WARWICK
THE UNIVERSITY OF WARWICK

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:
http://wrap.warwick.ac.uk/131784

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

SHORT TERM SPECTRAL ESTIMATION
WITH APPLICATIONS

A thesis presented for the degree of
Doctor of Philosophy in the Department
of Engineering Science, University
of Warwick, by,

L. BALMER, B.Sc., M.Sc., C.Eng., M.I.E.E.

April 1986

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I am indebted to my supervisor, Professor J. L. Douce, not only for his technical guidance, but for his constant encouragement throughout the period of the research leading to this thesis.

I would like to thank Alan Hume for his assistance with the Prime Computer installation and for his help with many other computing aspects of the project.

My thanks are due to my wife, Eileen, for her unfailing patience and to Eleanor for her courage in undertaking the typing of this thesis.

A part of the computer time used was funded by a Science and Engineering Research Council grant.

SUMMARY

This thesis investigates the effect of finite record length on spectral estimation with particular reference to the estimation of frequency response functions and to the application of frequency response methods to the problem of adaptive control.

Conventional analysis of the statistics of frequency response estimation has concentrated on the statistics of the auto and cross spectral estimates involved. The approach adopted here is to attribute the errors in the estimates to transient terms caused by the finite record lengths of input and output data. It is shown that these transient terms can be separated into a term correlated with the input and an uncorrelated term. The first term causes bias and the second can be regarded as an uncorrelated external disturbance.

Expressions for the bias of the estimate are obtained both in open-loop and closed-loop configurations. An expression for the variance of the estimator is obtained for the open-loop case only, but confidence intervals on the estimates are derived by both configurations.

A new parametric method of identification is investigated, the novelty being in the introduction of additional parameters to account for the transient terms. It is shown that the bias due to the finite record length can be removed completely in the noise free case. Methods similar to those used for parametric identification in the time domain are used to reduce the bias when noise is present.

Results obtained for non-parametric and parametric identification are applied to the area of adaptive control. Methods of applying frequency domain techniques to the problem of adaptive control are investigated. It is shown that in certain areas such an approach can have advantages over the corresponding time domain methods.

The results obtained, both in the area of frequency response estimation and in the field of adaptive control, are verified by simulation studies.

List of Symbols

As far as possible the following conventions have been followed.

Lower case variables denote time signals, $x(t)$, $y(t)$, and upper case, $X(j\omega)$, $Y(j\omega)$, denote their short term Fourier Transforms.

$\hat{}$ over a variable denotes an estimated value, e.g. $\hat{G}$

$-$ over a variable denotes mean value, e.g. $\bar{G}$

$*$ attached to a variable denotes complex conjugate, e.g. X^*, between variables it denotes convolution, e.g. $x(t)*y(t)$.

$a_0, a_1 \ldots a_n$	Coefficients in denominator of frequency response function.
$A(j\omega)$	Denominator of frequency response function.
$b_0, b_1 \ldots b_m$	Coefficients in numerator of frequency response function.
$B(j\omega)$	Numerator of frequency response function.
$D(j\omega)$	Controller frequency response function.
$c(t)$	Output signal of closed-loop system.
$e(t)$	Error signal of closed-loop system.
$E\{\ \}$	Expectation operator.
$g(t)$	System impulse response.
$G(j\omega)$	System frequency response function.
$h(t)$	Transient term due to finite record length.
K	Proportional gain in three term controller.
K_I	Integral gain in three term controller.
L	Number of data blocks in estimate.
$N(\ )$	Normal distribution.
$n(t)$	Noise signal.
$p(x)$	Probability density function.
$r(t)$	Input signal to closed-loop system.

$R_{xx}(\tau), R_{xy}(\tau)$	Auto and cross-correlation functions.
s	Laplace operator
t	Time
T	Record length
T_D	Derivative time constant in three term controller
$u(t)$	Control signal
$x(t)$	System input signal
$y(t)$	System output signal
$\underline{\alpha}$	Parameter vector
$\delta(t)$	Impulse function
λ	Forgetting factor
σ^2	Variance
τ	Time delay, argument in correlation function
Φ_{xx} Φ_{xy}	Auto and cross-spectral densities
ω	Angular frequency

CHAPTER 1

Introduction

The motivation for this project springs from an earlier research project undertaken by the author. This was in the area of adaptive control applied to a problem in metrology [Balmer (1), Balmer et al (2), Balmer and Douce (3)]. The performance index in the adaptive system concerned was based on time domain criteria; the analysis of the system was undertaken in the frequency domain. This led to considerations as to whether an adaptive control system based on a frequency response model and using frequency domain performance criteria would in general offer any advantages over a comparable system based in the time domain. In adaptive systems identification from short data lengths is important so as to cater for time variation of system or input parameters. Examination of the literature revealed that, although the problem of short record lengths had received considerable attention from the viewpoint of estimation of the power spectrum, corresponding results concerning the estimation of frequency response functions were less plentiful.

Hence, the aim of this project became threefold:-

(i) To review the methods used to estimate auto and cross-spectra from short record lengths.

(ii) To investigate the use of such methods for the estimation of frequency response functions.

(iii) To apply frequency domain techniques to the area of adaptive control.

The thesis is concerned with discrete time signals and data, although most of the discussion is equally applicable to continuous signals.

This introductory chapter gives a brief review of the methods that are available for power spectral estimation, giving emphasis to the effects of short record lengths. This leads into a review of the existing literature on frequency response estimation, again with the emphasis on the effect of short time records. The chapter concludes with an outline of the remainder of the thesis.

1.1 Power Spectra Estimation

An interesting historical perspective on spectral estimation is given in a paper by Robinson (4). This paper gives the birth of spectral estimation in ancient times with the empirical determination of the length of the day, the phases of the moon and the length of the year. It then traces development through the work of Newton, Fourier and into the twentieth century with the spectral theory of the atom. The three most common methods of empirical spectral analysis that have found favour today are based on methods that were first introduced in the first third of this century.

At the turn of the century Sir Arthur Schuster introduced the idea of the periodogram in the analysis of sunspot data. Given N observations of a time series x(n) then the periodogram $P(\omega)$ was defined as

$$P(\omega) = \frac{1}{N} \left| x(1)e^{j\omega} + x(2)e^{j2\omega} + \ldots\ldots x(N)e^{jN\omega} \right|^2$$

Peaks in the periodogram showed the locations of the frequencies of the underlying sinusoidal motion. As can be imagined, the labour involved in calculating the periodogram for large amounts of data was enormous.

In 1927 G. U. Yule introduced the concept of a finite parameter model to describe a random process. The process x(n) was assumed to have been generated by applying white noise to a system described by a difference equation. The coefficients in this equation were obtained by regression analysis, regressing the value x(n) against its past values. The equations describing the least squares solution to the regression problem are known as the Yule-Walker equations.

The third method was developed by Norbert Wiener and was given the title of Generalised Harmonic Analysis. Wiener showed that the power spectrum and the auto-correlation function of a time series formed a Fourier Transform pair, hence enabling the spectrum to be calculated via its auto-correlation function.

Although theoretically all these methods give identical results, in practice this is only true if the records being analysed are long. Results obtained from short time records are very erratic and often are of little practical use.

A turning point in the empirical analysis of time series came with the work of J. W. Tukey over the years 1949-1955. This work is collected in the book written with Blackman in 1958 [Blackman and Tukey (5)]. One of the major contributions presented was the idea of modifying or windowing the measured auto-correlation function before taking the Fourier Transform to obtain the power spectral density. The effect of finite data length can be included in the method by considering this as an operation on the true auto-correlation function by a triangular window.

Details of the steps in the Blackman-Tukey method are as follows:

(i) Derive an estimate $\hat{R}(\tau)$ of the auto-correlation function. There are two estimators in common use,

$$\hat{R}(\tau) = \frac{1}{T}\int_0^{T-\tau} x(t)x(t+\tau)dt$$

and

$$\hat{R}(\tau) = \frac{1}{T-\tau}\int_0^{T-\tau} x(t)x(t+\tau)dt$$

The second of these estimators is unbiased and, at first sight, would be the better to use. However, the first (biased) estimator has the lower variance and in general will produce the lower mean square error [Jenkins and Watts (6)].

(ii) Multiply the estimate obtained in (i) by a suitable weighting or window function $w(\tau)$.

(iii) Take the Fourier Transform of (ii) to obtain an estimate of the power spectral density

$$\Phi(\omega) = F\{\hat{R}(\tau)w(\tau)\}$$

Blackman and Tukey obtained expressions for the bias and variance of the estimate and introduced the useful concept of viewing the windowing process in the frequency domain. The estimated mean power can be written as a convolution between the true power spectrum and the Fourier Transform of the window. The result, at a single frequency f_0, for a rectangular window, is shown in Fig. 1.1.

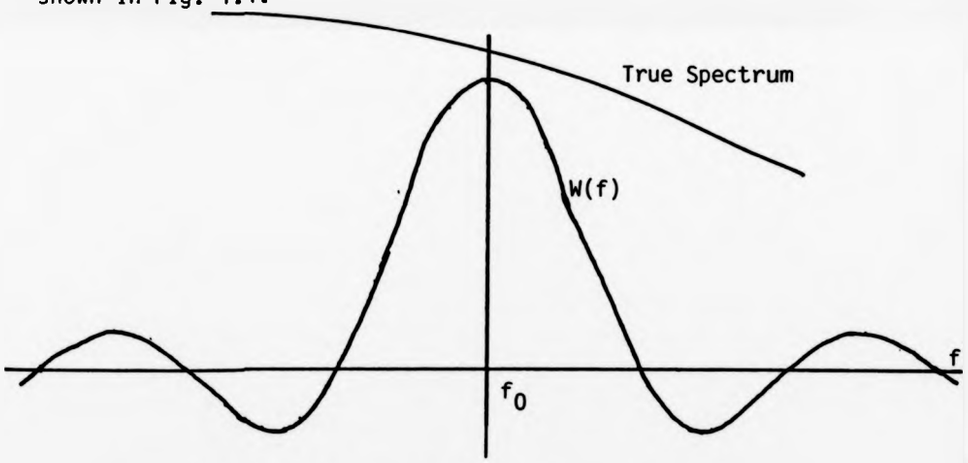


Fig. 1.1 Effect of Rectangular Time Window.

The estimated mean power at f_0 is the area under the curve formed by the product of the true spectrum and W(f). This method of viewing the process gives rise to the following useful interpretations.

(i) The true power at f_0 can only be obtained if the function W(f) is an impulse function $\delta(f - f_0)$. This would only be the case if the lag window was of infinite width.

(ii) The estimated mean power at f_0 is biased and this bias can be considered as being caused by two factors:

(a) Frequency components adjacent to f_0 "smear" the frequency f_0 leading to a lack of resolution.

(b) Frequency components in the side-lobes of W(f) can have a major influence at f_0 (even giving negative results).

It should be emphasised that these results apply to mean power for an ensemble average (time average for a stationary process).

As well as mean power, it is also important to obtain the factors determining the variance of the estimate. Although Blackman and Tukey do obtain expressions for the variance of the estimate at each frequency (and the covariance at differing frequencies), the question is treated in a broader manner by other authors [Jenkins and Watts (6)]. Jenkins and Watts show that for a rectangular (do-nothing) window the estimate of power is not consistent; its variance does not decrease to zero as the record length increases to infinity. Bartlett (7) proposed a solution that involved splitting the N point record into L separate records, each of length N/L points. An estimate of the spectral density $\Phi_i(\omega_k)$ is made for each block and these estimates combined to give an estimate $\Phi(\omega_k)$ according to

$$\bar{\Phi}(\omega_k) = \frac{1}{L} \sum_{i=1}^{L} \Phi_i(\omega_k)$$

This procedure reduces the variance of the estimate (by a factor of L if the signal is white noise). The variance obtained can be shown to be equivalent to applying a specific weighting (the Bartlett Window) to the complete record. However, this window has a greater bandwidth than the rectangular window and, hence, in general introduces greater bias.

Hence, the choice of window influences the resolution, bias and variance of resulting estimates. This gave rise to a large number of windows in an attempt to "trade off" the features of these properties. Harris (8) lists 24 windows and does an exhaustive comparison in terms of these and other factors.

This procedure for obtaining power spectral density estimates was known as the Blackman-Tukey (BT) or mean-lagged products method and was the most widely used method until the advent of the Fast Fourier Transform.

The publication of the paper by Cooley and Tukey (9) in 1965 on the Fast Fourier Transform (FFT) algorithm caused a major shift away from the Blackman-Tukey method. This algorithm provided a method of obtaining the discrete Fourier Transform that gave a very significant reduction in computing time compared with existing methods. Hence, the main obstacle to the use of the periodogram, as envisaged at the turn of the century was now removed. It also raised the possibility of performing spectral analysis in "real time", Wellstead (10).

The problems of frequency resolution, bias and variability still exist with the FFT method. The balance between resolution and bias was obtained by applying suitable weighting to the time data before applying the FFT - this was termed linear windowing (as opposed to quadratic windowing when the weighting was applied to the correlation function). The variability was reduced by either averaging in frequency over neighbouring elements of the periodogram [Bingham et al (11)] or by averaging over frequency components obtained from independent periodograms. This latter method was known as the weighted segment averaging or WSA method. Welch (12) proposed a method of averaging that involved overlapping the segments. This method was claimed to give results comparable

to the WSA method but required fewer computations. Welch's method, known as the weighted-overlap-segment-averaging (WOSA) method, became the standard method of spectral estimation.

Although the WOSA method became standard, the effect of linear windowing on the estimates was not as evident as the effect of quadratic windowing on the BT estimates. This has lead to alternative methods of viewing the application of the time window. For simplicity, the data is assumed to be sinusoidal but not necessarily of the same period as the observation time. Bertram (13) considers the amplitudes of the harmonics produced by such an analysis and considers the choice of window in terms of producing rapid convergance of this series.

Burgess (14) attributes the effects of leakage to discontinuities that exist at the end of the observation period due to the implied periodicity of the Discrete Fourier Transform. The effect of the time window is to remove the discontinuity and the discontinuities in as many higher derivatives as possible. However, his argument becomes very suspect when he states that the rectangular window must also be treated as periodic, thus removing the side-lobes due to this type of window. He confuses the issue further by requiring the input to be white but also sufficiently band limited to avoid aliasing problems.

The whole concept of linear windowing has been criticised by Yuen (15) from two viewpoints:

(i) The process of giving less weight to data points that have the same variance as any other point is statistically unsound. (This is not the case with quadratic windowing where lags in the estimated autocorrelation function that have most variance are given least weighting).

(ii) Although an equivalent quadratic window can easily be derived for a given linear window, the equivalence is only in bias, not variance. Any attempt to derive equivalent variance windows leads to equations that are too unwieldy to be informative.

Carter & Nutall (16) reply to these criticisms. They state that the low weighting of data at the extremes of a block is compensated for by overlapping the blocks and including the data more than once. Nutall (17) shows that the variance of the WOSA method is comparable to that obtained by the BT method for identical record lengths. Both Yuen, and Carter & Nutall (16) provide simulation results that support their respective arguments and the subject appears to be still one for discussion.

Nutall (18) proposes a method of spectral estimation that combines linear and lag weighting. This method involves obtaining a first stage spectral estimate using the WSA or WOSA method. However, the process does not stop there but involves taking the inverse transform of the power spectral estimate to give an auto-correlation function estimate. This is then subject to quadratic weighting as in the BT method and the weighted correlation function estimate is transformed to give the final power spectral density estimate. By choosing the overlaps, the form of the linear windowing and the form of the quadratic windowing, this method includes the BT, the WSA and the WOSA methods. The advantage of this method is that the linear window can be rectangular and no overlap employed, thus much reducing the computational effort required. The requirements of bias and variability can then be determined by a suitable quadratic window, this leading to some rather "unusual" quadratic windows. This is of particular advantage if "window carpentry" is used, i.e. initial estimates with a particular window shape are used to guide the choice of window for a more refined estimate.

The BT and WOSA methods are non-parametric methods - i.e. no form of model is assumed for the process generating the signal under investigation. As already stated, an alternative approach is to assume that the process investigated has been produced by applying white noise to a system described by a difference equation. The equations to be solved to obtain the coefficients in the difference equations are the Yule-Walker equations. Although these equations were formulated in the first third of the century, they received little attention until the sixties. This was probably due to a lack of an efficient method of solution using the computer resources available. In 1960 Durbin (19) proposed a method of solution that used a recursive algorithm first developed by Levinson in connection with the prediction problem and this algorithm became known as the Levinson-Durbin algorithm. This lead to a growth of interest in parametric methods of spectrum estimation, [Cadzow (20)]. In general these methods give better resolution than non-parametric methods, but are not as computationally efficient.

Many other methods of power spectrum estimation have been proposed [Kay and Marple (21)]. The one that has had the greatest influence on modern methods of estimation has been the maximum entropy method proposed by Burg (22) in 1967. In the methods of analysis described so far the effect of windowing has been to alter the observable data (or its estimated correlation function) by the weighting of the window and to assume all unobservable data is zero. Burg proposed leaving the observable data unaltered but to give values to data not directly observed. The basis of the values given is such that the spectral estimate obtained must be the most random (have maximum entropy) of any power spectra that is consistent with the measured data. A parametric model is assumed for the power

spectrum and this model, together with the measured data, forms the constraints on the maximisation. Using a measure of entropy based on information theory, this measure is maximised under these constraints and the solution of the resulting equations give the coefficients in the model. This method gives very good frequency resolution, but noise degrades the performance, causing bias. Hence, when noise is present other methods may give superior results.

All the estimation methods described have been applied to power spectrum estimation. More important, from the viewpoint of this thesis, is the estimation of cross-spectra and of frequency response functions.

1.2 Cross-Spectra and Frequency Response Estimation

Cross-spectra can be estimated by both the BT and the WOSA methods. In the BT method the estimated cross-spectra is the Fourier Transform of the estimated cross-correlation function. In the WOSA method it is obtained as $X(j\omega)Y^*(j\omega)$ where $X(j\omega)$ and $Y(j\omega)$ are the short term transforms of the data $x(t)$ and $y(t)$. Quadratic or linear windowing can be applied as in the estimation of power spectra.

One important difference in the estimation of cross-spectra is that the estimated cross-correlation function does not, in general, have its maximum at zero lag. Hence, the application of a window can truncate values having considerable influence. This effect can be reduced by "re-aligning" the records in time such that the cross-correlation function has its maximum at zero lag. The effect of the re-alignment is taken into account by appropriate phase shift of the cross-spectral density.

An estimate of the frequency response function can be obtained by regression analysis in the frequency domain [Jenkins and Watts (6)]. This gives the best estimate, in the least squares sense, of the frequency response $G(j\omega)$ as

$$\hat{G}(j\omega) = \frac{\hat{\Phi}_{xy}(j\omega)}{\hat{\Phi}_{xx}(j\omega)}$$

where $\hat{\Phi}_{xy}(j\omega)$ is an estimate of the cross-spectrum between input and output and $\hat{\Phi}_{xx}(j\omega)$ is an estimate of the input spectrum. The statistics of the estimate $\hat{G}(j\omega)$ are not easily expressed in terms of the statistics of $\hat{\Phi}_{xx}(j\omega)$ and $\hat{\Phi}_{xy}(j\omega)$. Goodman (23) and Akaike and Yamanouci (24) have studied the distribution of $\hat{G}(j\omega)$ in terms of bivariate complex normal distributions. Wellstead (10) applied these results to the estimation of open-loop and closed-loop frequency response functions in the presence of noise. Davall (25) derived alternative distributions for $\hat{G}(j\omega)$ in terms of the distribution of its magnitude and of its phase. These distributions are framed in terms of noise/signal power and assume the effects of finite record length are negligible. Douce (26) used these distributions to obtain the bias of closed-loop estimates obtained in the presence of noise.

An alternative estimator for frequency response functions was proposed by Allen and Rabiner in a series of papers (27, 28, 29). They showed that by summing sufficient overlapped windowed segments (the number depending on the window used) it is possible to remove the effect of the window. The resulting expansion at any data point x(n) is termed the overlap add expansion. The identification problem is first framed as a regression problem in the time domain. The resulting correlation functions are then written in terms of the overlap expansions. By taking Fourier Transforms, an estimate of

the frequency response can be obtained in terms of the auto and cross-spectra which differ from the classical estimates of these quantities. Taking the auto-spectral estimate, this is given by

$$\sum_k \sum_p X_k(j\omega)\ X^*_p(j\omega)$$

where $X_k(j\omega)$ and $X_p(j\omega)$ are given by the short term Fourier Transforms of the overlapping segments p and k. This differs from the classical case where p = k and it is the inclusion of these additional terms that is responsible for the removal of bias. This is obtained at the expense of additional computing requirements. The algorithm is known as the Short Term Unbiased Spectral Estimation or STUSE algorithm.

1.3 Outline of Thesis Content

Chapter 2 presents a new approach to the estimation of frequency response functions. It is shown that, in terms of short term Fourier Transforms, the transform of the output is equal to the true frequency response function times the transform of the input plus a transient term to account for the finite observation time. This transient term introduces bias and variance into the estimate. A model is developed in which the output is considered as the sum of the input multiplied by a biased frequency response function plus an independent equivalent noise signal. From this model the variance and distribution of the estimate are obtained. Results are presented for a simulated system showing the effect of different windows and different input spectra.

Chapter 2 considers open-loop systems and chapter 3 extends the ideas introduced to analyse closed-loop systems. The

transients now appear at input and output due to the feedback path. This introduces a further correlated component between input and output which increases the bias over that obtained in the open-loop system. A method of evaluating the closed-loop bias and of establishing confidence limits is considered and again results presented for simulated systems.

Chapter 4 considers the effects of transients when parametric identification is used. A model is assumed for the frequency response function and the parameters in the model are obtained by regression analysis in the frequency domain. The novelty in the approach is that additional parameters are included to account for the transient terms. It is shown that the bias due to the finite record length can be removed completely in the noise free case. Again, results are given for a simulated system.

Chapter 5 considers the application of frequency response methods to the problem of adaptive control. After a review of adaptive control systems using frequency response methods, several new adaptive schemes are proposed and the advantages and problems associated with each are discussed. Simulation results on these systems are presented in chapter 6.

Chapter 7 presents concluding comments and suggests some topics for further investigation.

References - Chapter 1

1. Balmer, L., "Control Aspects of an Optical Gauging Instrument", MSc Thesis, University of Warwick, 1977.

2. Balmer, L., Falkner, A.H. and Taberner, W.N., "Novel Electro-Optical Techniques in Metrology", Electronics and Power, pp.27-30, Jan. 1976.

3. Balmer, L. and Douce, J.L., "An Application of Hill Climbing Techniques in Measurement", IEEE Trans. on Automatic Control, Vol. AC-27, pp.89-94, Feb. 1982.

4. Robinson, E.A., "A Historical Perspective of Spectrum Estimation", Proc. IEEE, Vol.70, No.9, pp.885-907, Sept. 1982.

5. Blackman, R.B. and Tukey, J.W., "The Measurement of Power Spectra from the Point of View of Communications Engineering", Dover Books, 1959.

6. Jenkins, G.M. and Watts, D.G., "Spectral Analysis and its Applications", Holden Day, 1968.

7. Bartlett, M.S., "An Introduction to Stochastic Processes with Special Reference to Methods and Applications", Cambridge University Press, 1953.

8. Harris, F.J., "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform", Proc. IEEE, Vol.66, pp.51-83, Jan. 1978.

9. Cooley, J.W. and Tukey, J.W., "An Algorithm for Machine Calculation of Complex Fourier Series", Math. Comput., Vol.19, pp.297-301, April 1965.

10. Wellstead, P.E., "Aspects of Real-Time Digital Spectral Analysis", PhD Thesis, University of Warwick, 1970.

11. Bingham, C., Godfrey, M.D. and Tukey, J.W., "Modern Techniques of Power Spectrum Estimation", IEEE Trans. Audio Electoacoust., Vol. AU-15, pp.56-66, June 1967.

12. Welch, P.D., "The use of Fast Fourier Transform for the Estimation of Power Spectra: A Method based on Time Averaging over Short Modified Periodograms", IEEE Trans. Audio Electroacoust., Vol. AU-15, pp.70-73, June 1967.

13. Bertram, S., "Frequency Analysis Using the Discrete Fourier Transform", IEEE Trans. Audio Electroacoust., Vol. AU-18, pp.495-500, Dec. 1970.

14. Burgess, J.C., "On Digital Spectrum Analysis of Periodic Signals", J.Acoust. Soc. Am., Vol.58, No.3, pp.556-567, Sept. 1975.

15. Yuen, C.K., "Comments on Modern Methods for Spectral Estimation", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-27, pp.298-299, June 1979.

16. Carter, G.C. and Nutall, A.H., "On the Weighted Overlapped Segment Averaging Method for Power Spectral Estimation", Proc. IEEE, Vol.68, pp.1352-1354, Oct. 1980.

