-1}\, B(z^{-1}) = I + z^{-1}B_1^{-1}\, B_2 + \ldots + z^{-n+1}\, B_1^{-1}\, B_n$$

With this choice of F matrix, equation 7.5 becomes

$$A(z^{-1}) + B(z^{-1})(z\, B_1^{-1}\, B(z^{-1}))^{-1}\, G(z^{-1}) = 0$$

$$\therefore \quad A(z^{-1}) + z^{-1}\, B(z^{-1})\, B(z^{-1})^{-1}\, B_1\, G(z^{-1}) = 0$$

$$A(z^{-1}) + z^{-1}\, B_1\, G(z^{-1}) = 0$$

$$\therefore \quad G(z^{-1}) = -z\, B_1^{-1}\, A(z^{-1}) \qquad (7.8)$$

∴ If F & G are chosen such that 7.7 and 7.8 are satisfied the closed loop equation reduces to:

$$\underset{\sim}{Y}(t) = \underset{\sim}{e}(t)$$

The calculation of this control law would require the inversion of the matrix B_1 at each stage. This regulator would be similar to one constructed by Borisson[1] which was an extension to the multivariable case of the self tuning regulator suggested by Astrom for minimum phase systems.[2]

7.5 Equivalence of the tailored control law of section 3.6.3

For this control law the control is chosen such that

$$T(z^{-1})\, \underset{\sim}{Y}(t) = \underset{\sim}{e}(t) \qquad (7.9)$$

where $T(z^{-1})$ is a matrix of polynomials with no constant terms plus the unit matrix. Equation 7.3 implies that F and G must be chosen so that

$$(I+A(z^{-1})) + B(z^{-1})(I+F(z^{-1}))^{-1}\, G(z^{-1}) = T(z^{-1}) \qquad (7.10)$$

A solution for this in the case where B_1 is non singular is

$$(I+F(z^{-1})) = z\, B_1^{-1}\, B(z^{-1}) \qquad (7.11)$$

and by substituting

$$(I+A(z^{-1})) + z^{-1}\, B_1\, G(z^{-1}) = T(z^{-1})$$

$$G(z^{-1}) = z\, B_1^{-1}\, (T(z^{-1}) - I - A(z^{-1})) \qquad (7.12)$$

7.6 Equivalence of the pole shifting law of section 3.4

It will be noticed that the control laws of the previous two sections will be unstable if $B(z^{-1})$ has any zeros outside the unit circle. This instability can be avoided by a similar approach to that used in the single input single output case.

In the single input case the closed loop equation was

$$((1+A(z^{-1}))(1+F(z^{-1})) + B(z^{-1})\, G(z^{-1}))\, Y = (1+F)z^{-1}\, e \qquad (7.13)$$

Unfortunately to reach the equation 7.13 from the single input version of equation 7.3, the $(1+F(z^{-1}))^{-1}$ has to commute with $B(z^{-1})$, and so when B and F are matrices the same approach cannot be used. However, the inverse of the polynomial matrix $B(z^{-1})$ can be written as a product of a polynomial matrix $B'(z^{-1})$ and the inverse of a single polynomial in z^{-1} called $B^*(z^{-1})$

$$B(z^{-1})^{-1} = B^*(z^{-1})^{-1}\, B'(z^{-1}) \qquad (7.14)$$

$$B^*(z^{-1})^{-1} = \frac{1}{\text{Polynomial in } (z^{-1})} = \frac{1}{B^*(z^{-1})}$$

For examples for a two input two output system

$$B(z^{-1}) = \begin{bmatrix} B_{11}(z^{-1}) & B_{12}(z^{-1}) \\ B_{21}(z^{-1}) & B_{22}(z^{-1}) \end{bmatrix}$$

$$\text{and } (B(z^{-1}))^{-1} = \frac{1}{B_{11}(z^{-1})\, B_{22}(z^{-1}) - B_{12}(z^{-1})\, B_{21}(z^{-1})} \begin{bmatrix} B_{22}(z^{-1}) & -B_{12}(z^{-1}) \\ -B_{21}(z^{-1}) & B_{11}(z^{-1}) \end{bmatrix}$$

$$\therefore \quad B'(z^{-1}) = \begin{bmatrix} B_{22}(z^{-1}) & -B_{12}(z^{-1}) \\ -B_{21}(z^{-1}) & B_{11}(z^{-1}) \end{bmatrix} \quad \text{and} \quad B^*(z^{-1}) = B_{11}(z^{-1}) . B_{22}(z^{-1}) - B_{12}(z^{-1}) . B_{21}(z^{-1})$$