17. Nutall, A.H., "Spectral Estimation by means of Overlapped FFT Processing of Windowed Data", Naval Underwater Systems Center, New London, CT, Rep.4169, and suppl. TR41695.

18. Nutall, A.H. and Carter, C.G., "Spectral Estimation using Combined Time and Lag Weighting", Proc. IEEE, Vol.70, No.9, pp.1115-1125, Sept. 1982.

19. Durbin, J., "The Fitting of Time Series Models", Rev. Inst. Int. de Stat., Vol.28, pp.233-244, 1960.

20. Cadzow, J.A., "Spectral Estimation: An Overdetermined Rational Model Equation Approach", Proc. IEEE, Vol.70, No.9, Sept. 1982.

21. Kay, S.M. and Marple, S.L., "Spectrum Analysis - A Modern Perspective", Proc. IEEE, Vol.69, No.11, pp.1380-1419, Nov.1981.

22. Burg, J.P., "Maximum Entropy Spectral Analysis", Proc. 37th Meeting Society of Exploration Geophysicists (Oklahoma City, OK), 31st Oct. 1967.

23. Goodman, N.R., "On the joint estimation of the spectra, cospectrum and quadrature spectrum of a two dimensional stationary Gaussian process", PhD Thesis, Princeton University, 1957.

24. Akaike, H. and Yamanouchi, Y., "On the Statistical Estimation of Frequency Response Functions", Ann. Inst. Stat. Maths., Vol.14, Part I, pp.23-56, 1962.

25. Davall, P.W., "Applications of statistics in the spectral analysis of time-varying systems", PhD Thesis, University of Warwick, 1975.

26. Douce, J.L., "Bias of Frequency-Response Estimates in Closed-Loop Systems", Proc. IEE, Vol.127, Pt.D, No.4, pp.149-152, July 1980.

27. Rabiner, L.R. and Allen, J.B., "Short-time Fourier Analysis Techniques for FIR System Identification and Power Spectrum Estimation", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-27, No.2, pp.182-192, April 1979.

28. Allen, J.B. and Rabiner, L.R., "Unbiased Spectral Estimation and System Identification using Short-Time Spectral Analysis Methods", Bell System Technical Journal, Vol.58, No.8, pp.1743-1963, Oct. 1979.

29. Allen, J.B. and Rabiner, L.R., "A Unified Approach to Short-Time Fourier Analysis and Synthesis", Proc. IEEE, Vol.65, No.11, pp.1558-1569, Nov. 1977.

CHAPTER 2

Frequency Response Estimates based on Short Data Blocks

1 Open-Loop Estimates

2.1 Introduction

This chapter investigates the errors caused by finite record length on the estimation of the frequency response of a system. Only open-loop response will be considered in this chapter and the following chapter will extend the analysis to closed-loop systems. The effect of the choice of window will be considered as will that of combining estimates from more than one data block.

As stated in chapter 1, the conventional approach to the analysis of frequency response estimates has been via the auto and cross-spectral density functions. The approach adopted here is to attribute the errors in the frequency response estimates to transient terms that occur at the start and end of each data block. It is shown that certain terms in these transients correlate with the input data and this correlation causes bias in the estimate. The remaining terms in the transients are uncorrelated with the input and give a distribution of the estimate around its biased value. It is shown that these uncorrelated terms have the same effect as an uncorrelated noise source and results available for that situation can now be applied to obtain the distribution.

General expressions are developed for the bias of the frequency response estimate and it is shown that the bias is independent of the number of blocks used in the experiment.

A model is developed in which the output is considered as the sum of the input multiplied by a biased gain plus an independent noise signal. From this model the variance and distribution of the gain estimate are derived.

The terms in the transients that correlate with the input depend upon the input spectrum. Some of the results obtained apply for a white noise input only, others are true generally.

Simulation results will be presented to support the theoretical results derived.

Throughout the chapter extensive use will be made of the measured Short Term Fourier Transform. Given a sample of a process $x(t)$, over time range $0 < t < T$, this is defined as

$$X(j\omega) = \frac{1}{T}\int_0^T x(t)\, e^{-j\omega t} dt$$

where $\omega = 2\pi n/T$, n being an integer.

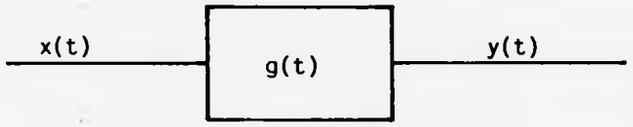
The measured autospectrum $\hat{\Phi}_{xx}(j\omega)$ is defined as

$$\hat{\Phi}_{xx}(j\omega) = X(j\omega)\, X^*(j\omega)$$

As defined, the units of $X(j\omega)$ are those of amplitude and those of $\hat{\Phi}_{xx}(j\omega)$ are those of amplitude squared or power. The usual units of $\hat{\Phi}_{xx}(j\omega)$ are those of a spectral density, i.e. power/unit frequency. Hence, $\hat{\Phi}_{xx}(j\omega)$ should be normalised by multiplying by a factor T. This is not carried out here as its inclusion leads to more cumbersome expressions.

2.2 Transients due to short term estimates

2.2.1 Nature of transients



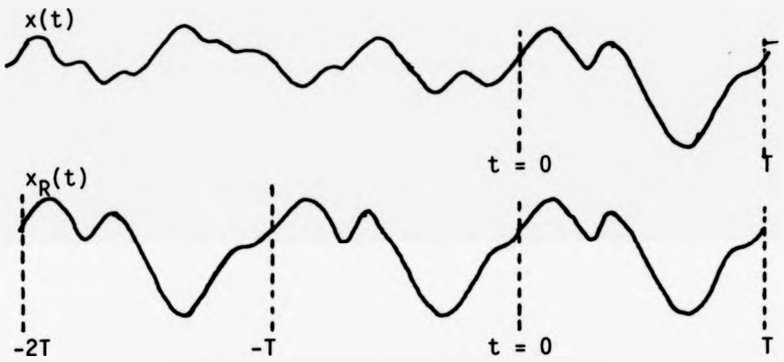


Fig. 2.1 Signals x(t) and $x_R(t)$

Consider the system shown in Fig. 2.1. Here x(t) is the statistically stationary input to the system measured over observation time t = 0 to t = T. $x_R(t)$ is a periodic signal equal to x(t) over that observation time with

$$x_R(nT + t) = x(t) \qquad n \text{ an integer}$$

Since $x_R(t)$ is periodic, the response $y_R(t)$ to this input is also periodic and the short term Fourier transforms of these signals at harmonic frequencies ω ($\omega = n\,2\pi/T$) are related by

$$Y_R(j\omega) = X_R(j\omega).\ G(j\omega) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ 2.1$$

Over the period of observation the input signal x(t) is equal to $x_R(t)$; the output y(t) is not however equal to $y_R(t)$.

This is due to differing responses to the initial conditions at $t = 0$ caused by the signals $x(t)$ and $x_R(t)$ prior to $t = 0$. Hence, the difference between the response $y(t)$ and $y_R(t)$ can be written

$$y(t) - y_R(t) = h(t) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 2.2$$

where $h(t)$ is a term that decays to zero as $t \to \infty$. It is useful to consider $h(t)$ as formed of two components.

One component, $h_0(t)$, will be defined as the output signal at $t = 0$ due to the input signal $x(t)$ prior to $t = 0$.

The other component will be the output signal at $t = 0$ due to the input signal $x_R(t)$ prior to $t = 0$. However, because $x_R(t)$ is periodic, this is also the output signal produced at $t = T$ due to the input signal $x_R(t)$ prior to $t = T$. Because $x(t) = x_R(t)$ over the interval $0 < t < T$, provided the impulse response of the system $g(t)$ tends to zero as $t \to T$, then this component is also the output at $t = T$ due to the input $x(t)$ prior to $t = T$. This component will be defined as $h_T(t)$.

It may be argued that the component $h_T(t)$ cannot be of importance as it occurs outside the range of measurement $0 < t < T$. However, as will be shown, by viewing $h(t)$ in this manner it provides physical insight into the way $h(t)$ produces bias and variance in the frequency response estimate.

To summarise these definitions,

$$h(t) = h_0(t) - h_T(t)$$

where

$h_0(t)$ is the output at $t = 0$ due to the input signal $x(t)$ prior to $t = 0$.

$h_T(t)$ is the output at $t = T$ due to tne input signal $x(t)$ prior to $t = T$. (Note that the time origin for the signal $h_T(t)$ is $t = T$).

These components are shown in Fig. 2.2.

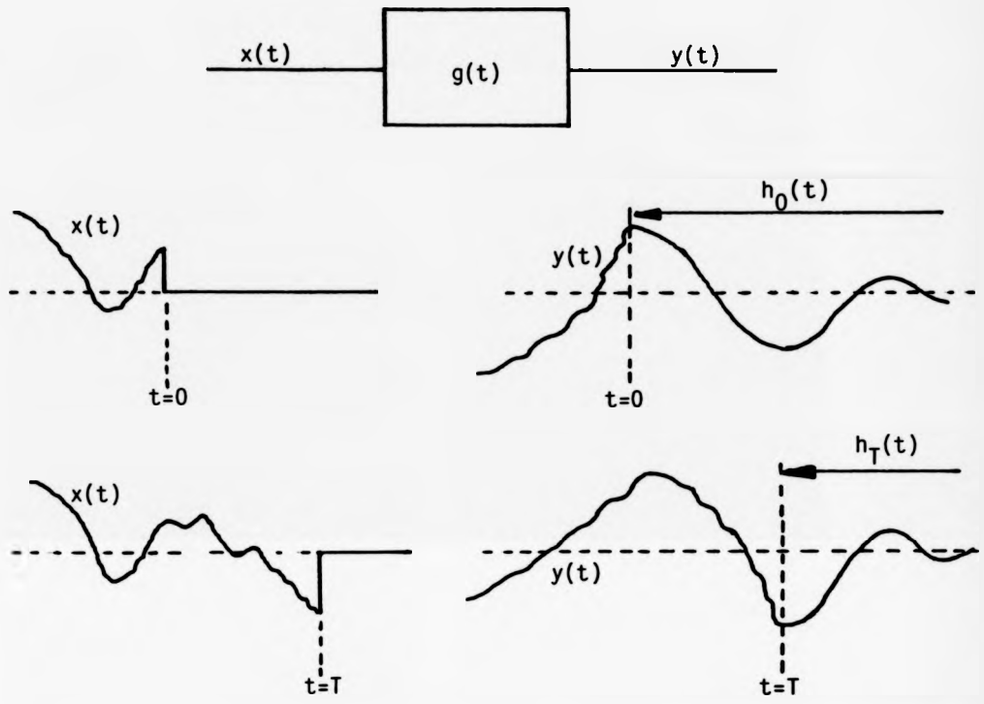


Fig.2.2 Components of transient term

The terms h(t), $h_0(t)$, $h_T(t)$, will be used extensively throughout this thesis. They, together with their short term Fourier Transforms, will be referred to as "transients".

Taking short term Fourier Transforms throughout eqn.(2.2)

$$Y(j\omega) - Y_R(j\omega) = H(j\omega)$$

and using eqn. (2.1) and the fact that $X(j\omega) = X_R(j\omega)$ gives

$$Y(j\omega) = G(j\omega)\, X(j\omega) + H(j\omega) \quad \text{.....(2.3)}$$

It is worth noting that this expression is exact even though the transforms are taken over finite time, i.e. the transient term has been truncated.

The time function h(t) corresponds to the output due to a particular set of initial values. Thus, the function can be expressed as a sum of natural modes of the system with amplitudes determined by past history. The sum of the short term Fourier Transforms of these modes determine the form of $H(j\omega)$. This form can be determined by use of the state variable representation of the system.

$$\dot{\underline{x}} = \underline{A}\,\underline{x} + \underline{B}u$$

$$y = \underline{C}\,\underline{x} + \underline{D}u$$

(Here u has been used as input signal and $\underline{x}$ as the state vector).

If the initial state of the system is $\underline{x}_0$ then the transient h(t) can be expressed

$$h(t) = \underline{C}\, e^{\underline{A}t}\, \underline{x}_0$$

The short term Fourier Transform $H(j\omega)$ is given by

$$H(j\omega) = \underline{C} \int_0^T e^{(\underline{A} - j\omega\underline{I})t}\, dt\, \underline{x}_0$$

$$= C(\underline{A} - j\omega\underline{I})^{-1}(e^{\underline{A}T}\, e^{-j\omega T} - \underline{I})\, \underline{x}_0$$

But $e^{j\omega T} = 1 \qquad (\omega = n2\pi/T)$

and $(e^{\underline{A}T} - \underline{I})$ is a constant matrix

$$(\underline{A} - j\omega\underline{I})^{-1} = \frac{\text{Adj}\ (\underline{A} - j\omega\underline{I})}{|\underline{A} - j\omega\underline{I}|}$$

If the system transfer function G(s) is written as B(s)/A(s) then

$$A(j\omega) = |\underline{A} - j\omega\underline{I}|$$

$$= a_0 + a_1 j\omega + \ldots\ldots\ldots\ldots a_n(j\omega)^n$$

and $\underline{C}(\underline{A} - j\omega\underline{I})(e^{\underline{A}T} - \underline{I})\, \underline{x}_0$

can be written as

$$H'(j\omega) = h_0 + h_1 j\omega + \ldots\ldots\ldots\ldots h_{n-1}(j\omega)^{n-1}$$

where $h_0 \ldots\ldots\ldots h_{n-1}$ are determined by the system constants and the system state at t = 0 and t = T. Hence

$$H(j\omega) = \frac{H'(j\omega)}{A(j\omega)} = \frac{h_0 + h_1 j\omega + \ldots\ldots\ldots\ldots h_{n-1}(j\omega)^{n-1}}{a_0 + a_1 j\omega + \ldots\ldots\ldots\ldots h_n(j\omega)^n} \quad \ldots\ldots(2.4)$$

Before the effects of the transients, on the estimation of the system frequency response function, are considered in detail a number of general points can be noted.

(i) The transients decay to negligible amplitude in a time determined by the system impulse response. As the block length T increases their transforms will become smaller in amplitude compared with those of the input and output signals. Hence, as T increases the effect of the transient terms on the frequency response estimate will decrease.

(ii) If the initial and final conditions are all equal then, $h_0(t) = h_T(t)$, and $H(j\omega) = 0$, no transient terms exist. An exact estimate of the frequency response function can be obtained. This is obviously the case when the input is periodic with period T.

This argument presents a neat alternative motivation for the introduction of time windows. From the viewpoint adopted here a window function is chosen so as to reduce the transient terms, especially where they are most likely to be at their maximum, at the start of the data block. Such a window also modifies the required terms and these effects must be balanced in the choice of window.

(iii) Using the concept of transients, the question can be asked as to whether symmetrical windows are necessarily the best for frequency response estimation. The transient term $h_0(t)$ is maximum

at the start of the output data block; reducing the amplitude of these points by a suitable window appears beneficial. However, there appears to be no reason to modify the input data in this region.

At the end of the output data block the effect of the transient $h_0(t)$ is small and now the output data requires no modification. However, it is the signal at the end of the input data block that gives the maximum contribution to the transient $h_T(t)$; again modification could be beneficial here. A possible window pair are shown in Fig. 2.3.

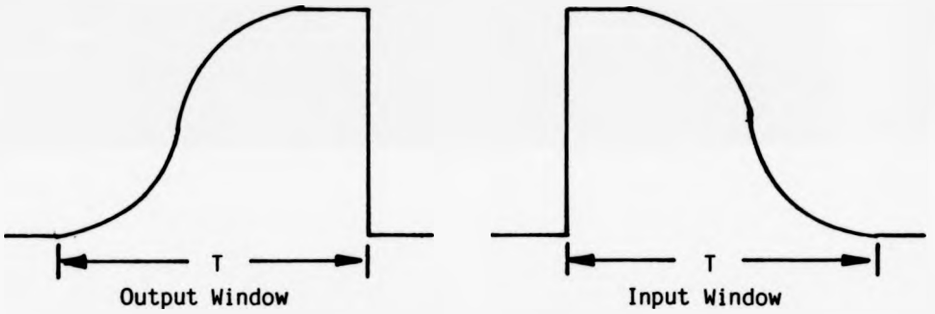


Fig. 2.3 Possible unsymmetrical window pair

(iv) If the input is restricted to white noise then the transient $h_0(t)$ is uncorrelated with the measured input over the block. This transient will give no bias in the frequency response estimate but will give variance to the estimate. This is no longer the case if the input is non-white, $h_0(t)$ now causes bias and variance.

(v) The transient $h_T(t)$ is correlated with the input over the block; this transient will cause bias and variance in the frequency response estimate.

(vi) If the transient term could be measured directly (by removal of the input after time T and measuring the subsequent response) an exact measure (in the absence of noise) of the frequency response could be obtained [Douce (1)].

2.2.2 Statistics of the transients

In order to investigate the effect of the transient terms on the frequency response estimates, the statistics of this transient are required. In particular, in order to determine the bias of the estimate, it is required to determine the component of the transient that correlates with the input signal. In order to determine the variance of the estimate, the component of the transient that is uncorrelated with the input is required.

The distribution of $H(j\omega)$ can be obtained by reference to eqn. (2.3).

$$Y(j\omega) = G(j\omega)\, X(j\omega) + H(j\omega) \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3)$$

Here $X(j\omega)$ and $Y(j\omega)$ are short term Fourier Transforms of samples taken from stationary random processes. The statistics of such variables have been investigated by Goodman (2) and he has shown that they have a complex normal distribution with zero mean. As eqn. (2.3) shows, $H(j\omega)$ is a linear combination of such variables - it must also be complex normal with zero mean.

It should be noted that if the input signal to a linear system is a stationary random process then the output is also a stationary random process. The transient however is not a stationary process and, in order to obtain its auto and cross-correlation functions a non-stationary analysis must be performed. This can be explained by reference to Fig. 2.4 [Papoulis (3)].

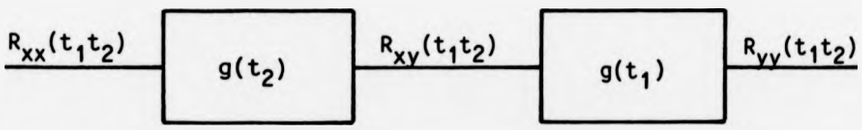


Fig. 2.4 Method of obtaining auto and cross-correlation functions for transient

$R_{xx}(t_1, t_2)$ is the autocorrelation function of the input signal. This is applied to the system treating t_1 as a parameter (the system operates on t_2) to give $R_{xy}(t_1, t_2)$.

$$R_{xy}(t_1, t_2) = R_{xx}(t_1, t_2) * g(t_2)$$

$R_{xy}(t_1, t_2)$ is now applied to the system treating t_2 as a parameter (the system now operates on t_1) to give $R_{yy}(t_1, t_2)$

$$R_{yy}(t_1, t_2) = R_{xy}(t_1, t_2) * g(t_1)$$

To calculate the auto and cross-correlation functions of the initial transient to a white noise input having a constant spectral density Φ_{xx} the procedure is as follows.

$$R_{xx}(t_1, t_2) = \Phi_{xx}\, \delta(t_1 - t_2) \qquad t_1 \leq 0, \quad t_2 \leq 0$$

$$R_{xx}(t_1, t_2) = 0 \qquad \text{otherwise}$$

Hence

$$R_{xy}(t_1, t_2) = \Phi_{xx}\, g(t_1 - t_2) \qquad t_1 \geq 0, \quad t_2 \geq 0$$

$$R_{xy}(t_1, t_2) = 0 \qquad \text{otherwise}$$

giving

$$R_{yy}(t_1, t_2) = \int_0^\infty g(t_1 - \tau)\, g(t_2 - \tau)\, d\tau$$

The expected value of the measured spectrum associated with these correlation functions is given by

$$E\{\Phi_{ab}(j\omega)\} = \int_0^T \int_0^T R_{ab}(t_1, t_2)\, e^{j\omega(t_1 - t_2)}\, dt_1 dt_2$$

2.2.3 Effect of Transients on the Frequency Response Estimate

Returning to eqn. (2.3) and now dropping the dependence of the terms on $j\omega$ gives

$$Y = GX + H$$

Multiplying through by X* and summing over L blocks gives

$$\sum_L X^*Y = G \sum_L X^*X + \sum_L X^*H \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.4)$$

The estimate used for frequency response is $\sum_L X^*Y / \sum_L X^*X$

$$\hat{G} = \frac{\sum_L X^*Y}{\sum_L X^*X} = G + \frac{\sum_L X^*H}{\sum_L X^*X}$$

Taking expected values gives the bias in this estimate as

$$E\left\{\frac{\sum_L X^*H}{\sum_L X^*X}\right\}.$$

In order to examine the properties of this estimate it is necessary to first examine a different estimate. Taking expected values in eqn. (2.4) and then dividing by $E\{\sum_L X^*X\}$ gives

$$\frac{E\{\sum_L X^*Y\}}{E\{\sum_L X^*X\}} = G + \frac{E\{\sum_L X^*H\}}{E\{\sum_L X^*X\}}$$

Interchanging the expectation and summation operators, the left hand side of this equation can be written

$$\frac{\sum_L E\{X^*Y\}}{\sum_L E\{X^*X\}}$$

However, X, Y for each block are samples from a stationary random process and each block is independent. Hence

$$\frac{\sum_L E\{X^*Y\}}{\sum_L E\{X^*X\}} = \frac{L\, E\{X^*Y\}}{L\, E\{X^*X\}} = G + k$$

where k is independent of the number of blocks. Hence,

$$\frac{E\{X^*H\}}{E\{X^*X\}} = k \quad \text{.....(2.5)}$$

This proves the interesting result that the bias of the estimate is independent of the number of blocks.

From Eqn. (2.5)

$$E\{X^*(H - kX)\} = 0 \quad \text{.....(2.6)}$$

Define a new random variable N such that

$$N = H - kX \quad \text{.....(2.7)}$$

As this variable is a linear combination of zero mean complex normal variables, it is also zero mean complex normal. Also from the last two equations,

$$E\{X^*N\} = 0$$

showing that X and N or orthogonal. As they are also normal with zero mean, this shows they are uncorrelated.

Re-writing eqn. (2.7) in terms of H and substituting into eqn. (2.3) gives

$$Y = GX + N + kX$$

or

$$Y = \overline{G}X + N \quad \text{.....(2.8)}$$

where $\overline{G} = G + k$

Hence the frequency response estimator introduced at the beginning of this section can be written

$$\hat{G} = \frac{\sum X^*Y}{\sum X^*X} = \overline{G} + \frac{\sum X^*N}{\sum X^*X}$$

Taking expected values gives

$$E\{\hat{G}\} = \overline{G} + E\{\frac{\sum X^*N}{\sum X^*X}\}$$

As shown, X and N are uncorrelated and the statistics of the random variable $\sum X^*N/\sum X^*X$ have been studied by Davall (4) where

in his case N represented uncorrelated noise. In particular he has shown that the expected value of this term is zero. Hence,

$$E\{\hat{G}\} = \bar{G}$$

The bias in the frequency response estimate can be evaluated either from $E\{\sum_L X^*Y\}/E\{\sum_L X^*X\}$ or from $E\{\frac{\sum_L X^*Y}{\sum_L X^*X}\}$

It is independent of the number of blocks and it is obviously easiest to obtain it from one block only.

Equation (2.8) can be represented as the block diagram of Fig. 2.5 where $\bar{G}$ is the biased value of G and N is zero mean, complex normal, uncorrelated with X.

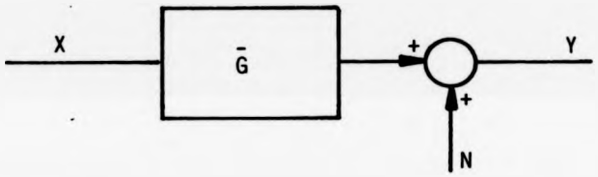


Fig. 2.5 Block diagram representing bias and equivalent noise

This result is very important; it shows that the effect of finite observation time can be accounted for by

(i) a biased value of G.

(ii) an equivalent uncorrelated noise source added at the output of the system.

As stated earlier, this case has been investigated and results have been obtained for the distribution of the estimate that can be applied with only slight modification to the situation here. In order to apply such results, it is necessary to calculate the biased estimate $\bar{G}$ and the variance of the equivalent noise source N.

2.3 Bias of the Estimate

As derived in the last section, the bias in the estimate can be obtained from two expressions,

either $$E\{\hat{G}\} = \frac{E\{X^*Y\}}{E\{X^*X\}}$$

or $$E\{\hat{G}\} = E\left\{\frac{X^*Y}{X^*X}\right\}$$

In this section two methods will be used to derive an expression for bias, the first depending on the first of the above expressions, the second upon the other expression.

The first method makes no assumptions about the system input but obtains the expected values of the measured auto and cross-spectral densities by relating them to the true auto and cross-correlation functions of the system.

The second method considers a sinusoidal input of random phase applied at t = 0. This approach is not as restrictive as may appear at first sight. Only the output terms that correlate with the input lead to bias and the input may be considered as one Fourier component in a more general input whose other components are uncorrelated with the one under consideration. By ignoring the input prior to t = 0 one is making the assumption that the general input is white noise.

The second method, although not as generally applicable as the first, does give a physical insight into the manner in which the bias arises due to correlation in the transient terms. This is even more apparent when applied to a specific system, as will be shown in section 2.5.

2.3.1 An expression for bias - Method 1

The estimate used will be

$$\bar{G} = \frac{E\{X^*Y\}}{E\{X^*X\}}$$

The expected values of the cross and auto spectra can be obtained by considering the Fourier Transforms of the cross and auto correlation functions after modification by a window to account for the finite block length [Jenkins & Watts (5)].

Restricting the input to white noise of constant spectral density Φ_{xx} (autocorrelation function $\Phi_{xx}\,\delta(\tau)$) then

$$E\{X^*X\} = \int_{-T}^{+T} \left(1 - \frac{|\tau|}{T}\right) \Phi_{xx}\,\delta(\tau) e^{-j\omega\tau}\,d\tau$$

$$= \Phi_{xx}$$

The cross-correlation function $R_{xy}(\tau)$ is given by

$$R_{xy}(\tau) = \Phi_{xx}\, g(\tau)$$

and $$E\{X^*Y\} = \Phi_{xx}\int_0^T \left(1 - \frac{\tau}{T}\right) g(\tau)\, e^{-j\omega\tau}d\tau$$

Hence $\overline{G}$ is given by

$$\overline{G} = \int_0^T \left(1 - \frac{\tau}{T}\right) g(\tau)\, e^{-j\omega\tau}d\tau \quad \text{.....(2.9)}$$

Two features of this expression lead to bias in the estimate; the upper limit of integration is finite and the second term within the brackets introduces an error term.

When the time of observation is sufficiently large relative to the settling time of the system the upper limit can be replaced by infinity and an approximate expression for the error in the gain estimate is

$$-\frac{1}{T}\int_0^\infty \tau g(\tau)\, e^{-j\omega\tau}d\tau$$

Using the Laplace Transform relationship

$$\int_0^\infty t\, f(t)\, e^{-st}dt = -\frac{dF(s)}{ds}$$

where F(s) is the Laplace Transform of f(t), this gives the error as

$$\frac{1}{T}\left.\frac{dG(s)}{ds}\right|_{s = j\omega}$$

or $$-\frac{j}{T}\frac{dG(j\omega)}{d\omega}$$

Hence $$\overline{G}(j\omega) = G(j\omega) - \frac{j}{T}\frac{dG(j\omega)}{d\omega} \qquad \text{.....(2.10)}$$

This expression can be interpreted graphically as shown in Fig. 2.6

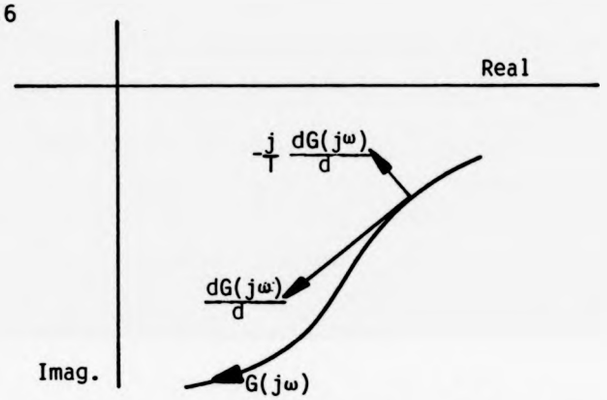


Fig. 2.6 Graphical interpretation of bias.

Here $G(j\omega)$ is plotted on the complex plane and $dG(j\omega)/d\omega$ is directed along the locus and multiplication by -j gives a clockwise rotation to direct the bias perpendicular to the locus. If $G(j\omega)$ is plotted at equal frequency intervals then the distance between adjacent frequency points gives an approximate indication of the magnitude of the bias.

2.3.2 An Expression for Bias - Method 2

In this method the estimate used is

$$\overline{G} = E\{\hat{G}\} = E\left\{\frac{X^*Y}{X^*X}\right\}$$

$$= E\left\{\frac{Y}{X}\right\} \text{ for one block.}$$

The input is considered as

$$x(t) = A\cos(\omega t + \theta)$$

where ω is a harmonic of the fundamental frequency and θ is a random phase angle uniformly distributed $0 - 2\pi$.

The input

$$x(t) = A\cos(\omega t + \theta)$$

is more conveniently written

$$x(t) = A \frac{e^{j(\omega t + \theta)} + e^{-j(\omega t + \theta)}}{2}$$

The short term Fourier Transform $X(j\omega)$ of this signal is

$$X(j\omega) = \frac{A}{2} e^{j\theta}$$

Writing the output $y(t)$ as a convolution of the input and the system impulse response $g(t)$,

$$y(t) = \int_0^t g(\tau)\, x(t - \tau)\, d\tau$$

This gives $Y(j\omega)$ as

$$Y(j\omega) = \frac{1}{T} \int_0^T \int_0^t g(\tau)\, x(t - \tau)\, d\tau\, e^{-j\omega t}\, dt$$

Interchanging the order of integration and altering the limits correspondingly,

$$Y(j\omega) = \frac{1}{T} \int_0^T g(\tau) \int_\tau^T x(t - \tau)\, e^{-j\omega t}\, dt\, d\tau$$

Substituting for $x(t - \tau)$ the inner integral becomes

$$\frac{A}{2} \int_\tau^T \left[e^{j(\omega[t - \tau] + \theta)}\, e^{-j\omega t} + e^{-j(\omega[t - \tau] + \theta)} e^{-j\omega t} \right] dt$$

Performing the integration and inserting limits gives

$$\frac{A}{2} \left[e^{j\theta}\, e^{-j\omega\tau}\, (T - \tau) + \frac{e^{-j\theta}\, e^{j\omega\tau}\, [e^{-2j\omega\tau} - 1]}{2j\omega} \right]$$

This gives $Y(j\omega)$ as

$$Y(j\omega) = \frac{A}{2}\int_0^T e^{j\theta}\, e^{-j\omega\tau}\,(1 - \tfrac{\tau}{T})\, g(\tau)\, d\tau$$

$$+ \frac{A}{2}\int_0^T \frac{e^{-j\theta}\, e^{j\omega\tau}\,[e^{-2j\omega\tau} - 1]\, g(\tau)}{2j\omega}\, d\tau$$

For one block $\hat{G}(j\omega) = Y(j\omega)/X(j\omega)$ giving

$$\hat{G}(j\omega) = \int_0^T g(\tau)\, e^{-j\omega\tau}\, d\tau - \frac{1}{T}\int_0^T \tau\, g(\tau)\, e^{-j\omega\tau}\, d\tau$$

$$+ \frac{e^{-2j\theta}}{T}\int_0^T \frac{(e^{-j\omega\tau} - e^{j\omega\tau})}{2j\omega}\, g(\tau)\, d\tau$$

Again assuming $g(\tau) \to 0$ as $t \to T$ the first term is the true response $G(j\omega)$. The second term is the bias term as evaluated previously. The integral in the last term becomes $[G(j\omega) - G^*(j\omega)]/2j\omega$ which equals $\mathrm{Im}\{G(j\omega)\}$.

Hence,

$$\hat{G}(j\omega) = G(j\omega) - \frac{j}{T}\frac{dG(j\omega)}{d\omega} + \frac{1}{\omega T}\, e^{-2j\theta}\, \mathrm{Im}\,[G(j\omega)] \quad \ldots\ldots\ldots (2.11)$$

For θ uniformly distributed the expected value of the last term is zero. This gives the same expression for bias as derived previously.

$$\bar{G} = E\{\hat{G}(j\omega)\} = G(j\omega) - \frac{1}{T}\frac{dG(j\omega)}{d\omega}$$

2.4 Variance of the estimate

Re-writing eqn. (2.8)

$$Y = \bar{G}X + N$$

then

$$\hat{G} = \frac{\sum_L X^*Y}{\sum_L X^*X} = \bar{G} + \frac{\sum_L X^*N}{\sum_L X^*X}$$

As stated in section (2.2), the expected value of the last term in the above expression is zero and it is the variance of this term that gives the variance in $\hat{G}$. Davall (4) has investigated the probability density of such a term by writing it in the form

$$\frac{\sum_{L} X^{*}N}{\sum_{L} X^{*}X} = M \angle\underline{\theta}$$

He then shows θ and M are independent with joint probability density function

$$p(M,\theta) = \frac{2LM\,(\sigma_x^2/\sigma_n^2)}{(1 + M^2\,\sigma_x^2/\sigma_n^2)^{L+1}}\,\frac{1}{2\pi} \qquad \text{.....(2.12)}$$

where σ_x^2 is the input variance and σ_n^2 is the variance of the equivalent noise input. It is worth noting that it is the independence of M and θ and the uniform distribution of θ that gives circular symmetry to the function and hence the zero mean value.

As the term has zero mean value, the variance in its magnitude is given by

$$\int_0^\infty p(M)\,M^2\,dM$$

$$= \int_0^\infty \frac{2LM^3\,(\sigma_x^2/\sigma_n^2)}{(1 + M^2\sigma_x^2/\sigma_n^2)^{L+1}}\,dM$$

This integral is listed in collections of definite integrals [Gradshteyn & Ryzik (6)] and it can be calculated giving

$$\frac{(\sigma_n^2/\sigma_x^2)}{L - 1} \qquad \text{.....(2.13)}$$

The method used by Davall (4) to evaluate the probability density function given in eqn. (2.12) uses the fact that the distribution of X and N is complex normal. This cannot be the case for $\omega = 0$ and the expression given in eqn. (2.12) is no longer true. An alternative method of obtaining the variance is now given which will include the case when $\omega = 0$.

As already shown

$$E\left[\frac{\sum_{L} X^*N}{\sum_{L} X^*X}\right] = 0$$

Hence the variance of this term is given by

$$\text{var.} = E\left[\left|\frac{\sum_{L} X^*N}{\sum_{L} X^*X}\right|^2\right]$$

$$= E\left[\frac{\sum_{i=1}^{L} X_i^* N_i \sum_{j=1}^{L} X_j N_j^*}{\left(\sum_{i=1}^{L} X_i^* X_i\right)^2}\right]$$

$$= E\left[\frac{\sum_{i=1}^{L} (X_i^* X_i)(N_i^* N_i) + \sum_{i=1}^{L}\sum_{\substack{j=1 \\ j\neq i}}^{L} (X_i^* X_j N_i N_j^* + X_i X_j^* N_i^* N_j)}{\left(\sum_{i=1}^{L} X_i^* X_i\right)^2}\right]$$

The expression can now be arranged as the sum of terms, each term being formed by a product of functions. One of the functions depends on X only, the other on N only and, as X and N are uncorrelated, the functions are uncorrelated. The expectation operator can be taken inside the summation sign and, because the functions are uncorrelated, they can be expressed as a product of expectations.

Hence

$$\text{var.} = \sum_{i=1}^{L}\left[E\left[\frac{X_i^* X_i}{\left(\sum_{k=1}^{L} X_k^* X_k\right)^2}\right] E(N_i^* N_i)\right]$$

$$+ \sum_{i=1}^{L}\sum_{\substack{j=1 \\ j\neq i}}^{L}\left[E\left[\frac{X_i^* X_j}{\left(\sum_{k=1}^{L} X_k^* X_k\right)^2}\right] E(N_i N_j^*) + E\left[\frac{X_i X_j^*}{\left(\sum_{k=1}^{L} X_k^* X_k\right)^2}\right] E(N_i^* N_j)\right]$$

However, as the components of N are uncorrelated,

$$E\{N_i N_j^*\} + E\{N_i^* N_j\} = 0$$

and $$E\{N_i^* N_i\} = \sigma_n^2$$

Interchanging the summation and expectation operations in the first term gives

$$\text{var.} = \sigma_n^2 E\left[\frac{\sum_{i=1}^{L} X_i^* X_i}{\left(\sum_{k=1}^{L} X_k^* X_k\right)^2}\right]$$

$$= \sigma_n^2 E\left[\frac{1}{\sum_{k=1}^{L} X_k^* X_k}\right]$$

Let $$u = \sum_{k=1}^{L} X_k^* X_k$$

When $\omega = 0$, $X_k = X_k^*$ and X_k has a zero mean normal distribution with variance σ_x^2.

When $\omega \neq 0$, the real and imaginary parts of X_k each have a zero mean, normal distribution with variance $\sigma_x^2/2$.

In either case the variable u has a x^2 distribution, when $\omega = 0$ it has L degrees of freedom, when $\omega \neq 0$ it has 2L degrees of freedom.

Hence,

$$p(u) = \frac{1}{2^{N/2}\sigma^N \Gamma(N/2)} \; u^{(N-2)/2} \; e^{-u/2\sigma^2} \qquad u > 0$$

Using the relationship

$$E\{g(x)\} = \int_{-\infty}^{+\infty} g(x)\, p(x)\, dx$$

then

$$E\left\{\frac{1}{u}\right\} = \frac{1}{2^{N/2}\sigma^N \Gamma(N/2)} \int_0^{\infty} u^{(N-4)/2} \; e^{-u/2\sigma^2} \; du$$

This can be evaluated with the aid of tables of integrals [Gradshteyn & Ryzik (6)] giving

$$E\left\{\frac{1}{U}\right\} = \frac{1}{2\sigma^2\left(\frac{N}{2} - 1\right)}$$

Hence the expressions for variance become

$$\text{var.} = \frac{\sigma_n^2}{\sigma_x^2} \cdot \frac{1}{(L - 2)} \qquad \text{when } \omega = 0 \quad \ldots\ldots\ldots\ldots (2.14)$$

$$= \frac{\sigma_n^2}{\sigma_x^2} \cdot \frac{1}{(L - 1)} \qquad \text{when } \omega \neq 0 \quad \ldots\ldots\ldots\ldots (2.15)$$

It can be noted that, although the expected value of the frequency response estimate is independent of L, the number of blocks, the variance of the estimate does depend upon L. Also the variance is infinite for a single block at all frequencies and is infinite at $\omega = 0$ for two blocks. Although in theory an estimate can be obtained for a single block, (two blocks for $\omega = 0$), in practice this estimate would be of little use.

If the probability density function p(M), as given by eqn. (2.12), is plotted on a line of constant θ passing through $\bar{G}$ the resultant function is as shown in Fig. 2.7.

The variance as calculated is not a very meaningful measure of the spread of this function. More useful in practice is the drawing of confidence regions.

The probability p_1 that a point will lie in a circle, radius M_1, centred on $\bar{G}$ is given by

$$p_1 = \int_0^{2\pi} \int_0^{M_1} p(M,\theta)\, dM d\theta$$

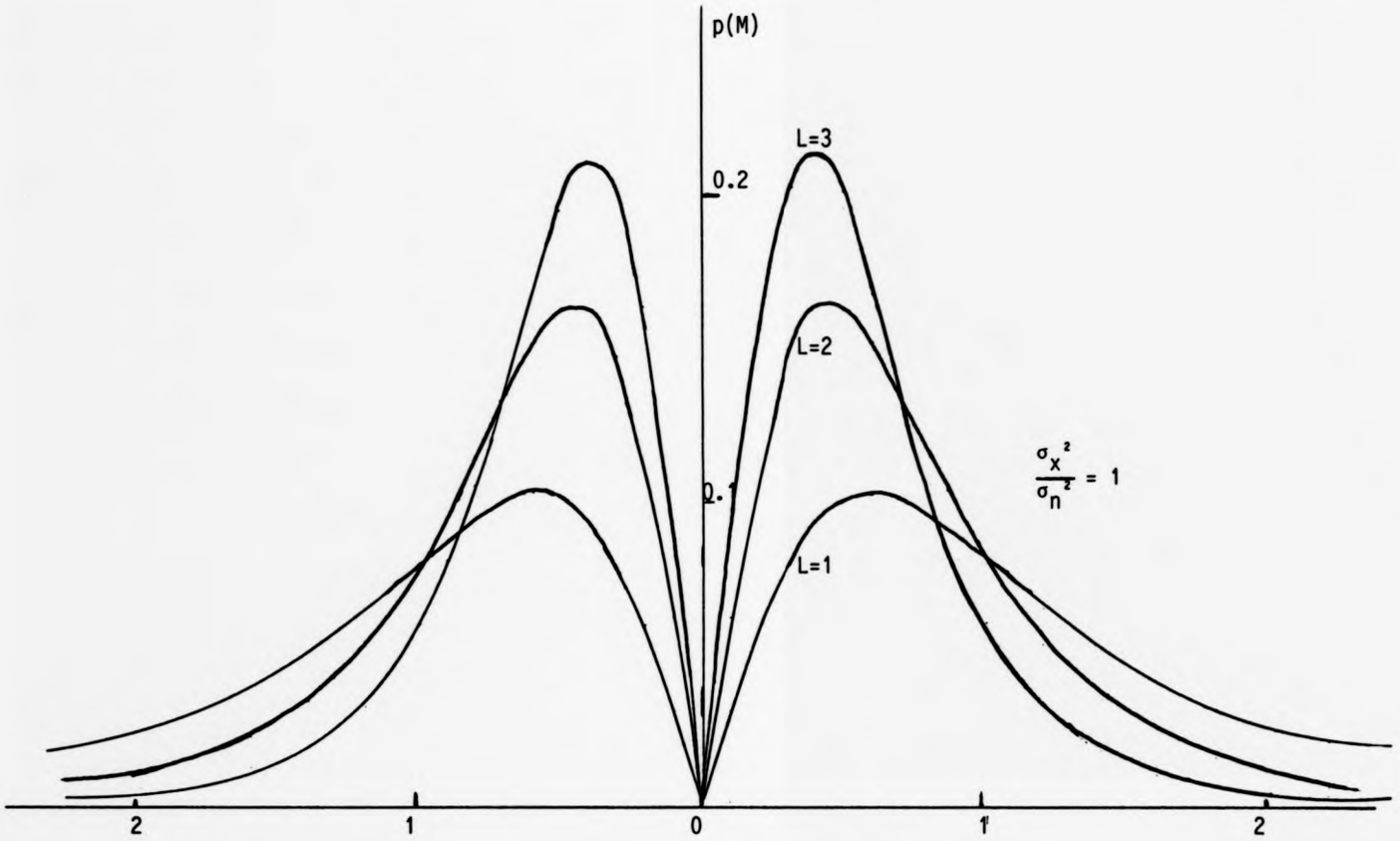


Fig. 2.7 Probability density function p(M,θ) for fixed θ

Performing the integration gives

$$p_1 = 1 - \frac{1}{\left(1 + \frac{\sigma_x^2}{\sigma_n^2} M_1^2\right)^L}$$

or re-arranging to give M_1 in terms of p_1

$$M_1 = \sqrt{\left(\frac{1}{(1 - p_1)^{\frac{1}{L}}} - 1\right)\frac{\sigma_n^2}{\sigma_x^2}} \quad \text{.....(2.16)}$$

In both the expressions for variance (eqns. 2.14 and 2.15) and confidence intervals (eqn. 2.16) it is required to obtain a value for σ_n^2/σ_x^2. This can be done by reference to eqn. (2.8)

$$Y = \bar{G}X + N$$

Multiplying through each side by its conjugate and taking expected values gives

$$E\{YY^*\} = |\bar{G}|^2 E\{XX^*\} + E\{NN^*\} + \bar{G} E\{XN^*\} + \bar{G}^* E\{X^*N\}$$

But $E\{XN^*\} = E\{X^*N\} = 0$ and $E\{YY^*\} = \sigma_y^2$,

$E\{X^*X\} = \sigma_x^2$, $E\{N^*N\} = \sigma_n^2$ giving

$$\sigma_y^2 = |\bar{G}|^2 \sigma_x^2 + \sigma_n^2$$

$$\frac{\sigma_n^2}{\sigma_x^2} = \frac{\sigma_y^2}{\sigma_x^2} - |\bar{G}|^2 \quad \text{.....(2.17)}$$

Note that σ_y^2 and σ_x^2 are the <u>measured</u> output and input powers respectively.

2.5 A specific example

The system chosen to illustrate the results obtained in this chapter has a first order transfer function. This in no way restricts the method but the resulting analytical solutions can be derived simply. Also the transients associated with a first order system are particularly simple (there only being one initial condition) hence one can obtain a physical insight into the nature of the transients and their effects upon the measured frequency response function.

It should be stressed that the method will handle the general n^{th} order system. More complex transfer functions will be considered in later examples with numerical solutions.

Theoretical results derived are compared to those obtained by digital simulation.

The system chosen has an open-loop transfer function G(s) given by,

$$G(s) = \frac{1}{1 + sT_1}$$

where T_1 = 5 seconds.

The observation time T was chosen to be the shortest that would still enable the approximate solution to be applied, hence $T \approx 5T_1$. The exact value of T was chosen as 25.6 seconds, this with a sampling interval of 0.1 second gave 256 samples in a block, a power of 2 as required by the FFT algorithm. This value of T gives a fundamental frequency of 0.039 Hz (angular frequency 0.245 rad/s) and results are obtained at harmonics of this frequency.

Initially results are obtained for white noise input and a rectangular or "do-nothing" window. Later results are presented for non-white input and a Hanning window.

2.5.1 Open-Loop Bias

Using the full expression for bias, eqn. 2.9,

$$\bar{G}(j\omega) = \int_0^T (1 - \frac{\tau}{T})\, g(\tau)\, e^{-j\omega\tau}\, d\tau$$

where ω is a multiple of the fundamental, $\omega = m\omega_0$.

The impulse response, $g(\tau)$, for the system under consideration is given by,

$$g(\tau) = \frac{e^{-\tau/T_1}}{T_1}$$

Substituting and taking the terms in the bracket separately (noting that $e^{-jm\omega_0 T} = 1$)

$$\int_0^T \frac{e^{-\tau/T_1}}{T_1}\, e^{-j\omega\tau}\, d\tau = \frac{(1 - e^{-T/T_1})}{(1 + j\omega T_1)}$$

The integral involving the second term can be evaluated using the standard form

$$\int xe^{ax} = \frac{(ax - 1)}{a^2}\, e^{ax} + \text{constant}$$

giving the second term as

$$-\int_0^T \frac{\tau}{T}\, \frac{e^{-\tau/T_1}}{T_1}\, e^{-j\omega\tau}\, d\tau$$

$$= \frac{1}{T_1}\, \frac{[(1 + j\omega T_1)\, T - T_1]e^{-T/T_1}}{(1 + j\omega T_1)^2} - \frac{T_1}{T}\, \frac{1}{(1 + j\omega T_1)^2}$$

If the first and second terms are combined this gives

$$\bar{G}(j\omega) = \frac{1}{1 + j\omega T_1} - \frac{T_1}{T}\, \frac{(1 - e^{-T/T_1})}{(1 + j\omega T_1)^2}$$

The approximate expression for $\bar{G}(j\omega)$ is given by eqn. 2.10

$$\bar{G}(j\omega) = G(j\omega) + \frac{1}{T}\, \frac{dG(s)}{ds}\Big|_{s = j\omega}$$

Noting that

$$\frac{dG(s)}{ds} = -\frac{1}{(1 + sT_1)^2}$$

and substituting $s = j\omega$ gives

$$\bar{G}(j\omega) = G(j\omega) - \frac{T_1}{T} \cdot \frac{1}{(1 + j\omega T_1)^2}$$

It can be seen that the two methods of calculating $\bar{G}(j\omega)$ give expressions that differ by the purely real term $(1 - e^{-T/T_1})$. As $T_1/T = 0.195$ this gives the term a value of 0.994 - the two expressions differ by approximately 0.6%.

Table 2.1 gives the theoretical bias against harmonic number and angular frequency. Simulation results are also given for 2 blocks and 4 blocks of data (theoretically the results should not depend upon block number). The results shown were obtained for 2000 experiments (repeating for another 2000 gave negligible difference).