Substituting for $B(z^{-1})$ in (7.3) gives

$$(I+A(z^{-1}))\underline{Y} + (B^*(z^{-1})^{-1} B'(z^{-1}))\ (I+F)^{-1} G(z^{-1})\underline{Y} = \underline{e} \qquad (7.15)$$

Multiplying both sides by $B(z^{-1})^{-1}$ gives

$$B^*(z^{-1})^{-1} B'(z^{-1})(I+A)\underline{Y} + (I+F(z^{-1}))^{-1} G(z^{-1})\underline{Y}$$

$$= B^*(z^{-1})B'(z^{-1})\ \underline{e} \qquad (7.16)$$

$$\therefore \quad (I+F(z^{-1}))\ B^*(z^{-1})^{-1} B'(z^{-1})(I+A)\underline{Y} + G(z^{-1})\underline{Y}$$

$$= (I+F(z^{-1}))\ B^*(z^{-1})^{-1} B'(z^{-1})\ \underline{e} \qquad (7.17)$$

but B^{*-1} will commute with the matrix $I+F(z^{-1})$ since it will commute with each term in it, because $B^*(z^{-1})$ is just a polynomial in (z^{-1}) .

$$\therefore \quad B^*(z^{-1})^{-1}(I+F(z^{-1}))\ B'(z^{-1})(I+A(z^{-1}))\underline{Y} + G(z^{-1})\underline{Y}$$

$$= B^*(z^{-1})^{-1}.(I+F(z^{-1}))\ B'(z^{-1})\ \underline{e} \qquad (7.18)$$

Multiplying both sides by $B^*(z^{-1})$ gives

$$(I+F(z^{-1}))\ B'(z^{-1})(I+A(z^{-1}))\underline{Y} + B^*(z^{-1})G(z^{-1})\underline{Y}$$

$$= (I+F(z^{-1}))\ B'(z^{-1})\ \underline{e} \qquad (7.19)$$

Equation (7.19) is the closed-loop equation for the system. Therefore the system will be stable if

$$(I+F(z^{-1}))\ B'(z^{-1})(I+A(z^{-1})) + B^*(z^{-1})\ G(z^{-1}) = T(z^{-1}) \qquad (7.20)$$

where $T(z^{-1})$ is a polynomial matrix as in the previous section, with all its zeros inside the unit circle. Therefore, if F and G are chosen to satisfy (7.20) the regulator will be a general stable pole shifting law. In the single input case $B^*(z^{-1}) = B(z^{-1})$ and

$B'(z^{-1}) = 1$ and equation (7.20) reduces to the equation for a general single input pole shifting law of Chapter 3.

CHAPTER 8 CONCLUSIONS

1) When a system is being identified in a closed loop situation a least squares estimator will not produce unique estimates unless the order of the controller is greater than the order of the system (Chapter 4).

2) A least squares estimator can be combined with a simple pole shifting control law to form a controller which tunes itself to take account of the dynamic behaviour of the system (Chapter 5).

3) One particular case of the pole shifting control law is a minimum variance control law as used by Astrom to form a self tuning regulator (Chapter 5).

4) The special case using the minimum variance control law can be unstable, even when the plant Laplace transfer function is minimum phase, since the z-transform's zeros depend on the plant poles as well as the zeros (Chapter 2).

5) The special case using the minimum variance control law can be conditionally stable if the system is of order three or more, or if there is a time delay of more than one half of the sample period (Chapter 2).

6) When a plant has time delays and the minimum variance control law is being used it is better to estimate the correct order of model, and extend for the control law than to use a higher order model as proposed by Astrom (Chapter 6).

7) A self tuning regulator with the general pole shifting law can easily be conditionally stable, i.e. if the loop gain is less than a lower value or greater than an upper value, instability may occur. It shall, however, always be possible to establish a stable region for a linear system as demonstrated in Chapters 3 or 5.

8) It may be possible to extend these self tuning controllers to a multivariable system (Chapter 7).

8.1 Suggestions for further work

There are three main areas for further work.