Table 2.1

Harmonic No. m	Frequency $=m\omega_0$	True Response $G(j\omega)$	Biased response $\bar{G}(j\omega)$		
			Theoretical	Simulation L=2	Simulation L=4
0	0	1.000-j0	0.805-j0	0.827-j0	0.806-j0
1	0.245	0.399-j0.490	0.415-j0.413	0.415-j0.414	0.415-j0.413
2	0.491	0.142-j0.349	0.162-j0.330	0.164-j0.329	0.165-j0.328
3	0.763	0.069-j0.253	0.080-j0.246	0.082-j0.247	0.080÷j0.247
4	0.982	0.040-j0.200	0.047-j0.193	0.045-j0.193	0.046-j0.192
5	1.227	0.026-j0.159	0.036-j0.157	0.029-j0.157	0.029-j0.156
6	1.473	0.018-j0.133	0.022-j0.132	0.018-j0.134	0.018-j0.134
7	1.718	0.013-j0.115	0.016-j0.114	0.014-j0.115	0.014-j0.115
8	1.963	0.010-j0.101	0.012-j0.100	0.011-j0.101	0.010-j0.101
9	2.209	0.008-j0.090	0.010-j0.090	0.008-j0.090	0.008-j0.090

Fig. 2.8 shows the true results and the biased estimates. Only the theoretical results have been shown, as the difference between these and results obtained by simulation does not show on the scales chosen.

As stated earlier in the chapter, the bias can be interpreted graphically - it is directed perpendicular to the locus. This is true for all lengths of record for the first order system. (It is correct for higher-order systems only when the record length greatly exceeds the system settling time). For this example the locus is a semicircle and the bias should be directed towards the centre. This can be confirmed by reference to Fig. 2.8 where lines have been drawn from the true points towards the centre point 0.5 + j0.

For this particular system the nature of the bias is open to a simple interpretation. Referring to eqn. 2.11,

$$\hat{G}(j\omega) = G(j\omega) - \frac{j}{T}\frac{dG(j\omega)}{d\omega} + \frac{1}{\omega T} e^{-2j\theta}\, \mathrm{Im}[G(j\omega)]$$

where θ is the phase angle of the measured input component at the frequency ω.

Substituting $$G(j\omega) = \frac{1}{1 + j\omega T_1}$$

this gives

$$\hat{G}(j\omega) = G(j\omega) - \frac{T_1}{T} \cdot \frac{1}{(1 + j\omega T_1)^2} - \frac{T_1}{T} \frac{e^{-2j\theta}}{1 + (\omega T_1)^2} \quad \text{.....(2.18)}$$

It can be seen that the magnitude of the final term is equal to the magnitude of the bias. Hence, for an ensemble of experiments <u>with fixed θ</u> the estimate $\hat{G}(j\omega)$ would lie on a circle as shown in Fig. 2.9.

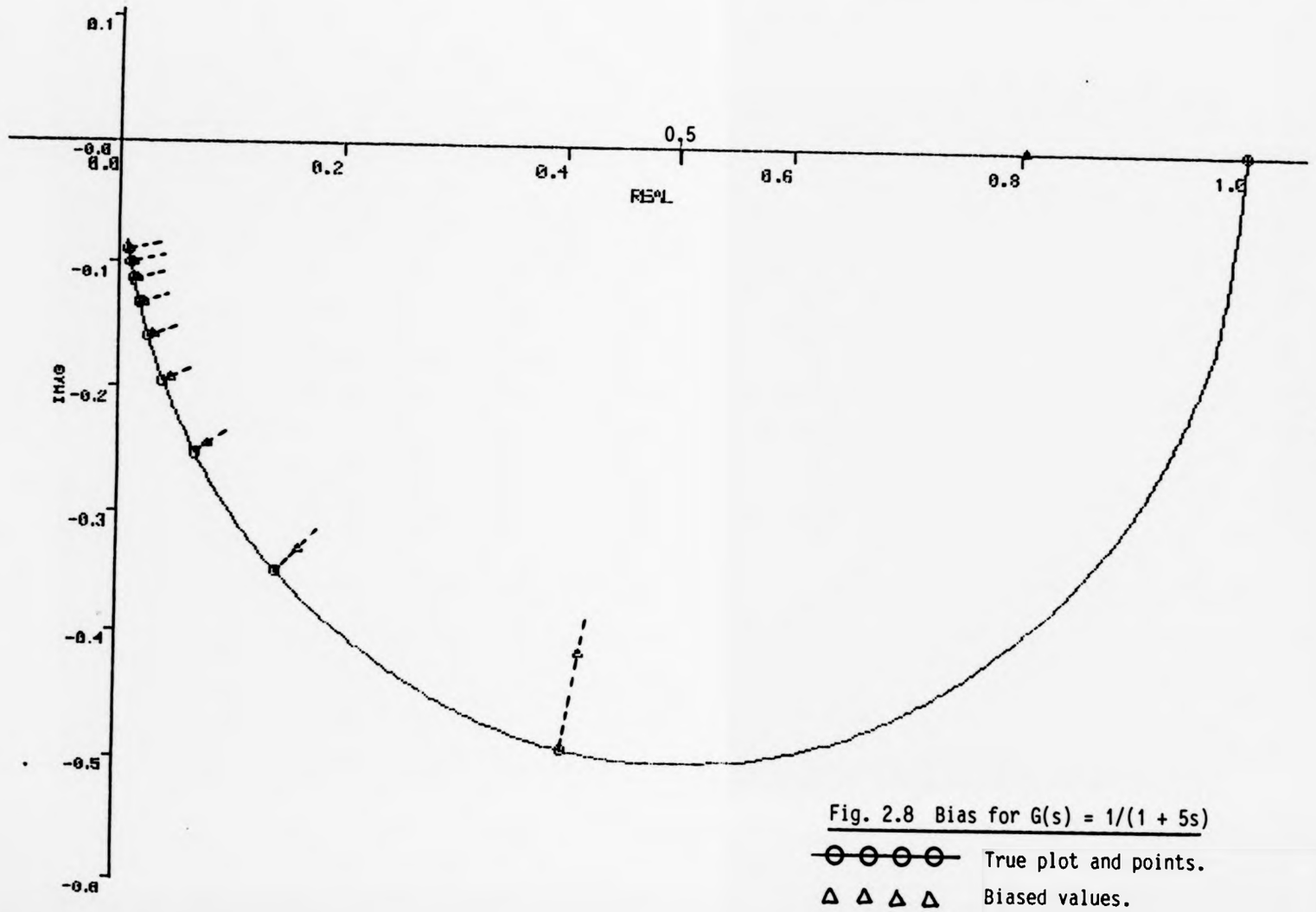


Fig. 2.8 Bias for G(s) = 1/(1 + 5s)

True plot and points.

Biased values.

Lines drawn towards point 0.5 + j0

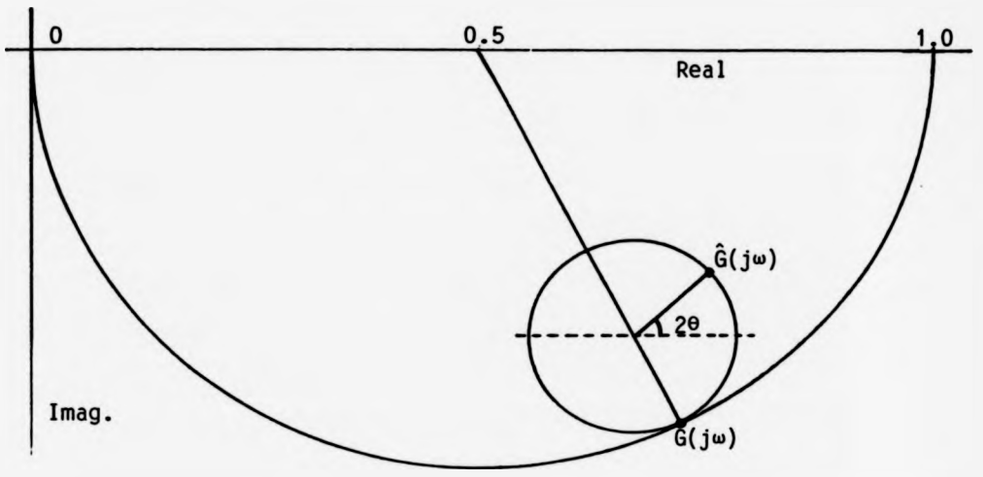


Fig. 2.9 Estimate $\hat{G}(j\omega)$ for an ensemble of experiments with fixed θ

For this particular system eqn. 2.18 can be derived from first principles by direct consideration of the transients. Consider an input

$$x(t) = A \cos(\omega t + \theta)$$

where A and θ are random variables. Let the output produced at $t = 0$ by this signal applied prior to $t = 0$ be h_0 (h_0 is a zero mean random variable). The resulting transient $h_0(t)$ is given by

$$h(t) = h_0 e^{-t/T_1}$$

The output produced at $t = T$ by the signal $x(t)$ is given by

$$h_T = |G(j\omega)| A \cos(\theta + \phi)$$

where $G(j\omega) = |G(j\omega)| \angle \phi$. The output signal produced after $t = T$ is given by

$$h_T(t) = [|G(j\omega)| A \cos(\theta + \phi)] e^{-t/T_1}$$

The complete transient h(t) is

$$h(t) = [h_0 - |G(j\omega)| A \cos(\theta + \phi)] e^{-t/T_1}$$

and taking Fourier Transforms this gives

$$H(j\omega) = \frac{1}{T} [h_0 - |G(j\omega)| A \cos(\theta + \phi)] \frac{T_1}{1 + j\omega T_1}$$

assuming that $e^{-T_1/T} = 1$.

Writing $\cos(\theta + \phi)$ as $(e^{j\theta}e^{j\phi} + e^{-j\theta}e^{-j\phi})$ and taking the Fourier Transform of the input as $Ae^{j\theta}/2$, this gives

$$\hat{G}(j\omega) = G(j\omega) + \frac{2}{A} \frac{T_1}{T} h_0 - \frac{T_1}{T} [G(j\omega)]^2 - \frac{T_1}{T} e^{-j2\theta} |G(j\omega)|^2$$

For an ensemble of experiments the mean value of h_0 is zero and this gives $\bar{G}(j\omega)$ identical to eqn. (2.18).

2.5.2 Variance and Confidence Limits

The variance of the estimate can be calculated using eqn. (2.15) for values of ω other than zero.

$$\sigma^2 = \frac{(\sigma_n/\sigma_x)^2}{L - 1}$$

and for $\omega = 0$ from eqn. (2.14)

$$\sigma^2 = \frac{(\sigma_n/\sigma_x)^2}{L - 2}$$

To use these formulae $(\sigma_n/\sigma_x)^2$ is required and this can be obtained from eqn. (2.16)

$$\frac{\sigma_n^2}{\sigma_x^2} = \frac{\sigma_y^2}{\sigma_x^2} - |\bar{G}(j\omega)|^2$$

$|\bar{G}(j\omega)|^2$ can easily be obtained from the values of $\bar{G}(j\omega)$ already calculated, and listed in table 2.1, and σ_y^2/σ_x^2 can be obtained by consideration of the measured autospectrum at the output.

For a white noise input having power per unit bandwidth σ_x^2 then elementary analysis shows that

$$R_{yy}(\tau) = \frac{\sigma_x^2}{2T_1} e^{-|\tau|/T_1}$$

The mean value of the measured autospectrum is given by

$$\sigma_y^2 = E\{\Phi_{yy}\}$$

$$= \int_{-T}^{+T} (1 - \frac{|\tau|}{T}) . \frac{\sigma_x^2}{2T_1} e^{-|\tau|/T_1} e^{-j\omega\tau} d\tau$$

giving

$$\frac{\sigma_y^2}{\sigma_x^2} = \frac{1}{1 + \omega^2 T_1^2} - \frac{T_1}{T} \frac{(1 - \omega^2 T_1^2)(1 - e^{-T/T_1})}{(1 + \omega^2 T_1^2)^2}$$

Table 2.2 gives values of σ_y^2/σ_x^2 and σ_n^2/σ_x^2 against harmonic number and angular frequency. Table 2.3 gives theoretical figures for variance with two blocks (L = 2) and four blocks (L = 4). These results are compared with those obtained by simulation. Simulation was for 2000 experiments and two simulation runs were undertaken. The results exhibited negligible differences except for $\omega = 0$; for $\omega = 0$ the two results are given. It is submitted that these are not incompatible with theoretical predictions.

Table 2.2

Harmonic Number m	Angular Frequency = $m\omega_0$	σ_y^2/σ_x^2	σ_n^2/σ_x^2
0	0.0	0.8050	0.1571
1	0.245	0.4142	0.0725
2	0.491	0.1621	0.0275
3	0.763	0.0802	0.0136
4	0.982	0.0470	0.0077
5	1.227	0.0306	0.0050
6	1.473	0.0215	0.0035
7	1.718	0.0159	0.0026
8	1.963	0.0122	0.0020
9	2.209	0.0097	0.0015

Table 2.3

Harmonic Number m	Angular Frequency = $m\omega_0$	Variance in $\hat{G}(j\omega)$			
		L = 2		L = 4	
		Theoretical	Simulation	Theoretical	Simulation
0	0	∞	0.4672	0.0785	0.0697
0	0	∞	0.7002	0.0785	0.0807
1	0.245	0.0725	0.0672	0.0242	0.0240
2	0.491	0.0275	0.0243	0.0092	0.0088
3	0.763	0.0136	0.0113	0.0045	0.0042
4	0.982	0.0077	0.0074	0.0025	0.0024
5	1.227	0.0050	0.0049	0.0016	0.0017
6	1.473	0.0035	0.0033	0.0012	0.0012
7	1.718	0.0026	0.0023	0.0008	0.0008
8	1.963	0.0020	0.0019	0.0007	0.0007
9	2.209	0.0015	0.0014	0.0005	0.0005

Confidence intervals can be obtained by use of eqn. (2.16) which applies at all frequencies except $\omega = 0$.

$$M_1 = \sqrt{\left(\frac{1}{(1 - p_1)^{1/L}} - 1\right)\frac{\sigma_n^2}{\sigma_x^2}}$$

Table 2.4 gives radii of circles representing 90% confidence intervals ($p_1 = 0.9$) for L = 2 and L = 4. Also given are simulated results giving the percentage of estimates that lay in these circles based on 4000 experiments. These circles are drawn with centres equal to the biased values and are portrayed in Fig. 2.10 for L = 2 and in Fig. 2.11 for L = 4.

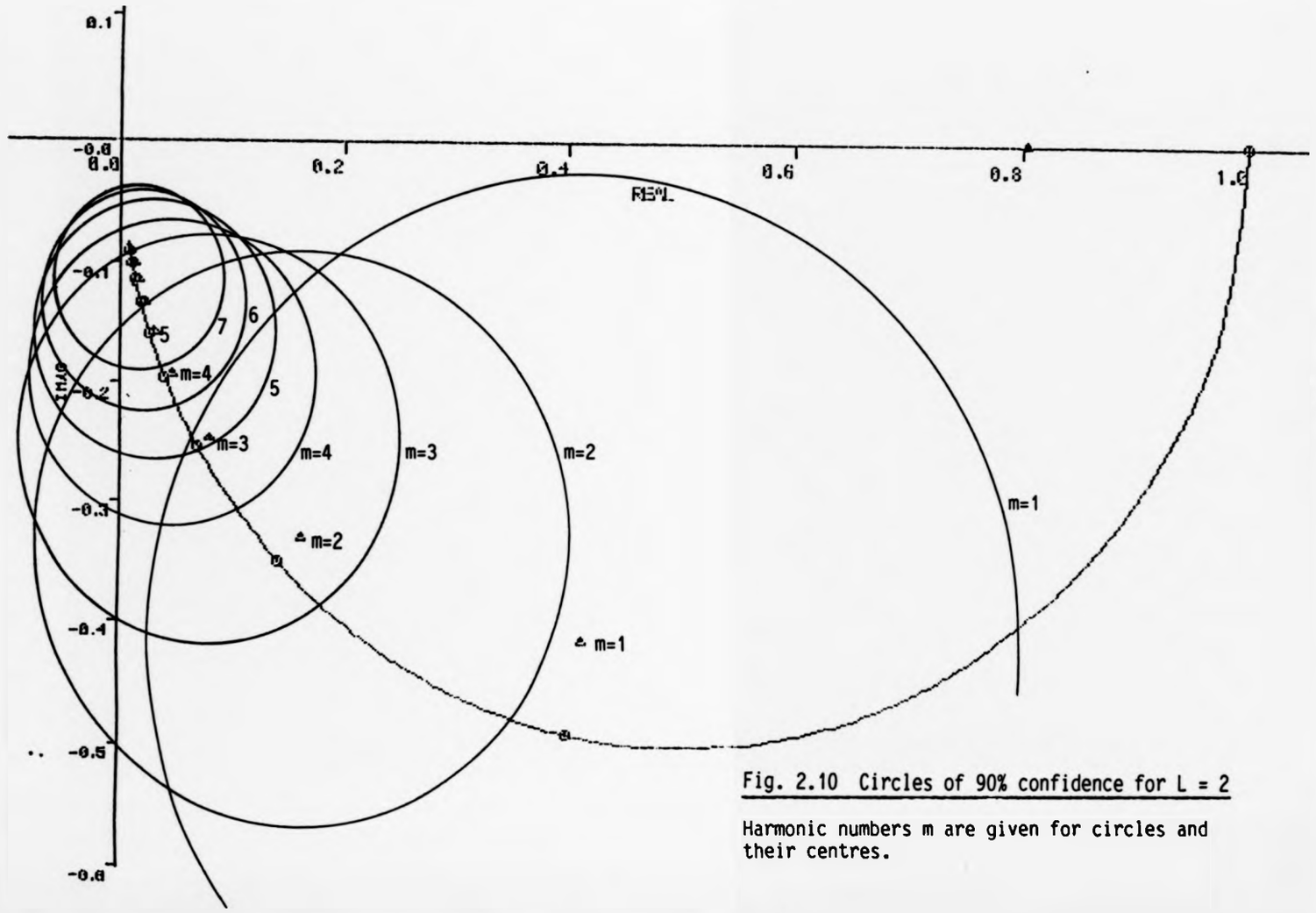


Fig. 2.10 Circles of 90% confidence for L = 2

Harmonic numbers m are given for circles and their centres.

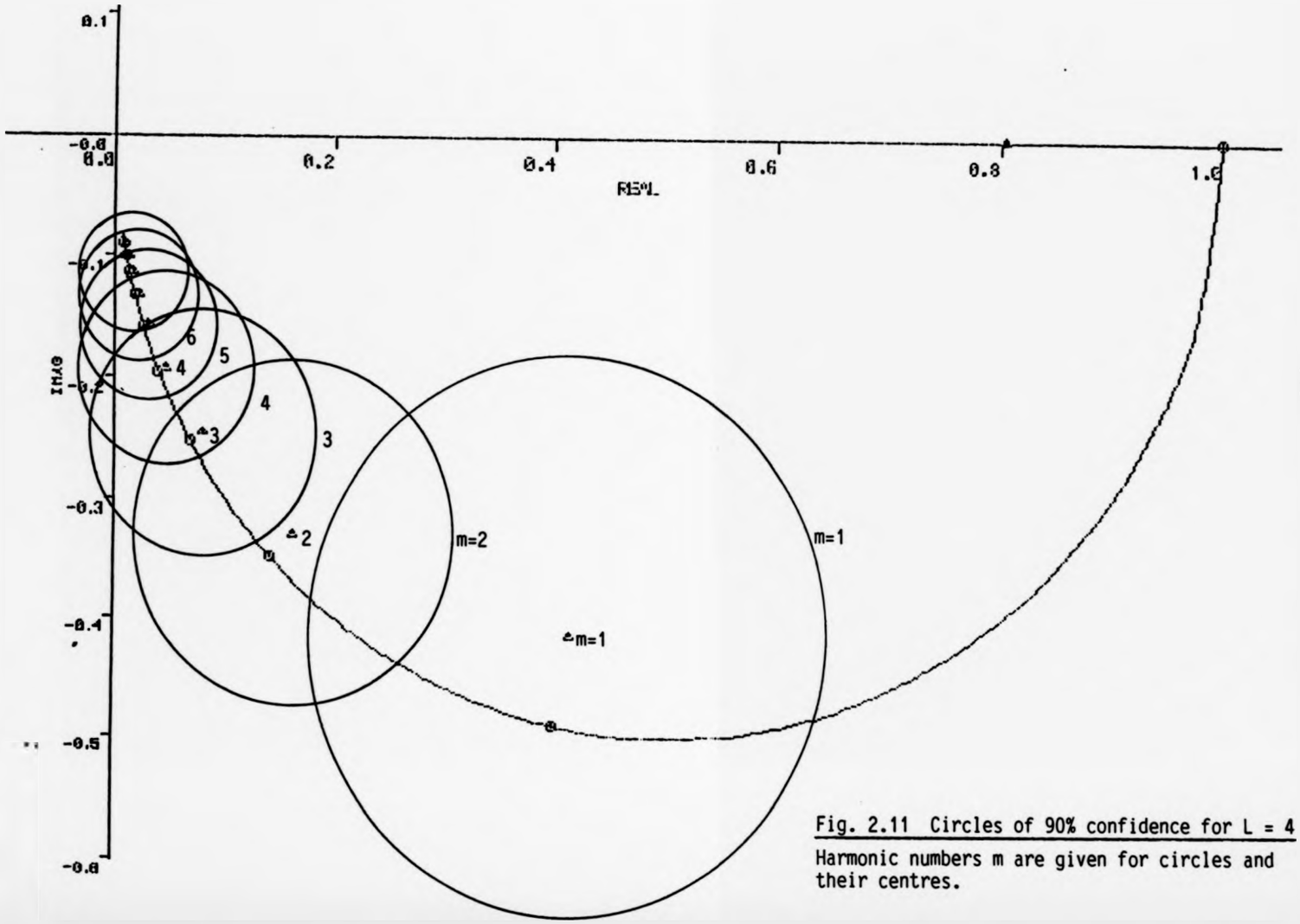


Fig. 2.11 Circles of 90% confidence for L = 4

Harmonic numbers m are given for circles and their centres.

Table 2.4

Harmonic Number m	Angular Frequency $\omega = m\omega_0$	Radius for 90% confidence		% of points within circle by simulation	
		L = 2	L = 4	L = 2	L = 4
1	0.245	0.395	0.237	84.8	90.1
2	0.491	0.243	0.146	91.1	89.7
3	0.763	0.171	0.102	91.0	90.1
4	0.982	0.129	0.077	89.6	89.5
5	1.227	0.103	0.062	90.0	90.0
6	1.473	0.087	0.052	88.7	88.7
7	1.718	0.075	0.045	89.2	89.3
8	1.963	0.066	0.039	89.4	90.5
9	2.209	0.057	0.034	89.5	88.7

2.5.3 Non-White Input

The expressions derived for bias and variance apply to signals taken from independent blocks. The expressions are true for successive blocks if the input is white noise but would not be correct for non-white input. However the assumption of independence gives results that are useful for comparison purposes with the non-white situation.

The general expression for the expected value of the frequency response estimate is given by

$$\bar{G} = \frac{\bar{\Phi}_{xy}}{\bar{\Phi}_{xx}} = \frac{\int_{-T}^{T}(1 - \left|\frac{\tau}{T}\right|) R_{xy}(\tau)\, e^{-j\omega\tau}\, d\tau}{\int_{-T}^{T}(1 - \left|\frac{\tau}{T}\right|) R_{xx}(\tau)\, e^{-j\omega\tau}\, d\tau} \quad \text{.....(2.19)}$$

For non-white input $R_{xx}(\tau) \neq \sigma^2\delta(\tau)$ and the simplifications used earlier cannot be applied. The specific example considered is shown in Fig. 2.12.

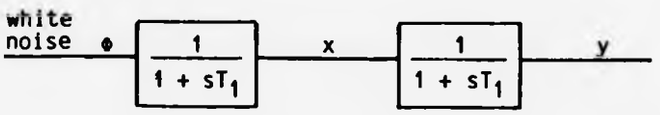


Fig. 2.12
Band limited noise into first order system

Here band limited white noise is used as the input to the first order system investigated earlier (for simplicity it is assumed that the time constant of the noise filter is equal to the time constant of the system; this in no way restricts the generality of the argument).

Now the signal at x has autocorrelation function

$$R_{xx}(\tau) = \frac{\Phi}{2T_1} e^{-|\tau|/T_1}$$

where Φ is the constant spectral density of the white noise before the filter.

The system impulse response h(t) is given by

$$h(t) = \frac{e^{-t/T_1}}{T_1}$$

and using the relationship

$$R_{xy}(\tau) = R_{xx}(\tau) * h(\tau)$$

this gives

$$R_{xy}(\tau) = \frac{\Phi}{4T_1} e^{\tau/T_1} \qquad \text{for } \tau < 0$$

$$R_{xy}(\tau) = \frac{\Phi}{2T_1^2} \left(\frac{T_1}{2} + \tau\right) e^{-\tau/T_1} \qquad \text{for } \tau > 0$$

The most convenient method of evaluating the numerator and denominator in eqn. (2.19) is to write $R_{xy}(\tau)$ as the sum of odd and even parts (valid for all τ). To evaluate the real part of the numerator only the even part need be considered and to evaluate

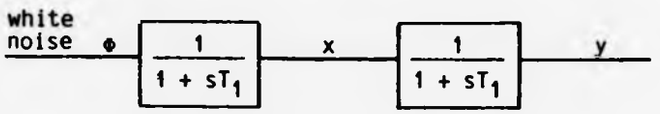


Fig. 2.12
Band limited noise into first order system

Here band limited white noise is used as the input to the first order system investigated earlier (for simplicity it is assumed that the time constant of the noise filter is equal to the time constant of the system; this in no way restricts the generality of the argument).

Now the signal at x has autocorrelation function

$$R_{xx}(\tau) = \frac{\Phi}{2T_1} e^{-|\tau|/T_1}$$

where Φ is the constant spectral density of the white noise before the filter.

The system impulse response h(t) is given by

$$h(t) = \frac{e^{-t/T_1}}{T_1}$$

and using the relationship

$$R_{xy}(\tau) = R_{xx}(\tau) * h(\tau)$$

this gives

$$R_{xy}(\tau) = \frac{\Phi}{4T_1} e^{\tau/T_1} \qquad \text{for } \tau < 0$$

$$R_{xy}(\tau) = \frac{\Phi}{2T_1^2} \left(\frac{T_1}{2} + \tau\right) e^{-\tau/T_1} \qquad \text{for } \tau > 0$$

The most convenient method of evaluating the numerator and denominator in eqn. (2.19) is to write $R_{xy}(\tau)$ as the sum of odd and even parts (valid for all τ). To evaluate the real part of the numerator only the even part need be considered and to evaluate

the imaginary part only the odd part need be used. This gives the following expressions (it has been assumed that $e^{-T_1/T} \ll 1$).

$$\text{Re } \{\bar{\Phi}_{xy}\} = \Phi \left[\frac{1}{2(1+\omega^2 T_1^2)} + \frac{(1-\omega^2 T_1^2)}{2(1+\omega^2 T_1^2)^2} - \frac{T_1}{T} \frac{(1-3\omega^2 T_1^2)}{(1+\omega^2 T_1^2)^3} \right]$$

$$\text{Im } \{\bar{\Phi}_{xy}\} = \Phi \left[\frac{-\omega T_1}{(1+\omega^2 T_1^2)} + \frac{T_1}{T} \frac{(3\omega T_1 - \omega^3 T_1^3)}{(1+\omega^2 T_1^2)^3} \right]$$

$$\text{Re } \{\bar{\Phi}_{xx}\} = \Phi \left[\frac{1}{(1+\omega^2 T_1^2} - \frac{T_1}{T} \frac{(1-\omega^2 T_1^2)}{(1+\omega^2 T_1^2)} \right]$$

$$\text{Im } \{\bar{\Phi}_{xx}\} = 0$$

With T_1 = 5 seconds, the results obtained from these formulae are compared with the results obtained from simulation in table 2.5. The results are for 4 blocks (L = 4) and the simulation results are based on 2000 experiments.

Table 2.5

Harmonic Number m	Angular Frequency $\omega = m\omega_0$	True Response $G(j\omega)$	Biased response $\bar{G}(j\omega)$	
			Theoretical	Simulation L = 4
0	0	1.000-j0	0.885-j0	0.882-j0
1	0.245	0.399-j0.490	0.514-j0.415	0.506-j0.424
2	0.491	0.142-j0.349	0.252-j0.332	0.242-j0.341
3	0.763	0.069-j0.253	0.169-j0.247	0.154-j0.247
4	0.982	0.040-j0.200	0.135-j0.193	0.123-j0.196
5	1.227	0.026-j0.159	0.119-j0.157	0.105-j0.155
6	1.473	0.018-j0.133	0.110-j0.132	0.096-j0.136
7	1.718	0.013-j0.155	0.104-j0.114	0.090-j0.114
8	1.963	0.010-j0.101	0.100-j0.100	0.086-j0.101

The variance of the estimates can be obtained using eqns. 2.14, 2.15 and 2.17. σ_y^2 can be obtained as

$$\sigma_y^2 = E\{\Phi_{yy}\} = \int_{-T}^{+T} (1 - \frac{|\tau|}{T}) R_{yy}(\tau) e^{-j\omega\tau} d\tau$$

and

$$R_{yy}(\tau) = R_{xy}(\tau) * h(-\tau)$$

This gives

$$R_{yy}(\tau) = \frac{\Phi}{4T_1^2} (T_1 + |\tau|) e^{-|\tau|/T_1}$$

and

$$\sigma_y^2 = \Phi\left[\frac{1}{2(1 + \omega^2 T_1^2)} + \frac{(1 - \omega^2 T_1^2)}{2(1 + \omega^2 T_1^2)^2} - \frac{T_1}{T} \frac{(1 - 3\omega^2 T_1^2)}{(1 + \omega^2 T_1^2)^2}\right]$$

Table 2.6 gives σ_y^2/σ_x^2, σ_n^2/σ_x^2, variance for L = 2 and L = 4, against harmonic number. For variance, theoretical and simulation results are given; again simulation was for 2000 experiments and two simulation runs were performed. Differences between runs were negligible except at $\omega = 0$ and both results are given at this frequency.

Table 2.6

Harmonic Number m	Frequency $m\omega_0$	σ_y^2/σ_x^2	σ_n^2/σ_x^2	Variance L = 2		Variance L = 4	
				Theory	Simulation	Theory	Simulation
0	0	0.885	0.108	∞	0.273	0.054	0.045
0	0	0.885	0.108	∞	0.278	0.054	0.054
1	0.245	0.514	0.078	0.078	0.075	0.026	0.026
2	0.491	0.252	0.078	0.078	0.075	0.026	0.024
3	0.763	0.169	0.079	0.079	0.065	0.026	0.024
4	0.982	0.135	0.080	0.080	0.065	0.027	0.023
5	1.227	0.119	0.080	0.080	0.070	0.027	0.025
6	1.473	0.110	0.080	0.080	0.065	0.027	0.024
7	1.718	0.104	0.080	0.080	0.066	0.027	0.023
8	1.963	0.100	0.080	0.080	0.067	0.027	0.024

As can be seen from this example, the labour involved in calculating the biased estimate and its variance is considerable. The system was first order with the input signal obtained by passing white noise through a first order filter. For higher order systems and for an input having a more complex spectrum the calculations become so involved that it is not feasible to obtain results from direct calculation. Prof. J. L. Douce (University of Warwick) has developed a program that calculates the bias and variance of the estimate for a general frequency response function using any prescribed input spectrum and the program can take into account the effect of data windows at input and output. I am indebted to Prof. Douce for the use of this program to check the results derived in this example, and for its use to obtain the results in the next example.

2.5.4 Effect of Hanning Window

This example is the first order system with white noise input that was investigated previously; however a Hanning Window is used to modify the input and output data. Introduction of this window modifies the measured input and output spectral density. (For the special case of white noise input the measured input spectral density remains as before).

The expected value of the cross spectral density now becomes

$$\bar{\Phi}_{xy} = \int_{-T}^{+T} R_{xy}(\tau)\, e^{-j\omega\tau} \left[\int_{-\infty}^{+\infty} w(t_2)\, w(t_2 - \tau)\, dt_2 \right] d\tau$$

where for the Hanning Window

$$w(t) = (1 - \cos 2\pi t/T)$$

and the inner integral becomes

$$T \left[(1 - \tfrac{\tau}{T})(1 + \tfrac{1}{2} \cos \omega_0 \tau) + \frac{3}{4\pi} \sin \omega_0 \tau \right]$$

where $\omega_0 = 2\pi/T$.

Using $R_{yy}(\tau)$ in place of $R_{xy}(\tau)$ gives a similar expression for $\bar{\Phi}_{yy}$.

The calculation of $\bar{G}(j\omega)$ and its variance now follows the previous examples. In this case the results were obtained using the program quoted in the previous section. The results and those obtained by simulation are given in Table 2.7.

Table 2.7

Harmonic Number m	Frequency $m\omega_0$	σ_n^2/σ_x^2	Biased Estimate $\bar{G}(j\omega)$		Variance L = 2		Variance L = 4	
			Theoretical	Simulation	Theoretical	Simulation	Theoretical	Simulation
0	0	0.180901	0.7582-j0	0.7473-j0	∞	0.729890	0.090451	0.083154
1	0.245	0.087742	0.4744-j0.3951	0.4633-j0.4008	0.087742	0.091205	0.029247	0.032102
2	0.491	0.015334	0.1828-j0.3548	0.1712-j0.3572	0.015334	0.015522	0.005111	0.004892
3	0.763	0.003112	0.0868-j0.2588	0.0761-j0.2591	0.003112	0.002906	0.001037	0.001052
4	0.982	0.000940	0.0528-j0.1987	0.0426-j0.1991	0.000940	0.000940	0.000313	0.000308
5	1.227	0.000375	0.0372-j0.1604	0.0267-j0.1603	0.000375	0.000369	0.000125	0.000117
6	1.473	0.000178	0.0288-j0.1342	0.0217-j0.1344	0.000178	0.000172	0.000060	0.000058
7	1.718	0.000095	0.0238-j0.1153	0.0210-j0.1148	0.000095	0.000092	0.000032	0.000034
8	1.963	0.000055	0.0206-j0.1010	0.0194-j0.1003	0.000055	0.000057	0.000018	0.000012
9	2.209	0.000034	0.0183-j0.0898	0.0179-j0.0881	0.000034	0.000031	0.000012	0.000010

References - Chapter 2

1. Douce, J.L., "A Novel Direct Method for Frequency Response Measurement", to be published - Proc. IEE.

2. Goodman, N.R., "On the joint estimation of the spectra, cospectrum and quadrature spectrum of a two dimensional stationary Gaussian process", PhD Thesis, Princeton University, 1957.

3. Papoulis, A., "Probability, Random Variables and Stochastic Processes", McGraw-Hill, 1965.

4. Davall, P.W., "Applications of statistics in the spectral analysis of time-varying systems", PhD Thesis, University of Warwick, 1975.

5. Jenkins, G.M. and Watts, D.G., "Spectral Analysis and its Applications", Holden Day, 1968.

6. Gradshteyn, I.S. and Ryzhik, I.M., "Table of Integrals series and products", Academic Press, 1965.

CHAPTER 3

Frequency Response Estimates based on Short Data Blocks
2. Closed-Loop Estimates

3.1 Introduction

In the previous chapter it was shown how the transients associated with short-term estimates lead to bias and variance in these estimates. The approach was restricted to open-loop systems; in this chapter it is extended to include closed-loop systems.

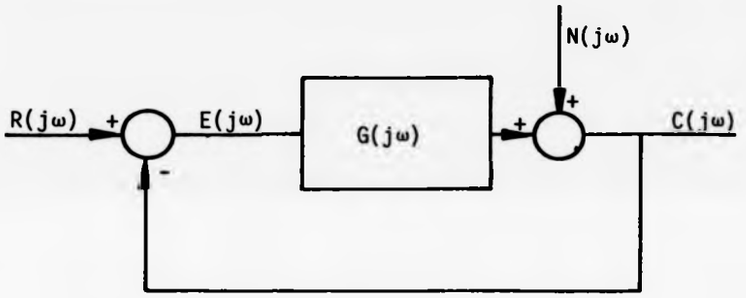


Fig. 3.1 General closed-loop system

The general system investigated is shown in Fig. 3.1. Here $R(j\omega)$, $E(j\omega)$ and $C(j\omega)$ represent the measured Fourier components at system input, error and output and $N(j\omega)$ represents an external noise source uncorrelated with the input.

It has been shown elsewhere [Wellstead (1)] that the frequency response estimator used in the open-loop system

$$\hat{G}(j\omega) = \frac{\sum\hat{\Phi}_{ec}}{\sum\hat{\Phi}_{ee}}$$

leads to bias when used in closed-loop systems when noise is present in the system. This is the case even when long records are used. To eliminate the bias under these conditions the estimator used under closed-loop conditions is

$$\hat{G}(j\omega) = \frac{\sum \hat{\Phi}_{rc}}{\sum \hat{\Phi}_{re}}$$

This is very different in form from the expression used for open-loop systems. One important difference is that the denominator is complex. This, as will be shown, gives an estimate that, unlike the open-loop estimate, has bias dependent upon the number of blocks L. Physically this can again be attributed to the transient terms. However, in the closed-loop situation, due to feedback, these are present both at the output and input of the forward path. The short-term correlation between these terms and the input gives additional bias which decreases as the number of blocks is increased.

The bias in the closed-loop estimate is obtained by consideration of the transfer functions C(s)/R(s) and E(s)/R(s). The effect of the transient terms on the measured frequency response associated with these functions can be obtained by the methods presented in chapter 2. These transients can be incorporated into the estimator used for $\hat{G}(j\omega)$ under closed-loop conditions. By consideration of the probability density function associated with the transient terms an expression is obtained for the expected value of $\hat{G}(j\omega)$.

An expression for the variance of $\hat{G}(j\omega)$ is not obtained. Although an expression for the probability distribution associated with $\hat{G}(j\omega)$ could be derived, this would be too involved to obtain an analytical expression for variance. However, confidence intervals are obtained for $\hat{G}(j\omega)$. This is done by transforming the circles defining the confidence limits for open-loop measurement to closed-loop conditions.

The general analysis presented is restricted to unity feedback systems. Although this is easily extended to the case

where there is a constant in the feedback path, difficulties arise in the case of a more general feedback transfer function. These are due to the effect of the energy storage associated with the feedback function producing transient terms in addition to those produced by the forward path function. However, a more limited method of obtaining the bias is presented which is applicable to the situation where the number of data blocks is sufficient to reduce the variance to negligible proportions.

Simulation results are presented that support the theoretical results derived.

3.2 The closed-loop estimator

Referring to Fig. 3.1, the estimator used is

$$\hat{G} = \frac{\sum \hat{\Phi}_{rc}}{\sum \hat{\Phi}_{re}}$$

$$= \frac{\sum R^*C}{\sum R^*E} \qquad \text{.....(3.1)}$$

where for convenience the dependence upon ω is dropped and the summation is taken over L blocks.

To obtain the bias in this estimator it is required to evaluate

$$E\{\hat{G}\} = E\left\{\frac{\sum R^*C}{\sum R^*E}\right\}$$

This expected value is dependent upon the number of blocks L and

$$E\left\{\frac{\sum R^*C}{\sum R^*E}\right\} \neq \frac{E\{\sum R^*C\}}{E\{\sum R^*E\}}$$

as for the corresponding estimator in the open-loop case.

The approach adopted is to consider the transfer functions C/R and E/R using the ideas considered in the previous chapter. For long records it is easily shown that

$$C = \frac{G}{1+G} R + \frac{N}{1+G}$$

$$E = \frac{1}{1+G} R - \frac{N}{1+G}$$

For records of length such that the transient terms are significant the total effect of the transient term at the system output is denoted by H and we may write

$$C = \frac{G}{1+G} R + \frac{N}{1+G} + H \quad \text{.....(3.2)}$$

H can be evaluated for this condition, but it should be noted that the closed-loop transfer function now determines the modes of the system and this is also the transfer function associated with the variance of the equivalent noise term.

Due to the unity feedback the transient term in the error is the negative of that at the output giving

$$E = \frac{1}{1+G} R - \frac{N}{1+G} - H \quad \text{.....(3.3)}$$

Eqn. (3.2) and (3.3) can be re-written

$$C = \frac{GR}{1+G} + \frac{N + (1+G)H}{1+G} \quad \text{.....(3.4)}$$

$$E = \frac{R}{1+G} - \frac{N + (1+G)H}{1+G} \quad \text{.....(3.5)}$$

Considering the numerator in the second term of these expressions

$$N + (1 + G)H$$

then using eqn. (2.7) this can be written

$$N + (1 + G)[k'R + N_e] \quad \text{.....(3.6)}$$

where k' is the closed-loop bias and N_e has been written as the equivalent uncorrelated noise source as defined in section 2.3.

The expression (3.6) can be written

$$N + (1 + G)N_e + (1 + G)k'R$$

As the external noise disturbance N is assumed to be uncorrelated with N_e the first two terms in the above expression can be combined into a single random variable which has zero mean and variance $|1 + G|^2\sigma_{Ne}^2 + \sigma_N^2$.

It is a significant result that the effect of uncorrelated noise can be incorporated into the analysis by modification of N_e. In view of this basic result, the disturbance N will, in the following discussion, be set to zero.

Hence, putting N = 0 into eqn. (3.4) and (3.5) and incorporating (3.6) gives

$$C = \frac{GR}{1 + G} + k'R + N_e \quad \text{.....(3.7)}$$

$$E = \frac{R}{1 + G} - k'R - N_e \quad \text{.....(3.8)}$$

From the estimator given as eqn. (3.1)

$$\hat{G} = \frac{\frac{G}{1 + G}\sum R^*R + k'\sum R^*R + \sum R^*N_e}{\frac{1}{1 + G}\sum R^*R - k'\sum R^*R - \sum R^*N_e}$$

$$= \frac{G + (k' + \sum R^*N_e/\sum R^*R)(1 + G)}{1 - (k' + \sum R^*N_e/\sum R^*R)(1 + G)}$$

This can be written

$$\hat{G} = G + \frac{(1 + G)^2(k' + \sum R^*N_e/\sum R^*R)}{1 - (k' + \sum R^*N_e/\sum R^*R)}$$

Following the method used in chapter 2 and writing

$$\frac{\sum R^*N_e}{\sum R^*R} = Me^{j\theta}$$

then the expression for $\hat{G}$ can be written

$$\hat{G} = G + (1 + G)\frac{k' + Me^{j\theta}}{(\frac{1}{1 + G} - k') - Me^{j\theta}} \quad \text{.....(3.9)}$$

3.3 Bias of the estimate in Closed-Loop

Taking expected values in eqn. (3.9) gives

$$\bar{G} = E\{\hat{G}\}$$

$$= G + (1 + G)\, E\left[\frac{k' + Me^{j\theta}}{(\frac{1}{1+G} - k') - Me^{j\theta}}\right] \quad \text{.....(3.10)}$$

The expectation can be written as (noting that M and θ are independent)

$$\int_0^\infty \int_0^{2\pi} \frac{k' + Me^{j\theta}}{(\frac{1}{1+G} - k') - Me^{j\theta}}\; p(\theta)\, p(M)\, d\theta\, dM \quad \text{.....(3.11)}$$

Taking first the inner integral with $p(\theta) = 1/2\pi$ and making the substitution $z = e^{j\theta}$, the integral becomes

$$\frac{1}{2\pi j} \int_{\substack{\text{unit}\\ \text{circle}}} \frac{1}{z}\; \frac{k' + Mz}{(\frac{1}{1+G} - k') - Mz}\, dz$$

This integral can be evaluated using residue theory, writing the integral as

$$-\frac{1}{z} \cdot \frac{(z + k'/M)}{(z - (\frac{1}{1+G} - k')/M)}$$

This has a pole at $z = 0$ with residue

$$\frac{k'}{\frac{1}{1+G} - k'}$$

and a pole at $z = (\frac{1}{1+G} - k')/M$ with residue

$$\frac{-1}{1 - k'(1+G)}$$

The integration is around the unit circle as shown in Fig. 3.2.

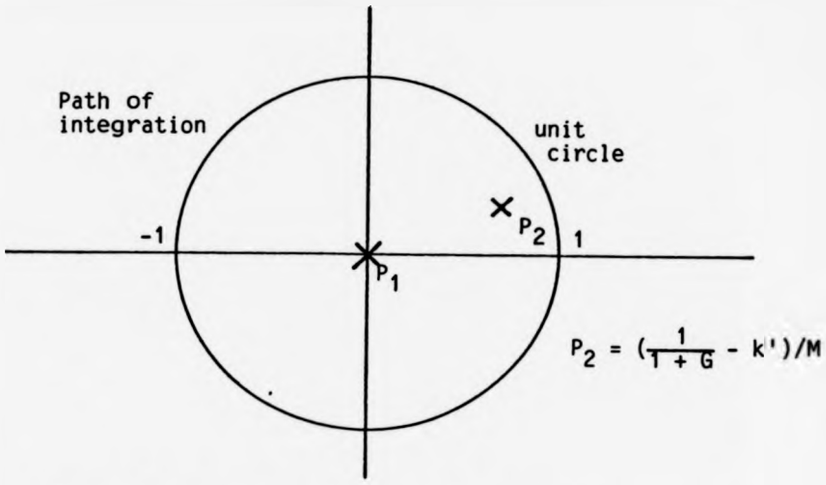


Fig. 3.2 Integration around unit circle

The value of the integral depends on whether P_2 is within the unit circle.

$$\text{For } \left| \frac{(\frac{1}{1+G} - k')}{M} \right| > | \text{ or } \quad M < \left| \frac{1}{1+G} - k' \right|$$

The pole is outside the unit circle and for $M > \left| \frac{1}{1+G} - k' \right|$ the pole is inside the unit circle. Using the relationship that the integral $= 2\pi j \sum$ residues gives

$$\frac{k'}{\frac{1}{1+G} - k'} \quad \text{if} \quad M < \left| \frac{1}{1+G} - k' \right|$$

and

$$\frac{k'}{\frac{1}{1+G} - k'} - \frac{1}{1 - k'(1+G)} = -1$$

if

$$M > \left| \frac{1}{1+G} - k' \right|$$

Returning to the complete integral as given in (3.11), the range of integration with respect to M can be divided 0 to $| 1/(1+G) - k' |$ and $| 1/(1+G) - k' |$ to ∞. This gives the integral

$$\left|\frac{1}{1+G} - k'\right| \int_0^{\frac{k'}{\frac{1}{1+G} - k'}} p(M)dM - \int_{\left|\frac{1}{1+G} - k'\right|}^{\infty} p(M)dM \quad \text{.....(3.12)}$$

Denoting by p_1 the integral

$$p_1 = \int_{\left|\frac{1}{1+G} - k'\right|}^{\infty} p(M)dM$$

which is the probability that M has value greater than $|1/(1 + G) - k'|$ the expression in (3.12) can be written

$$\frac{k'}{\frac{1}{1+G} - k'}(1 - p_1) - p_1$$

$$= \frac{k' - \frac{p_1}{1+G}}{\frac{1}{1+G} - k'}$$

The integral defining p_1 can be evaluated from the probability density function as given in chapter 2.

$$p_1 = \int_{\left|\frac{1}{1+G} - k'\right|}^{\infty} \frac{2LM(\sigma_R^2/\sigma_{Ne}^2)}{(1 + M\sigma_R^2/\sigma_{Ne}^2)^{L+1}} dM$$

$$= \frac{1}{(1 + \left|\frac{1}{1+G} - k'\right|^2 \sigma_R^2/\sigma_{Ne}^2)^L}$$

Returning to eqn. (3.10) gives an expression for $\bar{G}$

$$\bar{G} = G + (1 + G)\frac{k' - \frac{p_1}{1+G}}{\frac{1}{1+G} - k'} \quad \text{.....(3.13)}$$

Unlike the open-loop estimate, the bias in this estimate does depend upon the probability p_1 and hence on the number of blocks L. p_1 can vary from zero (when σ_R^2/σ_{Ne}^2 large or L large) to unity

(when σ_R^2/σ_{Ne}^2 small or L small). Eqn. (3.13) is the equation of a straight line joining the points

$$\bar{G} = G + (1 + G)\frac{k'}{\frac{1}{1+G} - k'}$$

$$= \frac{\frac{1}{1+G} + k'}{\frac{1}{1+G} - k'} \quad \text{when } p_1 = 0$$

to $\bar{G} = -1$ when $p_1 = 1$

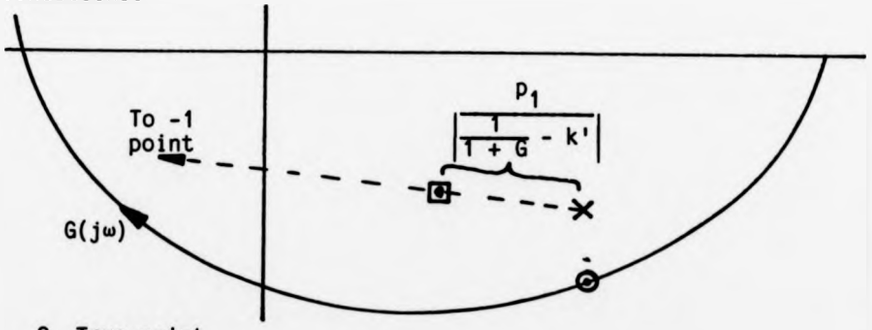
To locate $\bar{G}$ on this line it is easiest in practice to measure the distance for a given p_1 from the point for $p_1 = 0$. This is given by

$$\left| G + (1 + G)\left[\frac{k'}{\frac{1}{1+G} - k'}\right] - G - (1 + G)\left[\frac{k' - p_1}{\frac{1}{1+G} - k'}\right] \right|$$

$$= \frac{p_1}{\left|\frac{1}{1+G} - k'\right|}$$

This approach is similar to that used by Douce [2] to investigate bias in closed-loop systems due to uncorrelated noise.