1) To investigate the effects of nonlinearities other than limiters.

2) To apply these self tuning regulators to real systems. This was attempted but problems about the availability of certain pieces of machinery prevented success.

3) Multivariable self tuning regulators could be attempted using a general pole shifting law.

APPENDIX A

Notation

Polynomials in the delay operator z^{-1}

$A(z^{-1}) = a_1 z^{-1} + \ldots + a_{na} z^{-na}$	- The open loop transfer function of the system has a denominator equal to $1+A(z^{-1})$.
$B(z^{-1}) = b_1 z^{-1} + \ldots + b_{nb} z^{-nb}$	- Numerator of the open loop transfer function of the system.
$C(z^{-1}) = 1 + C_1 z^{-1} + \ldots + C_{nc} z^{-nc}$	- This is filter on the disturbance, allowing the modelling of coloured noise disturbances.
$F(z^{-1}) = f_1 z^{-1} + f_2 z^{-2} + \ldots + f_{nf} z^{-nf}$	- The denominator of the control law is $1+F(z^{-1})$.
$G(z^{-1}) = g_o + g_1 z^{-1} + \ldots + g_{ng} z^{-ng}$	- The numerator of the control law.
$P(z^{-1}) = 1 + p_1 z^{-1} + \ldots + p_r z^{-k}$	- Polynomial introduced in 2.5 such that: $P(z^{-1})(1+A(z^{-1})) = 1 + z^{-k} A'(z^{-1})$
$R(z^{-1}) = r_o + r_1 z^{-1} + \ldots + r_{nr} z^{-nr}$	- Acts on setpoints
$T(z^{-1}) = 1 + t_1 z^{-1} + \ldots + t_{nt} z^{-nt}$	- Closed loop characteristic polynomial.
$A_s(z^{-1}), B_s$	Similar to the model $A(z^{-1})$ and $B(z^{-1})$, but with coefficients which fit the system exactly.
$A'(z^{-1})\ B'(z^{-1})$	Similar to $A(z^{-1})$ and $B(z^{-1})$ but modified for time delays.

APPENDIX A (contd.)

$T'(z^{-1})$	Polynomial used in the calculation of the control law. This is the same as $T(z^{-1})$ if the system is linear.
$B^*(z^{-1})$	A polynomial used in the multivariable case.

<u>Scalars</u>

d	A constant offset on the system
e	A white noise disturbance.
k	Time delay as a multiple of sample interval.
ℓ	constant in calculation of control law.
na, nb, nc, nf, ng, nr, nt	highest power of z^{-1} in the polynomials A,B,C,F,G,R and T
q	Forward shift operator
setpoint	setpoint for system
t	time
U(t)	Input to system
s	Laplace operator
Y(t)	Output of system
z^{-1}	Delay operator
Δt	Sample interval
τ	time delay
B_i	zeros of the open loop system.

System equation without time delays

$$(1 + A(z^{-1}))\ Y = B(z^{-1})\ U + C(z^{-1})\ e$$

System equation with time delays

$$(1 + A(z^{-1}))\ Y = z^{-k}\ B(z^{-1})\ U + C(z^{-1})\ e$$

or $$(1 + z^{-k}\ A'(z^{-1}))\ Y = z^{-k}\ B'(z^{-1})\ U + P(z^{-1})\ C(z^{-1})\ e$$

where $(1 + A(z^{-1}))\ P(z^{-1}) = (1 + z^{-k}\ A'(z^{-1}))$

and $B(z^{-1})\ P(z^{-1}) = B'(z^{-1})$

Control equation

$$(1 + F(z^{-1}))\ U = G(z^{-1})\ Y + R(z^{-1})\ \text{setpoint}$$

control chosen by

$$(1 + A(z^{-1}))(1 + F(z^{-1})) + z^{-k}\ B(z^{-1})\ G(z^{-1}) = T'(z^{-1})$$

APPENDIX B

To obtain the minimum variance stable controller

Astrom and Wittenmark in their paper reference 3 from Chapter 3, state that the minimum variance control is given by:

$$U = \frac{G^*(q^{-1}) \quad y(t)}{F^*(q^{-1})\ B^{+*}(z^{-1})} \tag{B.1}$$

where $F^*(q^{-1})$ and $G^*(q^{-1})$ satisfy

$$B^-(q^{-1}) = A^*(q^{-1})\ F^*(q^{-1}) + q^{-k}\ G^*(q^{-1})\ B^{-*}(q^{-1}) \tag{B.2}$$

and $B^-(x)B^+(x)$ is a factorization of the polynomial $B(x)$ such that $B^+(x)$ has all its zeros inside the unit circle and $B^-(x)$ has all its zeros outside or on the unit circle, and $B^-(o) = 1$.