Fig. 3.3 illustrates the bias produced under closed-loop conditions.



G True point

$\frac{G}{\frac{1}{1+G} - k'}$ Biased point for $p_1 = 0$

$\frac{G}{\frac{1}{1+G} - k'} + k' - p_1$ Biased point for p_1

Fig. 3.3 Bias in closed-loop configuration

The biased value of G when obtained using a closed-loop estimator and a large number of blocks has been shown to be

$$\bar{G} = \frac{\frac{1}{1+G} + k'}{\frac{1}{1+G} - k'}$$

But $G/(1+G)$ is the closed-loop transfer function and $1/(1+G)$ is the error transfer function and $\pm k'$ is the respective bias. Hence the ratio of the biased values of these functions gives $\bar{G}$.

An approximate relationship between the biased value of G obtained in open-loop and closed-loop conditions can be obtained as follows. The bias of the closed-loop system of transfer function

$$\frac{C}{R} = G_C(s) = \frac{G}{1+G}$$

is from chapter 2

$$\text{bias} = \frac{1}{T}\frac{dG_C(s)}{ds}$$

i.e. $$k' = \frac{1}{T}\frac{d}{ds}\left(\frac{G}{1+G}\right)$$

$$= \frac{1}{T}\frac{G'}{(1+G)^2} \quad \text{where } G' = \frac{dG}{ds}$$

Denoting the bias in the value of G from open-loop and closed-loop measurements as k and $k_{C/L}$ respectively

$$G + k_{C/L} = \frac{\frac{1}{1+G} + \frac{1}{T}\frac{G'}{(1+G)^2}}{\frac{1}{1+G} - \frac{1}{T}\frac{G'}{(1+G)^2}}$$

$$= \frac{G(1+G) + k}{(1+G) - k}$$

$$k_{C/L} = \frac{k}{1 - \frac{k}{1+G}}$$

Providing $|k| << |1+G|$ $\quad k_{C/L} \approx k$

3.4 Confidence Limits on Estimates

One method of obtaining the variance and confidence limits of the estimate $\hat{G}$ would be to proceed via the probability density function for $\hat{G}$ as in the open-loop case. This involves transforming the probability density function $p(M,\theta)$ of $Me^{j\theta}$ according to the transformation given by eqn. (3.14)

$$\hat{G} = G + (1 + G).\frac{1}{\frac{1}{(1 + G)(k' + Me^{j\theta})} - 1} \qquad \text{.....(3.14)}$$

Although this can be done, it is very involved and the resulting expression for the probability density function is too intractable to obtain its variance and to calculate confidence intervals. It is easier to restrict the investigation to confidence intervals and obtain these by transformation of the circles $Re^{j\theta}$ where R has been obtained from eqn. (2.16). This is similar to the approach adopted by Wellstead (1) and Davall (3) when considering the effect of noise on closed-loop frequency response estimates. In order to interpret the transformation eqn. (3.14) can be re-written

$$\hat{G} = -1 + \frac{1}{(\frac{1}{1 + G} - k') - Re^{j\theta}}$$

This equation can be interpreted as a number of elementary transformations on the circle $Re^{j\theta}$

1. Movement of the centre to point $(\frac{1}{1 + G} - k')$
2. Inversion
3. Translation by -1.

The resultant circles are shown in Fig. 3.4 and the following points can be made in relation to them

(i) The centres of the circles lie on a straight line passing through the points

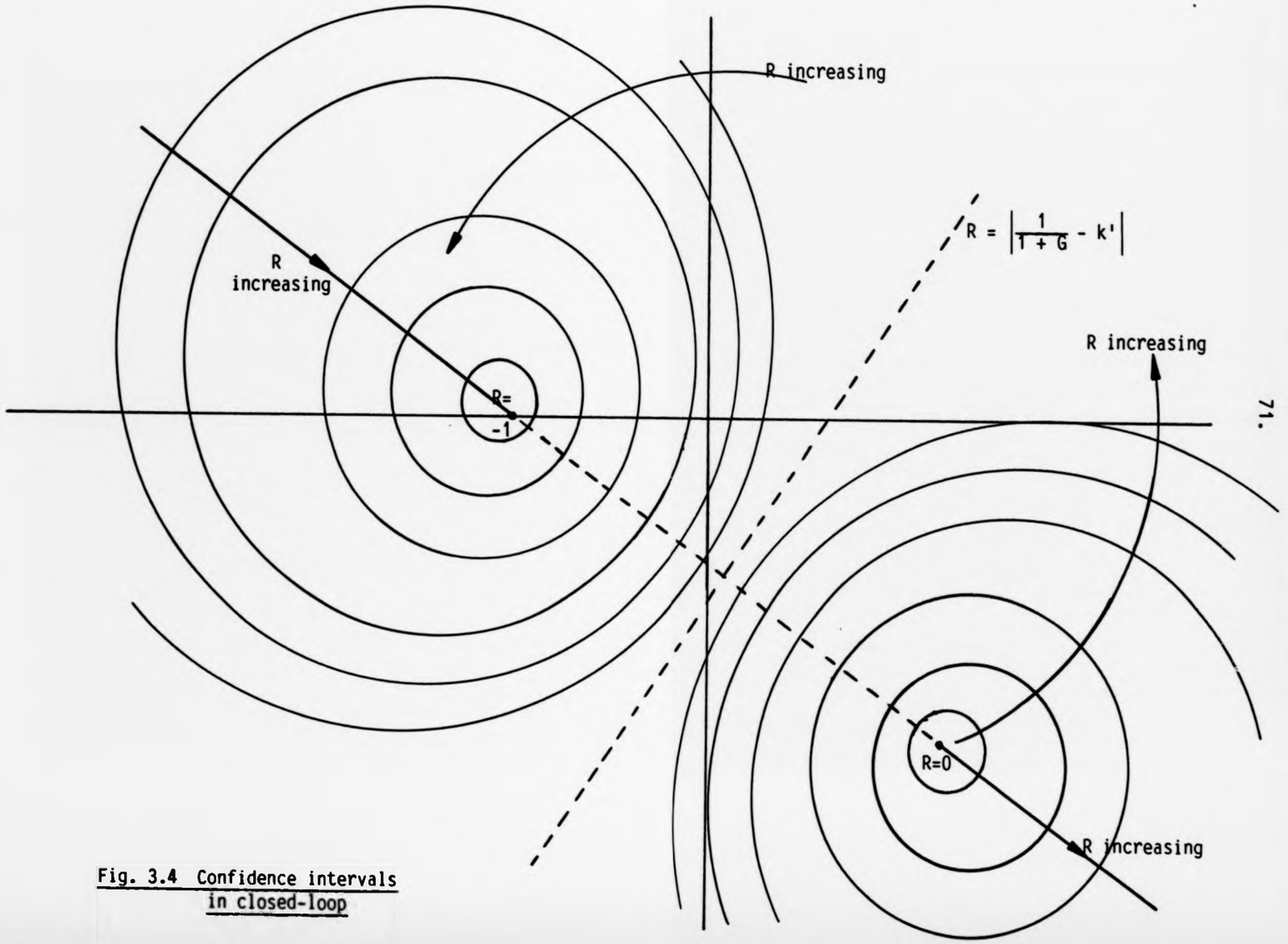


Fig. 3.4 Confidence intervals in closed-loop

$$\frac{\frac{G}{1+G} + k'}{\frac{1}{1+G} - k'} \qquad \text{when } R = 0$$

and $\quad -1 + j0 \qquad$ when $R = \infty$

(ii) The radii of the circles R' are given by

$$R' = \frac{R}{\left|\frac{1}{1+G} - k'\right|^2 - R^2} \qquad \text{.....(3.15)}$$

(iii) The distance of the centres from the -1 point, D, are given by

$$D = \frac{\left|\frac{1}{1+G} - k'\right|}{\left|\frac{1}{1+G} - k'\right|^2 - R^2} \qquad \text{.....(3.16)}$$

A positive sign in the expression for D is interpreted as the distance being measured from the -1 point towards the point for R = 0.

(iv) If $R < \left|\frac{1}{1+G} - k'\right|$ the circles map interior to interior. If $R > \left|\frac{1}{1+G} - k'\right|$ they map interior to exterior.

It should be noted that the centres of these circles do not coincide with the biased values $\bar{G}$ except for the case R = 0 (although $\bar{G}$ and the circle centres do lie on the same straight line). As R is increased the points for $\bar{G}$ move towards the -1 point, whereas the centres of these circles move away from the -1 point. Although this statement appears paradoxical, it is due to the fact that the probability, for a given R, is not uniform with θ and the expected value is not the centre of the circle.

The formulae for R and D, eqns. (3.15) and (3.16) do not apply at $\omega = 0$ due to the differing probability density function p(M) at this frequency.

3.5 A Specific Example

For comparison purposes the same example was chosen as was investigated in chapter 2. However, now the system was used with unity feedback as shown in Fig. 3.5.

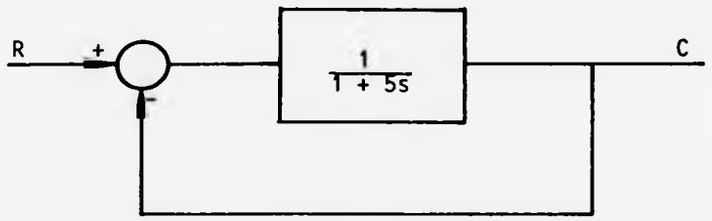


Fig. 3.5 First order system in closed-loop configuration

Again the observation time, T, was chosen as 25.6 seconds. In order to calculate the bias in the estimate and to obtain confidence limits it is required to obtain the mean square value of the equivalent noise source σ_{Ne}^2. This is obtained from consideration of the closed-loop transfer function.

$$\frac{C}{R}(s) = \frac{0.5}{1 + 2.5s}$$

The bias in this estimate is

$$k' = \frac{1}{T} \frac{d(c/R)}{ds}\Big|_{s=j\omega}$$

$$= - \frac{0.0488}{(1 + j2.5\omega)^2}$$

Hence the biased closed-loop estimate is

$$\bar{G}_{c/L} = \frac{0.5}{(1 + j2.5\omega)} - \frac{0.0488}{(1 + j2.5\omega)^2}$$

The ratio σ_c^2/σ_R^2 can be obtained as in the example of chapter 2 (note that $T_1 = 2.5$ and there is a factor $(0.5)^2$ due to the non-unity d.c. gain).

$$\frac{\sigma_c^2}{\sigma_R^2} = (0.5)^2 \left[\frac{1}{1 + \omega^2 T_1^2} - \frac{T_1}{T} \frac{(1 - \omega^2 T_1^2)(1 - e^{-T/T_1})}{(1 + \omega^2 T_1^2)^2} \right]$$

The ratio σ_{Ne}^2/σ_R^2 is given by

$$\frac{\sigma_{Ne}^2}{\sigma_R^2} = \frac{\sigma_C^2}{\sigma_R^2} - \left| \bar{G}_{C/L} \right|^2$$

The biased estimate for the forward path transfer function $\bar{G}$ is given by

$$\bar{G} = G + (1 + G) \frac{k' - \frac{p_1}{1 + G}}{\frac{1}{1 + G} - k'}$$

where

$$p_1 = \frac{1}{(1 + \left| \frac{1}{1 + G} - k' \right|^2 \frac{\sigma_R^2}{\sigma_{Ne}^2})^L}$$

Table 3.1 gives values of $\bar{G}$ against frequency. The results are for one block, L = 1, and two blocks, L = 2, and theoretical results are compared with those obtained by simulation. Results have not been given for $\omega = 0$ as the formulae are not applicable at this frequency. The simulation was taken over 5000 experiments and the agreement between theoretical results and those obtained by simulation is good, especially for the case where L = 2 when the confidence limit circles have smaller radii. It should be noted for L = 2 the difference between these results and those obtained open-loop (as given in table 2.1) is small.

Table 3.1

Harmonic Number m	Angular Frequency $m\omega_0$	$\bar{G}(j\omega)$ for L = 1		$\bar{G}(j\omega)$ for L = 2	
		Theoretical	Simulation	Theoretical	Simulation
1	0.245	0.3631-j0.4007	0.3672-j0.4044	0.4089-j0.4141	0.4066-j0.4111
2	0.491	0.1463-j0.3253	0.1368-j0.3181	0.1615-j0.3296	0.1625-j0.3296
3	0.763	0.0730-j0.2445	0.0763-j0.2469	0.0801-j0.2461	0.0800-j0.2463
4	0.982	0.0429-j0.1918	0.0425-j01946	0.0469-j0.1926	0.0464-j0.1928
5	1.227	0.0280-j0.1568	0.0275-j0.1600	0.0306-j0.1571	0.0308-j0.1573
6	1.473	0.0197-j0.1322	0.0172-j0.1307	0.0215-j0.1324	0.0213-j0.1321
7	1.718	0.0145-j0.1141	0.0159-j0.1140	0.0159-j0.1142	0.0159-j0.1142
8	1.963	0.0112-j0.1003	0.0120-j0.1010	0.0122-j0.1004	0.0124-j0.1006

Confidence intervals for estimates of $\bar{G}$ under closed-loop conditions can be obtained from eqns. (2.16) and (3.15). However, in (2.16) σ_n^2/σ_x^2 becomes σ_{NE}^2/σ_R^2 and this can now be written

$$R = \sqrt{\left(\frac{1}{(1-p)^{1/L}} - 1\right)\frac{\sigma_{Ne}^2}{\sigma_R^2}}$$

Here R is the radius of a circle giving a confidence interval of 100p% for the closed-loop estimate C/R. This radius is transformed into the radius of the circle R' giving the corresponding confidence interval on $\bar{G}$ by eqn. (3.15).

$$R' = \frac{R}{\left|\frac{1}{1+G} - k'\right|^2 - R^2}$$

The distance of the centre of this circle from the -1 point is given by

$$D = \frac{\left|\frac{1}{1+G} - k'\right|}{\left|\frac{1}{1+G} - k'\right|^2 - R^2}$$

Table 3.2 gives radii of 90% confidence interval circles and the distance from the point $\bar{G}$ corresponding to $L = \infty$. Results are given for L = 2 and L = 4 and theoretical results are compared with those obtained by simulation.

As can be seen, the distance of the centres of the confidence interval circles from the point $\bar{G}$, corresponding to $L = \infty$ is small and little error would result if the centres were assumed to lie at that point.

The radii of the circles can be compared to the corresponding open-loop circles given in table 2.4. Because the input is identical to the open-loop case the output power is reduced due to feedback in this example; this reduces the radii of the circles. The transformation from R to R' however increases the

radius. Hence, the confidence intervals are greater than the corresponding open-loop intervals at low frequencies, but smaller at high frequencies.

The results obtained from simulation gave very good agreement with theory; these were obtained for 4000 experiments.

Table 3.2

Harmonic Number m	Frequency $m\omega_0$	Theoretical Results				Simulation. % of points inside circle	
		L = 2		L = 4			
		Distance	Radii	Distance	Radii	L = 2	L = 4
1	0.245	0.1195	0.4358	0.0409	0.2485	92.07	90.45
2	0.491	0.0362	0.2123	0.0128	0.1250	90.70	89.57
3	0.763	0.0161	0.1345	0.0057	0.0800	89.65	89.70
4	0.982	0.0089	0.0981	0.0032	0.0586	90.03	89.23
5	1.227	0.0057	0.0773	0.0020	0.0462	89.65	89.58
6	1.473	0.0039	0.0638	0.0014	0.0382	90.20	89.25
7	1.718	0.0029	0.0544	0.0010	0.0326	89.98	90.28
8	1.963	0.0022	0.0474	0.0008	0.0284	89.38	89.38
9	2.209	0.0017	0.0420	0.0006	0.0252	89.38	89.28

References - Chapter 3

1. Wellstead, P.E., "Aspects of real-time digital spectral analysis", PhD Thesis, University of Warwick, 1970.

2. Douce, J.L., "Bias of frequency-response estimates in closed-loop systems", Proc. IEE, Vol.127, Pt.D, No.4, pp.149-152, July 1980.

3. Davall, P.W., "Applications of statistics in the spectral analysis of time-varying systems", PhD Thesis, University of Warwick, 1975.

CHAPTER 4

A New Approach to Parametric Identification

4.1 Introduction

The identification techniques described in chapters 2 and 3 have been non-parametric methods. In such methods no assumptions need be made regarding the form of the system model. However, as outlined in chapter 1, parametric methods have proved very successful when applied to auto-spectra estimation. In the parametric method a model is assumed for the process generating the spectra and the identification technique obtains estimates for the parameter values in this model. The main application of this approach to frequency response estimation has been for the off-line fitting of measured frequency data to a rational model.

This chapter investigates the use of parametric methods to short term frequency response estimation. The novelty in the approach lies in the fact that additional terms are included in the model to account for the transient terms discussed in previous chapters. Although this approach is original when applied to the frequency domain, similar problems have been investigated in the time domain. These have occurred when records from different tests have been concatenated and recursive estimation techniques have been applied to the complete record [Gawthrop (1)] or when one record has been passed through a recursive estimator a number of times [Solbrand et al (2)]. It is shown that in the noise free case exact estimates can be obtained from records that are significantly shorter than the system time. Addition of noise causes bias in the estimate, but this can be removed in the open-loop case by using recursive methods similar to those employed in the time domain.

The method meets problems when estimates are updated by using data from successive blocks. This is because the transient terms, and the parameters introduced to account for them, change in successive blocks. Two methods of combining data from successive blocks are discussed.

Simulation results are presented to verify the theories developed.

4.2 Open-Loop Estimation

The starting point for the parametric method is based on eqn. (2.3)

$$Y(j\omega) = G(j\omega)X(j\omega) + H(j\omega)$$

where $Y(j\omega)$, $X(j\omega)$ and $H(j\omega)$ are short term Fourier Transforms of the output signal, input signal and transient term respectively. As shown in section 2.2.1, this equation is exact even though the transforms are taken over a finite time.

In the open-loop configuration the addition of noise is represented by the addition of an extra term in the above equation.

$$Y(j\omega) = G(j\omega)X(j\omega) + H(j\omega) + N(j\omega) \quad \text{.....(4.1)}$$

It should be noted that the noise signal prior to $t = 0$ will also produce transient terms and $N(j\omega)$ is not an accurate representation of the effects of the noise. However, as the noise term is considered unobservable, this does not affect the following argument.

In order to apply parametric methods, the functions $G(j\omega)$ and $H(j\omega)$ have to be written in parametric form.

$$G(j\omega) = \frac{B(j\omega)}{A(j\omega)}$$

where

$$B(j\omega) = b_0 + b_1 j\omega + b_2(j\omega)^2 + \ldots\ldots\ldots b_m(j\omega)^m$$

and $$A(j\omega) = a_0 + a_1 j\omega + a_2(j\omega)^2 + \ldots\ldots\ldots a_n(j\omega)^n$$

The analysis is more convenient if the coefficient a_0 is made equal to unity and this can obviously be done by dividing both numerator and denominator of $G(j\omega)$ by a_0. This cannot be done if $a_0 = 0$ (this case represents integration in the $G(j\omega)$ function) and some modification of the approach will be required.

As shown in 2.2.1, the component $H(j\omega)$ can be written as

$$H(j\omega) = \frac{H'(j\omega)}{A(j\omega)}$$

$$= \frac{h_0 + h_1 j\omega + h_2(j\omega)^2 + \ldots\ldots\ldots h_{n-1}(j\omega)^{n-1}}{A(j\omega)}$$

Substituting these expressions for $G(j\omega)$ and $H(j\omega)$ into eqn. (4.1) gives

$$Y(j\omega_i) = \frac{B(j\omega_i)}{A(j\omega_i)} X(j\omega_i) + \frac{H'(j\omega_i)}{A(j\omega_i)} + N(j\omega_i) \ldots\ldots\ldots\ldots (4.2)$$

where ω_i has been written for ω to emphasise that only harmonic multiples of the fundamental frequency are considered.

The problem of parametric identification can be stated - given the data $X(j\omega_i)$ and $Y(j\omega_i)$ at each frequency point ω_i, what values of the parameters $a_1 \ldots\ldots a_n$, $b_0 \ldots\ldots b_m$, $h_0 \ldots\ldots h_{n-1}$ will give a "best fit" to this data. This is a problem in regression analysis [Draper & Smith (3)] and the problem has been widely investigated for identification in the time domain [Eykhoff (4)]. However, the data here is complex and, although some work has been done [e.g. Levy (5), Lawrence & Rogers (6)] on the fitting of frequency response functions to complex data, this

has been intended for off-line applications and the effects of finite observation time have not been considered.

Defining a complex error term at each frequency ω_i as $E(j\omega_i)$ then

$$E(j\omega_i) = Y(j\omega_i) - \frac{B(j\omega_i)}{A(j\omega_i)} X(j\omega_i) - \frac{H'(j\omega_i)}{A(j\omega_i)}$$

$$\frac{Y(j\omega_i)A(j\omega_i) - B(j\omega_i)X(j\omega_i) - H'(j\omega_i)}{A(j\omega_i)} \quad \ldots\ldots\ldots(4.3)$$

From eqns. (4.2) and (4.3) it can be seen that

$$E(j\omega_i) = N(j\omega_i) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4.4)$$

It should be noted that this equation is only true if the model is correct. Any inaccuracies in the assumed form, or in the order of the model, will introduce additional errors.

The best fit, in a least squares sense, is obtained by minimising the function

$$S = \sum_{i=1}^{R} |E(j\omega_i)|^2$$

with respect to the parameters a, b, h (R denotes the number of frequency points). Minimisation of this function leads to a set of non-linear equations [Bard (7)] and methods of solution will be considered later.

If eqn. (4.2) is multiplied through by $A(j\omega_i)$ an alternative error criterion can be formulated. Redefining

$$E'(j\omega_i) = Y(j\omega_i)A(j\omega_i) - B(j\omega_i)X(j\omega_i) - H'(j\omega_i) \quad \ldots\ldots\ldots(4.5)$$

it can be seen that

$$E'(j\omega_i) = A(j\omega_i)N(j\omega_i) = A(j\omega_i)E(j\omega_i)$$

If the noise term $N(j\omega_i)$ is zero at all frequencies then it can be seen that $E(j\omega_i) = E'(j\omega_i) = 0$. The sum of squares

criterion using $E'(j\omega_i)$ becomes

$$\sum_{i=1}^{R} |E'(j\omega_i)|^2 = \sum_{i=1}^{R} |A(j\omega_i)|^2 |E(j\omega_i)|^2$$

The advantage of the modified function is that its minimisation leads to equations that are linear, i.e. a linear regression problem. This will give correct estimates if there is no noise present; however, in the presence of noise it weights the least squares criterion with a weighting function $|A(j\omega_i)|^2$. In general this function will increase with frequency and undue weighting will be given to higher frequency points; a good fit will not be obtained at low frequencies. If, however, the function A(s) has zeros near to the imaginary axis $|A(j\omega_i)|^2$ could vary widely throughout the frequency range and large errors could be introduced.

The linear solution does, however, provide a starting point for algorithms to solve the non-linear regression problem, as will be shown later.

4.3 Linear Regression in Open-Loop

The problem is most conveniently expressed in vector form. Re-stating the definitions for $A(j\omega_i)$, $B(j\omega_i)$ and $H(j\omega_i)$ as

$$A_i = 1 + a_1 j\omega_i + a_2 (j\omega_i)^2 + \ldots\ldots\ldots a_n (j\omega_i)^n$$

$$B_i = b_o + b_1 j\omega_i + b_2 (j\omega_i)^2 + \ldots\ldots\ldots b_n (j\omega_i)^m$$

$$H'_i = h_o + h_1 j\omega_i + h_2 (j\omega_i)^2 + \ldots\ldots\ldots h_{n-1} (j\omega_i)^{n-1}$$

and define the vectors

$$\underset{\sim}{p}_i = [1, j\omega_i, (j\omega_i)^2 \ldots\ldots\ldots\ldots\ldots (j\omega_i)^n]^T$$

$$\underline{q}_i = [j\omega_i, (j\omega_i)^2 \ldots\ldots\ldots\ldots (j\omega_i)^m]^T$$

$$\underline{r}_i = [1, j\omega_i, (j\omega_i)^2 \ldots\ldots\ldots\ldots (j\omega_i)^{n-1}]^T$$

$$\underline{a} = [a_1, a_2 \ldots\ldots\ldots\ldots\ldots\ldots a_n]^T$$

$$\underline{b} = [b_0, b_1 \ldots\ldots\ldots\ldots\ldots\ldots b_m]^T$$

$$\underline{h} = [h_0, h_1 \ldots\ldots\ldots\ldots\ldots\ldots h_{m-1}]^T$$

where $[\ ]^T$ denotes the transpose.

Eqn. (4.5) can now be written in vector form and for R frequency points gives the set of equations.

$$\begin{matrix} E_1' \\ \vdots \\ E_R' \end{matrix} = \begin{matrix} Y_1 \\ \vdots \\ Y_R \end{matrix} [1 + \underline{a}^T\underline{q}_1] - \underline{b}^T\underline{p}_1 X_1 - \underline{h}^T\underline{r}_1$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$E_R' = Y_R[1 + \underline{a}^T\underline{q}_R] - \underline{b}^T\underline{p}_R X_r - \underline{h}^T\underline{r}_R$$

These equations can now be written

$$\underline{E}' = \underline{Y} - \underline{u}\,\underline{\alpha} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.6)$$

where $\underline{E}' = [E_1', E_2' \ldots\ldots\ldots\ldots\ldots E_R']$

$$\underline{Y} = [Y_1, Y_2 \ldots\ldots\ldots\ldots\ldots Y_R]$$

$$\underline{u} \qquad \begin{bmatrix} \underline{p}_1^T X_1 , & -\underline{q}_1^T Y_1 , & \underline{r}_1^T \\ \vdots & \vdots & \vdots \\ \underline{p}_R^T X_R , & -\underline{q}_R^T Y_R , & \underline{r}_R^T \end{bmatrix}$$

$$\underline{\alpha} = [b_0, b_1 \ldots b_m, a_1, a_2 \ldots a_n, h_1, h_2 \ldots h_{n-1}]^T$$

The function to be minimised can now be written

$$\sum_{i=1}^{R} |E_i'|^2 = \underline{E}'^{*T}\underline{E}'$$

$$= (\underline{Y} - \underline{u}\underline{\alpha})^{*T}(\underline{Y} - \underline{u}\underline{\alpha})$$

$$= \underline{Y}^{*T}\underline{Y} - \underline{\alpha}^T\underline{u}^{*T}\underline{Y} - \underline{Y}^{*T}\underline{u}\underline{\alpha} + \underline{\alpha}^T\underline{u}^{*T}\underline{u}\underline{\alpha}$$

The conditions for a minimum are derived by setting the derivatives of the above expression with respect to all the parameters to zero. This is most conveniently carried out by differentiation with respect to the vector $\underline{\alpha}$. This gives the expression

$$- \underline{u}^{*T}\underline{Y} - \underline{u}^{T}\underline{Y}^{*} + \underline{u}^{*T}\underline{u}\underline{\alpha} + \underline{u}^{T}\underline{u}^{*}\underline{\alpha}$$

and if this equals zero for $\underline{\alpha} = \hat{\underline{\alpha}}$, then $\hat{\underline{\alpha}}$ is the best estimate, in a minimum least squares sense, of $\underline{\alpha}$.

$$\hat{\underline{\alpha}} = (\underline{u}^{*T}\underline{u} + \underline{u}^{T}\underline{u}^{*})^{-1} (\underline{u}^{*T}\underline{Y} + \underline{u}^{T}\underline{Y}^{*}) \quad \ldots\ldots\ldots\ldots\ldots (4.7)$$

It should be noted that, although the matrix $\underline{u}$ is complex and includes terms dependent upon frequency in a non-linear manner, the regression is linear in the parameters $\underline{\alpha}$. The result obtained is of the same form as that obtained for least squares methods in the time domain and standard expressions for bias and variance can be applied [Eykhoff (4)]. In particular it means that the estimates will be biased due to the non-white nature of the term $A(j\omega)N(j\omega)$ introduced to produce a linear regression problem.

In order to give less weighting to frequencies where the input component is small, the error can be weighted by the magnitude of the input at each frequency (weighted least squares regression).

The expression to be minimised now becomes

$$\underline{E}'^{*T}\,\underline{W}\,\underline{E}'$$

where $\underline{W}$ is a weighting matrix given by

$$\underline{W} = \begin{bmatrix} |X_1|^2 & & & \\ & |X_2|^2 & & \\ & & \ddots & \\ & & & |X_R|^2 \end{bmatrix}$$

$$= \begin{bmatrix} X_1 & & \\ & X_2 & \\ & & \ddots \\ & & & X_R \end{bmatrix} \begin{bmatrix} X_1^* & & \\ & X_2^* & \\ & & \ddots \\ & & & X_R^* \end{bmatrix}$$

Substituting $\underline{E}'$ from eqn. (4.6) gives sum of squares error as

$$[\underline{Y} - \underline{\underline{u}}\underline{\alpha}]^{*T} \begin{bmatrix} X_1 & & \\ & X_2 & \\ & & \ddots \\ & & & X_R \end{bmatrix} \begin{bmatrix} X_1^* & & \\ & X_2^* & \\ & & \ddots \\ & & & X_R^* \end{bmatrix} [\underline{Y} - \underline{\underline{u}}\underline{\alpha}]$$

By performing the multiplications of the two outer terms this can be written

$$[\hat{\underline{\Phi}}_{xy} - \underline{\underline{v}}\underline{\alpha}]^{*T} [\hat{\underline{\Phi}}_{xy} - \underline{\underline{v}}\underline{\alpha}]$$

where

$$\hat{\underline{\Phi}}_{xy} = [X_1^*Y_1, X_2^*Y_2, \ldots\ldots\ldots X_R^*Y_R]^T$$

and

$$\underline{\underline{V}} = \begin{bmatrix} \underline{p}_1^T\hat{\Phi}_{xx}(j\omega_1), & -\underline{q}^T\hat{\Phi}_{xy}(j\omega_1), & \underline{r}_1^T \\ \vdots & \vdots & \vdots \\ \underline{p}_R^T\hat{\Phi}_{xx}(j\omega_R), & -\underline{q}_R^T\hat{\Phi}_{xy}(j\omega_R), & \underline{r}_R^T \end{bmatrix}$$

It should be noted that $\underline{\alpha}$ is still defined as before

$$\underline{\alpha} = [b_0, b_1 \ldots b_m, a_1, a_2 \ldots a_n, h_1, h_2 \ldots h_{n-1}]^T$$

however, the parameters $h_1 \ldots h_{n-1}$ are now the parameters in the polynomial representing the cross-spectra between X and H.

$$\hat{\Phi}_{xH}(j\omega) = X^*(j\omega)H(j\omega) = h_0 + h_1(j\omega) + \ldots. h_{n-1}(j\omega)^{n-1}$$

Hence, parameter identification can be performed using auto and cross-spectra. As in the time domain, this will reduce variance at the expense of bias. This method will be reconsidered in section 4.6 when the problem of combining data blocks is considered.

4.4 Closed-Loop Identification

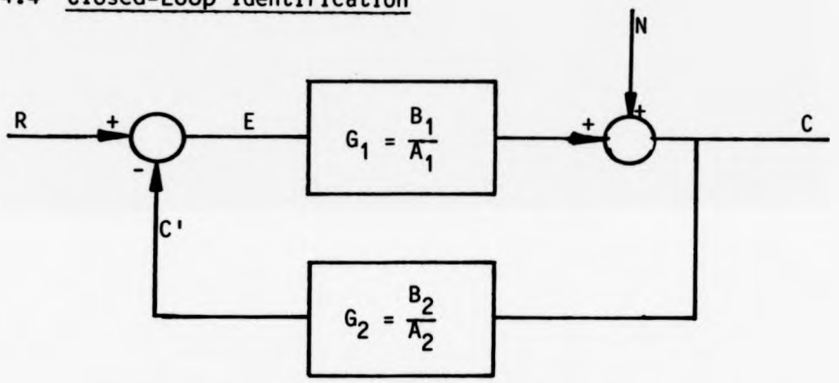


Fig. 4.1 Closed-loop configuration

With reference to Fig. 4.1, the forward block and the feedback block can be treated separately. Each gives an output which is its input multiplied by its frequency response function plus a term due to the input prior to t = 0. This gives the equations

$$C = E.\frac{B_1}{A_1} + \frac{H_1}{A_1} + N \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4.8)$$

$$C' = C\frac{B_2}{A_2} + \frac{H_2}{A_2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4.9)$$

where the dependence on frequency has been dropped for convenience. Eqn. (4.8) is the equation that is required to obtain the parameters

of the forward path. However, it is instructive to solve these equations to obtain the relationship between C and R. Using the equation

$$E = R - C' \quad \text{.....(4.10)}$$

together with eqns. (4.8) and (4.9) gives

$$C = \frac{B_1A_2}{A_1A_2 + B_1B_2} R + \frac{A_2H_1 - H_2B_1}{A_1A_2 + B_1B_2} + N\frac{A_1A_2}{A_1A_2 + B_1B_2}$$

Writing $A_2H_1 - H_2B_1$ as H, then

$$C = \frac{G_1}{1 + G_1G_2} R + \frac{H}{A_1A_2 + B_1B_2} + \frac{N}{1 + G_1G_2}$$

In this expression $G_1/(1 + G_1G_2)$ is the closed-loop transfer function and $H/(A_1A_2 + B_1B_2)$ is the transient associated with the input R and the closed-loop transfer function.

The term at the output due to the input prior to t = 0 can be viewed in two ways.

(i) A term

$$\frac{H}{A_1A_2 + B_1B_2}$$

The system is regarded as a closed-loop system with input R. This term has modes that are associated with the closed-loop transfer function of the system with zero input for t > 0.

(ii) A term

$$\frac{H_1}{A_1}$$

This term is associated with the forward path transfer function G_1 and its modes are associated with the poles of this transfer function. Due to the unity feedback, the input to the forward path is the output transient obtained in (i) (with a sign reversal).

Equations (4.8), (4.9), (4.10) can also be solved to give the signal E. This is given by

$$E = \frac{1}{1 + G_1G_2}R - \frac{(A_1H_2 + B_2H_1)}{A_1A_2 + B_1B_2} - \frac{G_2N}{1 + G_1G_2}$$

Again the contribution to this expression due to the signal prior to t = 0 has modes determined by the closed-loop poles of the system.

For the special case of unity feedback $A_2 = B_2 = 1$ and $H_2 = 0$. As expected, these terms are then identical (except for the sign reversal) at both C and E.

Although eqn. (4.8) can be used to obtain the parameters of the forward path, the transient terms at the input and output of the forward path transfer function are correlated. This can cause additional bias in the parameter estimates.

4.5 Non-Linear Parameter Estimation Methods

These methods are based on the minimisation of the function given in section 4.2. This can be written

$$S(\alpha) = \sum_{i=1}^{R} |E_i|^2$$

$$= \sum_{i=1}^{R} \left| \frac{Y_iA_i - B_iX_i - H_i'}{A_i} \right|^2 \qquad \text{.....(4.11)}$$

Differentiation of this function with respect to the required parameters would lead to a non linear set of equations and no analytical solution is in general possible. The general problem of non-linear parameter estimation is of current research interest [Bard (7)] and many approaches can be adopted in order to obtain a solution. Most of these involve some form of iterative technique to obtain values of the parameters $\underline{\alpha}$ that will minimise the function

given in (4.11). However, it is recognised that there is no universal "best" method and a method giving satisfactory results with one problem may not do so with a different problem.

Two methods for the minimisation of (4.11) have been investigated; one of these is a general method applicable to non-linear least squares regression problems; the other method is dependent upon the particular nature of the problem.

4.5.1 Modified Newton Method

This method is based on the linearisation of the function $S(\underline{\alpha})$, initially around an initial parameter estimate $\underline{\alpha}_o$ A linear regression can now be applied to find the optimum change in parameters $\Delta\underline{\alpha}$. This in turn determines a new estimate $\underline{\alpha}_o + \Delta\underline{\alpha}$ and the process can be repeated until the solution converges. In practice this basic algorithm has been extensively modified to improve robustness and provide more rapid convergence [Gill and Murray (8)].

For this particular problem the form of $S(\underline{\alpha})$ enables an improvement to be made to the algorithm. Although differentiation of $S(\underline{\alpha})$ with respect to the $\underline{a}$ parameter produces non-linear equations, the equations are linear when differentiation is with respect to the $\underline{b}$ and $\underline{h}$ parameters. This property can be utilised as follows. Consider the initial estimate $\underline{\alpha}_o$ required to initiate the algorithm and suppose the initial choice $\underline{a}_o$ of the $\underline{a}$ parameters has been made. Rather than choose the accompanying $\underline{b}_o$ and $\underline{h}_o$ parameters arbitrarily, it would be better to choose parameters that give a minimum value of $S(\underline{\alpha})$ conditional to the fixed $\underline{a}_o$. This must give a value of $S(\underline{\alpha}_o)$ closer to the absolute minimum than any other choice of $\underline{b}_o$ and $\underline{h}_o$. However, these values of $\underline{b}_o$ and $\underline{h}_o$ can

be obtained by linear regression. Although this has been stated for the initial estimate, it is true for any point in the iteration. The modified algorithm is shown in Fig. 4.2.

The advantages of this method are,

(i) initial estimates need only be made for the $\underset{\sim}{a}$ parameters.

(ii) as the number of parameters in the iterative routine have been reduced, it produces a more robust algorithm with a faster rate of convergence.

A disadvantage is that the partial derivatives $\partial S(\underset{\sim}{\alpha})/\partial \underset{\sim}{\alpha}$ are now not available and these must be approximated by small differences.

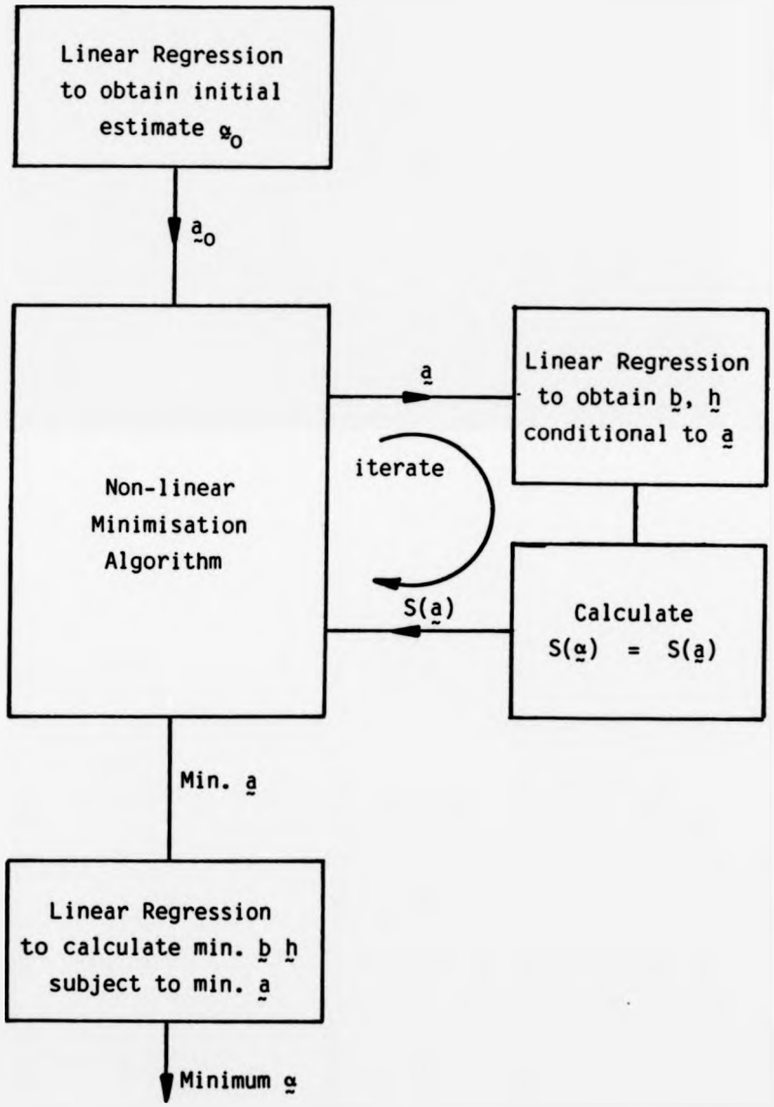


Fig. 4.2 Algorithm combining linear and non-linear regression.

4.5.2 Generalised Least Squares Method

This method is similar to the Generalised Least Squares method used for identification in the time domain. Re-writing eqn. (4.3) the function to be minimised as

$$S(\underset{\sim}{\alpha}) = \sum_{i=1}^{R} \left| \frac{Y_i A_i - B_i X_i - H_i'}{A_i} \right|^2$$

In the linear regression (section 4.3) the denominator $|A_i|$ is taken as a weighting factor leading to biased estimates. However, a weighted linear regression can be performed with $|A_i|$ held at any constant value. This suggests an iterative technique where the denominator is held constant, at a given step, with a value determined by the $\underset{\sim}{a}$ parameters calculated at the previous step. The algorithm can be illustrated by the flow chart shown in Fig. 4.3. Although the convergence of this algorithm cannot be guaranteed, it can be seen should the algorithm converge $A_k = A_{k-1}$ and the true minimum is obtained.

In this algorithm all the data $X(j\omega)$ and $Y(j\omega)$ is used at each iteration and each iteration requires the solution of eqn. (4.7) which may be ill-conditioned. In the least squares identification method in the time domain recursive methods of solution are used to avoid the solution of the linear equations [Hsia (9)]. A similar method can be applied here, but modification is required to deal with the data in complex form.

The k^{th} frequency point introduces a new row $\underset{\sim}{u}^T(k)$ to the $\underset{\sim}{u}$ matrix.

$$\underset{\sim}{u}^T(k) = [\underset{\sim}{p}_k^T X_k, \; - \underset{\sim}{q}_k^T Y_k, \; \underset{\sim}{r}_k^T \;]$$

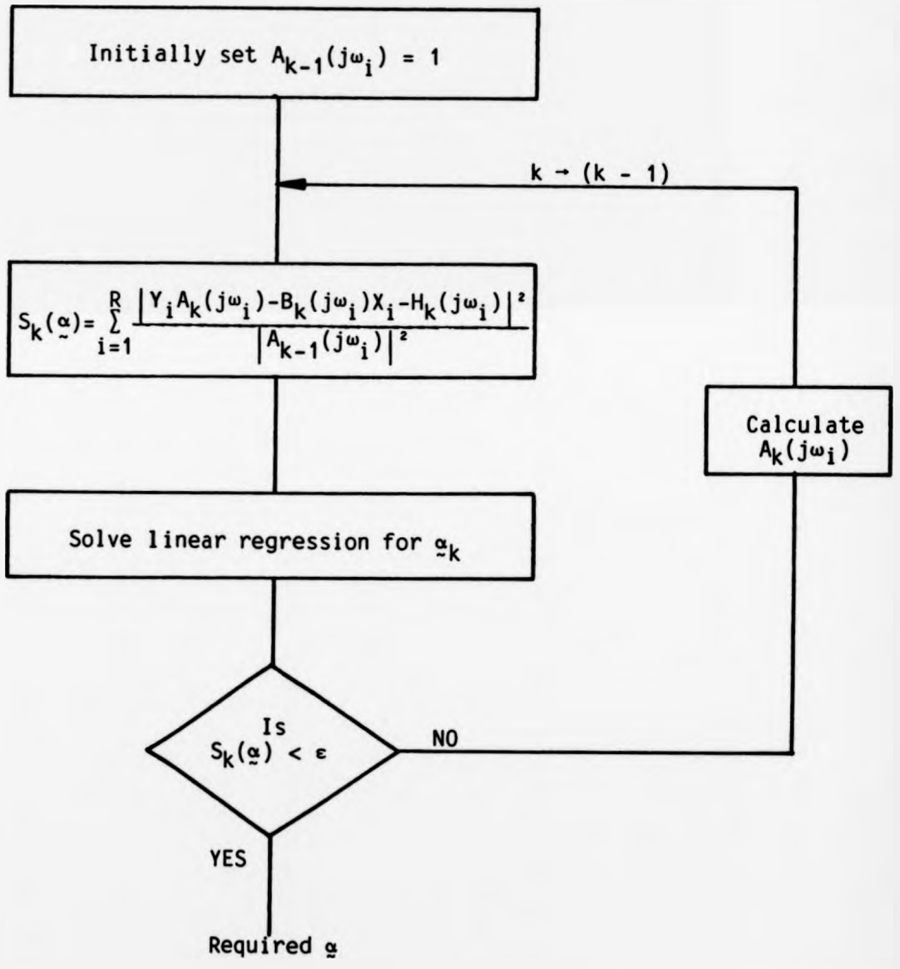


Fig. 4.3 Generalised Least Squares in Frequency

If this vector is broken into its real and imaginary parts

$$\underline{u}_R(k) = \text{Re}\{\underline{u}(k)\}$$

$$\underline{u}_I(k) = \text{Im}\{\underline{u}(k)\}$$

then the analysis follows that in the time domain leading to the equations

$$\underline{\alpha}_k = \underline{\alpha}_{k-1} + \frac{2}{|A_k|^2}\,\underline{P}_k[\underline{u}_R(k)\{\text{Re}(Y_k) - \underline{u}_R^T(k)\underline{\alpha}_{k-1}\} + \underline{u}_I(k)\{\text{Im}(Y_k) - \underline{u}_I^T(k)\underline{\alpha}_{k-1}\}] \quad \ldots\ldots\ldots(4.12)$$

where

$$\underline{Q}_k = \underline{P}_{k-1}\left[\underline{I} - \left(\frac{\underline{u}_R(k)\,\underline{u}_R^T(k)\,\underline{P}_{k-1}}{|A_k|^2/2 + \underline{u}_R^T\,\underline{P}_{k-1}\underline{u}_R}\right)\right]$$

$$\underline{P}_k = \underline{Q}_k\left[\underline{I} - \left(\frac{\underline{u}_I(k)\,\underline{u}_I^T(k)\,\underline{Q}_k}{|A_k|^2/2 + \underline{u}_I^T\,\underline{Q}_k\,\underline{u}_I}\right)\right]$$

Because the data is complex, each frequency point requires two iterations before an updated estimate $\underline{\alpha}_k$ is obtained.

In order to start the algorithm an initial estimate $\underline{P}_o$ is required. As in the time domain approach, $\underline{P}_o$ is set $\underline{P}_o = \beta\underline{I}$ where β is a large scalar quantity. Initially A_o is set to unity as discussed earlier.

4.6 Combination of Data Blocks

The problem is how best to combine data from previous blocks (and previous parameter estimates) with data from the current block in order to improve the estimates. At first sight it would appear that each new frequency point in the current block would add one more row to the $\underline{u}$ matrix in the non-recursive method or update the $\underline{P}$ matrix in the recursive method. However, the

transient terms, and hence the values of $\underline{h}$, are different in each block and this method cannot be used.

The most direct method of updating the estimates would be to average the new values of $\underline{a}$ and $\underline{b}$ with those obtained from previous blocks (this would also have the advantage that only the parameter values need be stored from block to block). However, the estimates are obtained by an effective matrix inversion and the variance of the estimates would be reduced if the new data could be combined with the previous data before the inversion. The problem becomes one of retaining the data pertinent to the $\underline{a}$ $\underline{b}$ parameters but discarding that associated with the $\underline{h}$ parameters.

Two methods have been investigated:

(i) Modification of the data, to remove the $\underline{h}$ terms from the system equation. Data from each block can then be combined as in conventional regression analysis.

(ii) Selective resetting of terms in the $\underline{P}$ matrix as it is updated with data from a new block.

4.6.1 Modification of output data

Returning to eqn. (4.2)

$$Y = \frac{B}{A} X + \frac{H'}{A} + N \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4.2)$$

Suppose the estimation routine is performed over the first data block and estimates $\hat{\underline{\alpha}}$ obtained. Using the estimated values $\hat{\underline{h}}$ $\hat{\underline{a}}$ to calculate $\hat{H}'$, $\hat{A}$, eqn. (4.2) can be written

$$Y - \frac{\hat{H}'}{\hat{A}} = \frac{B}{A} X + N$$

or

$$Y' = \frac{B}{A} X + N$$

where

$$Y' = Y - \frac{\hat{H}'}{\hat{A}}$$

The output has been modified by subtraction of the transient terms. The $\underline{u}$ matrix now becomes

$$\underline{u} = \begin{bmatrix} \underline{p}_1^T \underline{x}_1 , & -\underline{q}_1^T \underline{y}_1' \\ \vdots & \vdots \\ \underline{p}_R^T \underline{x}_R , & \underline{q}_R^T \underline{y}_R' \end{bmatrix}$$

and the corresponding $\underline{\alpha}$ matrix $[b_0 \ldots b_m, a_1 \ldots a_n]^T$ (the h terms being omitted).