In this notation the system is represented by

$$A^*(q^{-1})\ y(t) = B^*(q^{-1})\ U(t-k) + \varepsilon(t)$$

The * on the polynomials indicates an operation on the polynomial.

if $\quad A(x) = a_o x^n + a_1 x^{n-1} + \ldots + a_n$

then $\quad A^*(x) = a_o + a_1 x + \ldots + a_n x^n$

q is the forward shift operator.

Combining equation B.1 and B.3 gives the closed loop equation B.4

$$A^*(q^{-1})\ y(t) + \frac{G^*(q^{-1})\quad B^*(q^{-1})\quad y(t)}{F^*(q^{-1})\ B^{+*}(q^{-1})} = \varepsilon(t) \tag{B.4}$$

Using the factorization of B gives

$$B^{+*}(q^{-1})(A^*(q^{-1})\ F^*(q^{-1}) + G^*(q^{-1})\ B-*(q^{-1}))\ y(t)$$

$$= F^*(q^{-1})\ B^{+*}(q^{-1}) \tag{B.5}$$

Therefore the characteristic polynomial of the closed loop system can be found from equations B.5 and B.2 to be

$$T(q^{-1}) = B^{+*}(q^{-1})\ B^{-}(q^{-1}) \qquad \text{B.6}$$

Therefore if the zeros of the system are $B_1 \ldots B_{nb-1}$ and just the first j of these are inside the unit circle the roots of $T(q^{-1})$ in the plane are

$$B_1, \ldots, B_j, \quad \frac{1}{B_{j+1}}, \ldots, \frac{1}{B_{nb-1}}$$

APPENDIX C

Summary of regions of stability and conditional stability

Many systems will fall into several of the following categories in which case the least stable result applies.

Open Loop system	Control		
	$T'(z^{-1})=B'(z^{-1})$ Min. variance	$T'(z)$ product of zeros less than 1	$T'(z)=1$ simple control
Unstable: (1+A) has roots outside unit circle	Conditionally stable	Conditionally stable	Conditionally stable
Non minimum phase $B(z^{-1})$ with roots outside the unit circle as described in Chapter 2	Unstable	Stable	Stable
3 or more poles near the unit circle	Conditionally stable	Conditionally stable	Conditionally stable
At least one pole near to a zero	Stable	Stable	Conditionally stable
2 poles and a time delay of more than one sample period	Conditionally stable	-	-

Please Return

APPENDIX D

Choice of Closed Loop Pole Positions

There are three factors which compete in the choice of closed loop pole positions. The first is the stability and conditional stability of the closed loop system, the second is the variances of the output and return difference for the closed loop system, the third is the complexity of the algorithm for choosing the closed loop pole position. The constraint that the closed loop system is stable must be satisfied, hence the closed loop poles must be within the unit circle. This implies that a very simple strategy such as choosing the positions to be the same as the open loop zeros cannot be used for all systems. This implies that for minimum variance control a more complicated control algorithm involving polynomial factorisation will be required. (e.g. for non minimum phase system). Similarly the requirement that the system should not be conditionally stable conflicts with low output variance, (section 3.6).

Several different strategies could be tried.

1. For minimum variance systems. The closed loop poles can be chosen to be the open loop zeros together with an extra pole somewhere around 0.9 to decrease the input variance by relaxing the control, i.e. first decide how fast a response can be reasonably expected and chose the closed loop poles to give such a response. This use of the open loop zeros has the advantage of removing most of the computation of the control law.

2. For systems which either are open loop stable or have all the open loop zeros within the unit circle some combination of the open loop pole polynomial and the open loop zero polynomial could be used in a similar manner to that used by Dr. Clarke (Chapter 1, reference 5).

3. For systems which are open loop unstable and may have zeros outside the unit circle some fixed polynomial could be tried with one or more poles near to the +1 point.