This procedure is repeated in the next data block, the output again being modified by the transient terms. Extra rows can now be added to the $\underline{u}$ matrix. This now becomes

$$\underline{u} = \begin{bmatrix} \underline{p}_1^T \underline{x}_1 , & -\underline{q}_1^T \underline{y}_1' \\ \vdots & \vdots \\ \underline{p}_R^T \underline{x}_R , & -\underline{q}_R^T \underline{y}_R' \\ \hline \underline{p}_1^T \underline{x}_1 , & -\underline{q}_1^T \underline{y}_1' \\ \vdots & \vdots \\ \underline{p}_R^T \underline{x}_R & -\underline{q}_R^T \underline{y}_R' \end{bmatrix}$$

First data block

Second data block

In practice the update between data blocks can be most easily performed by updating the matrices $\underline{u}^{*T}\underline{u}$, $\underline{u}^T\underline{u}^*$, $\underline{u}^{*T}\underline{y}$, $\underline{u}^T\underline{y}^*$.

This algorithm has the disadvantage that when the transient is calculated in each block the data used in previous blocks is not used. The algorithm can be modified to overcome this disadvantage.

For the first data block solve for $\underline{\alpha}$ as before and subtract the transient terms to obtain $\underline{y}'$. Form the matrix $\underline{V}_1$ given by

$$\underline{V}_1 = \begin{bmatrix} \underline{p}_1^T x_1 , & -\underline{q}_1^T y_1' \\ \vdots & \vdots \\ \underline{p}_R^T x_R , & -\underline{q}_R^T y_R' \end{bmatrix}$$

This is the matrix in the equation

$$\underline{Y}' = \underline{V}_1\underline{\alpha} + \underline{N}$$

where $\underline{\alpha} = [b_o \ldots b_m, a_1 \ldots a_n]$.

From the matrix $\underline{V}_1$ form the matrix $\underline{u}_1$.

$$\underline{u}_1 = \begin{bmatrix} & \vdots \underline{0} \\ \underline{V}_1 & \vdots \\ & \vdots \underline{0} \end{bmatrix}$$

This matrix fits the equation

$$\underline{Y}' = \underline{u}_1\underline{\alpha} + N$$

where $\alpha = [b_o \ldots b_m, a_1 \ldots a_n, h_o \ldots h_{n-1}]$. Note that this system of equations cannot be solved to give $\underline{\alpha}$ due to the absence of $\underline{h}$ terms.

For the second data block construct the $\underline{u}$ matrix and form the $(\underline{u}^{*T})_2$ matrix according to

$$(\underline{u}^{*T}\underline{u})_2 = \underline{u}^{*T}\underline{u} + \underline{u}_1^{*T}\underline{u}_1$$

and similarly for $(\underline{u}^T\underline{u}^*)_2$, $(\underline{u}^T\underline{Y}^*)_2$, $(\underline{u}^{*T}\underline{Y})_2$

Solve equation for $\underline{\alpha}$ and obtain $\underline{Y}$ by subtracting transients. Then

$$\underline{V}_2 = \begin{bmatrix} \underline{p}_1^T x_1 , & -\underline{q}_1^T y_1' \\ \vdots & \vdots \\ \underline{p}_R^T x_R , & -\underline{q}_R^T y_R' \end{bmatrix}$$

Hence

$$\underline{u}_2 = \begin{bmatrix} & \underline{V}_2 & \vdots \underline{0} \\ & & \vdots \\ & & \vdots \underline{0} \end{bmatrix}$$

For the k_{th} block the procedure is

Form $(\underline{u}^{*T}\underline{u})_k = \underline{u}^{*T}\underline{u} + \sum_{j=0}^{k} \underline{u}_j^{*T}\underline{u}_j$

Similarly for matrices $(\underline{u}^T\underline{u}^*)_k$, $(\underline{u}^T\underline{y}^*)_k$, $(\underline{u}^{*T}\underline{y})_k$.

Solve for $\hat{\underline{\alpha}}$ using the above matrices.

Correct for transients to obtain modified matrices $\underline{V}_k$ and $\underline{u}_k$.

4.6.2 Resetting of terms in the P matrix

Eqn. (4.12) is the equation used to update the estimates at each (complex) data point. This equation is similar to the one used for recursive estimation in the time domain [Hsia (9)] and it can be given a physical interpretation.

New parameter estimate = old parameter estimate

+ $\underline{P}_k$ (error in fitting previous estimate to new data)

The matrix $\underline{P}_k$ weights the fitting error in the updating of $\underline{\alpha}_k$. At the end of a data block it is known that the parameters $\underline{h}$ change and a large error will occur if the old values are fitted to new data. For this reason it seems intuitively correct to reset the elements in the $\underline{P}$ matrix, corresponding to the $\underline{h}$ terms, to their initial values at the end of each data block. The other terms in the $\underline{P}$ matrix are left unaltered.

A similar scheme has been used in the time domain [Hagglund (11)]. Here the purpose was to detect the rapid change in a parameter value due to a failure situation.

4.7 Simulation Results

In the following examples the correct transfer function has the form

$$G(s) = \frac{b_3 s^3 + b_2 s^2 + b_1 s + b_0}{a_3 s^3 + a_2 s^2 + a_1 s + a_0}$$

4.7.1 No noise, linear regression

All results in this section are for 1 experiment consisting of 1 block.

Example 1.

$$G(s) = \frac{1}{(s+1)^3} = \frac{1}{s^3 + 3s^2 + 3s + 1}$$

The parameters in this system were identified in an open-loop configuration when the input was band limited noise. Measurements were made for different block lengths, the number of frequency points used being decreased accordingly.

Example 2.

$$G(s) = \frac{(s-2)(s-3)}{(s+1)^3} = \frac{s^2 - 5s + 6}{s^3 + 3s^2 + 3s + 1}$$

Zeros added to the system of example 1 to give a non minimum phase system. Again system measurements made open-loop and band-limited noise input.

Example 3

$$G(s) = \frac{8}{(s+1)^3} = \frac{8}{s^3 + 3s^2 + 3s + 1}$$

This is the system of example 1 with the gain increased to 8. In this example, the system is measured in a closed-loop configuration with unity feedback and a gain of 8 is that required to give marginal stability.

The results for these three examples are shown in table 4.1, where the frequency spacing ($\Delta\omega = 2\pi/T$) for block length T and the number of frequency points used are also shown.

Example	Block Length sec.	Δω rad/s	No.of freq-uency points	b_0	b_1	b_2	a_1	a_2	a_3
1	True parameter values			1.0	-	-	3.0	3.0	1.0
	25.6	0.245	10	0.999	-	-	2.999	3.008	1.004
	12.8	0.491	5	0.999	-	-	3.015	2.984	1.005
	6.4	0.982	4	0.782	-	-	1.833	2.495	0.641
2	True parameter values			6.0	-5.0	1.0	3.0	3.0	1.0
	25.6	0.245	10	6.003	-4.987	1.003	3.005	3.004	1.011
	12.8	0.491	5	5.955	-4.997	1.137	2.899	2.966	0.884
3	True parameter values			8.0	-	-	3.0	3.0	1.0
	25.6	0.245	10	8.085	-	-	2.946	3.047	0.985
	12.8	0.491	5	8.179	-	-	3.052	3.091	1.017

Table 4.1

These results show good agreement between the estimated parameters and their theoretical values except when the block length was 6.4 seconds. Except in this case, the results are in agreement with the expectation that the method gives unbiased results. The special case in which the block length is 6.4 seconds is an extreme example of a very short block length. The fundamental frequency $\omega = 2\pi/T = 0.98$ is very nearly equal to the corner frequency of the system, $\omega = 1$. This leads to considerable difference in magnitude in the terms of the $\underline{U}$ matrix, hence ill-conditioned equations leading to poor results.

Example 4

This example shows the effect of ignoring the transient term (parameters h_0, h_1, h_2) in the estimation algorithm. Examples

1 and 3 were repeated without these terms present; the results obtained, together with the originals, are shown in Table 4.2.

Example	Block Length sec.	Transient Term	b_0	a_1	a_2	a_3
1	25.6	Included	0.999	2.999	3.008	1.004
	25.6	Not included	0.708	2.173	2.388	0.788
3	25.6	Included	8.085	2.946	3.047	0.985
	25.6	Not included	-1.095	-0.373	-3.346	-1.550

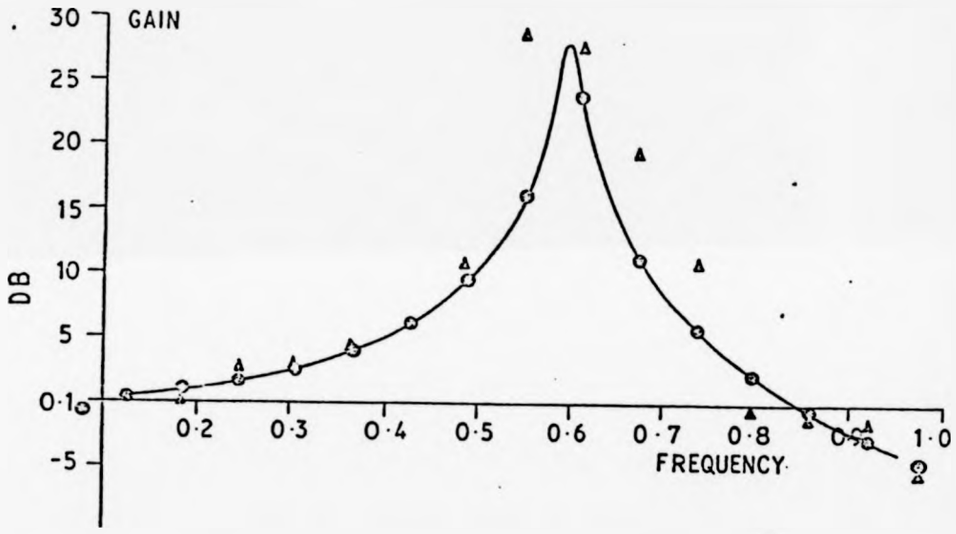
Table 4.2

As can be seen, omission of transient terms causes errors in the order of 20-30% in example 1. In example 3, however, the estimated parameters bear little relationship to their correct values. In example 1 the time constant associated with the transient terms is 1 second and the block length is approximately 25 times this. However, in example 3 the transient term does not decay (system unstable) and not including it in the estimation has a very significant effect.

Example 5

$$G(s) = \frac{0.36}{s^2 + 0.024s + 0.36}$$

This represents a second order lightly damped system with $\omega_n = 0.6$, $\zeta = 0.02$ giving a decay time constant for the transient of $1/\omega_n\zeta = 83.3$ sec. The parameters were estimated over a block length of 102.4 seconds and Fig. 4.4 shows the points on the frequency response calculated from these estimates. Also shown are points calculated from the ratio of cross to auto spectra, a Hanning window being used on input and output data.



Lightly damped second order system.

———— theoretical response

● ● ● ● ● points obtained by regression method

▲ ▲ ▲ ▲ ▲ points obtained by classical method

Fig. 4.4 Lighty damped second order system

4.7.2 Effect of Noise on parameter estimates

Fig. 4.5 shows a comparison of linear and non-linear regression when noise is present. The open-loop system is

$$G(s) = \frac{1}{(1 + s)^3}$$

with an observation time of 25.6 seconds. Fig. 4.5 shows ensemble averages (from 2000 experiments) as the ratio noise/signal power was increased. As can be seen, the bias using linear regression increases as the noise power increases. The non-linear regression removes the bias in the parameter estimates b_0, a_1, a_2, although the parameter a_3 still shows bias.

The same system was used to compare the bias of the estimates under open-loop and closed-loop conditions. The extended least squares method was used for estimation and Fig. 4.6 shows ensemble averages plotted against increasing noise power. As can be seen, the method gives small bias under open-loop conditions which increases under closed-loop conditions.

4.7.3 Combination of data blocks

Fig. 4.7 gives results that show a comparison of the methods described in section 4.6 for updating the estimates as further data becomes available. Again the system was

$$G(s) = \frac{1}{(1 + s)^3}$$

and the estimates were obtained under closed-loop unity feedback conditions using extended least squares. The input was band limited and the noise present had half the power of the input. The block length was 25.6 seconds.

As can be seen, both methods give estimates that converge to a constant value. The method of updating the $\underset{\sim}{P}$ matrix gives

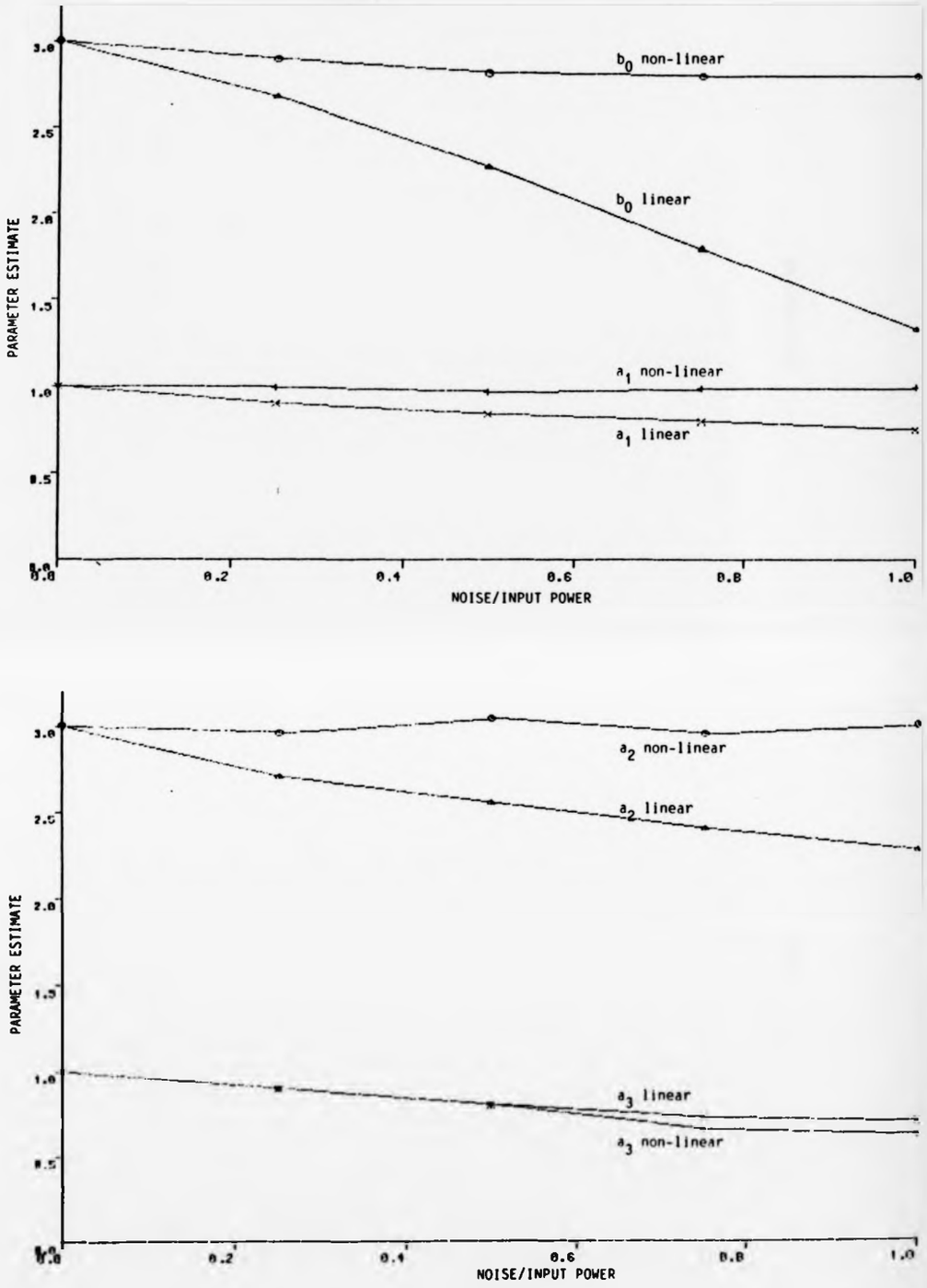


Fig. 4.5 Graphs of Parameter Estimates/Noise Power

Comparison of mean value of estimates for linear and non-linear regression for increasing noise power.

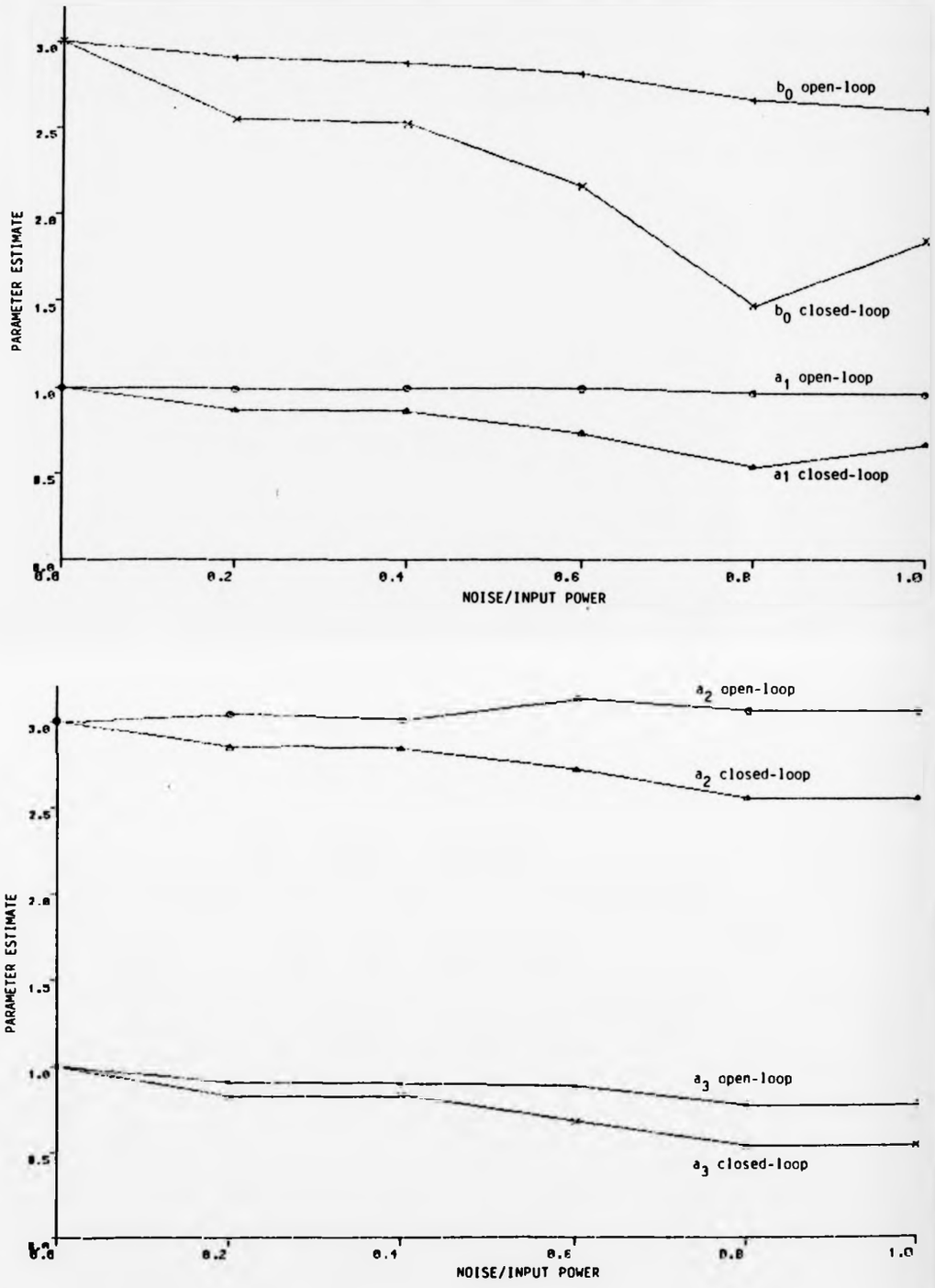


Fig. 4.6 Graphs of Parameter Estimates/Noise Power

Comparison of mean value of estimates obtained in open-loop and closed-loop configurations for increasing noise power.

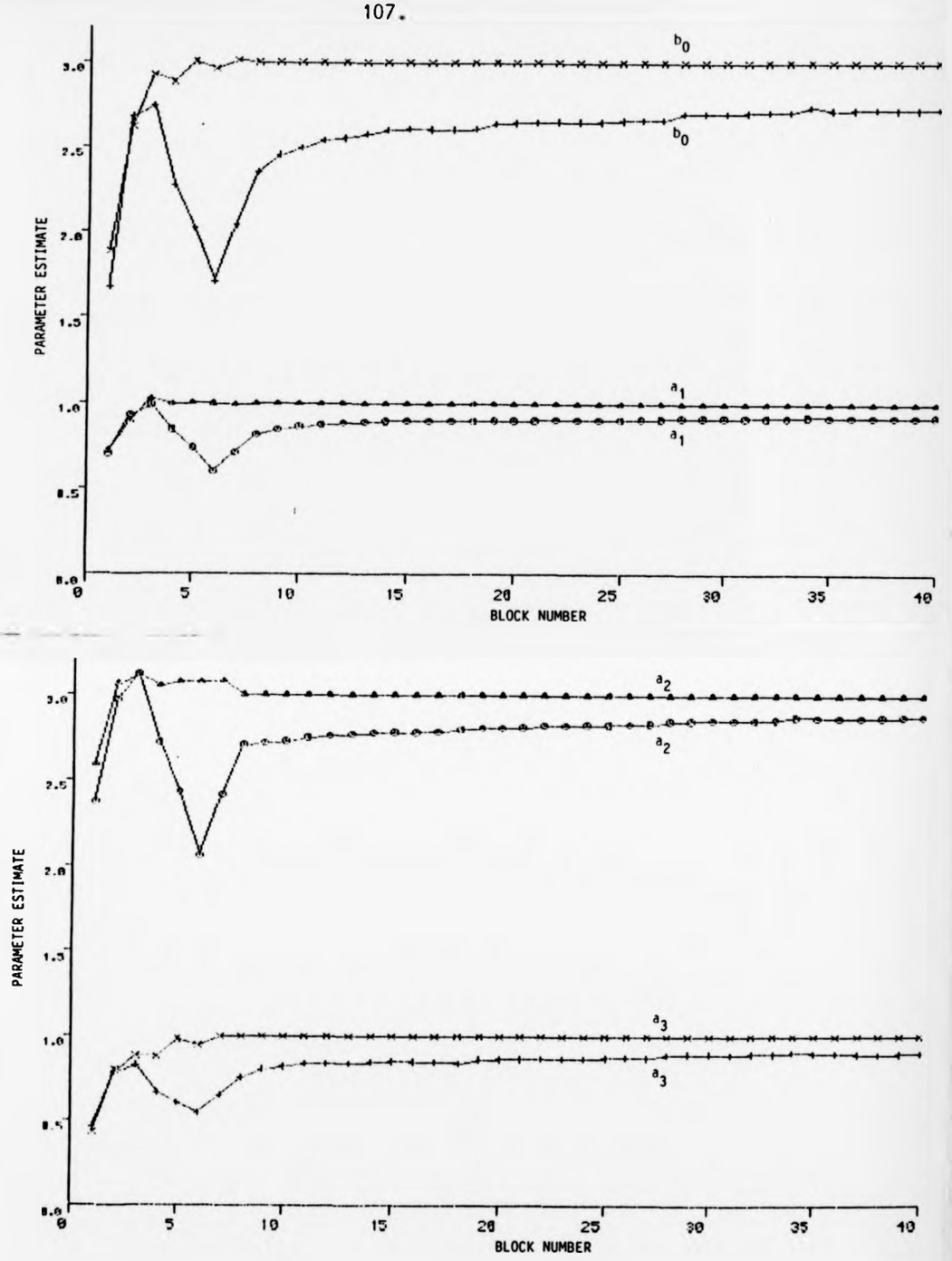


Fig. 4.7 Graph showing updating of parameter estimates with block number

Resetting of elements in P matrix

Modification of data by transient.

more rapid convergence and negligible bias. The method of correcting the data for the transient terms is slower converging and shows some bias.

References - Chapter 4

1. Gawthrop, P.J., "Parametric Estimation from noncontiginous data", Proc. IEE - D, Vol.131, No.6, pp.261-266, Nov. 1984.

2. Solbrand, G., Ahlen, A. and Ljung, L., "Recursive Methods for off-line Identification", Int.J.of Control, Vol.41, No.1, pp.177-191, 1985.

3. Draper, N.R. and Smith, H., "Applied Regression Analysis", John Wiley & Sons, 1966.

4. Eykhoff, P., "System Identification - Parameter and State Estimation", John Wiley & Sons, 1974.

5. Levy, E.C., "Complex Curve Fitting", IRE Trans., AC-4, pp.37-43, 1959.

6. Lawrence, P.J. and Rogers, G.J., "Sequential Transfer Function Synthesis from Measured Data", Proc. IEE, Vol.126, pp.104-106, 1979.

7. Bard, Y., "Non linear parameter estimation", Academic Press, 1974.

8. Gill, P.E. and Murray, W., "Algorithms for the solution of the non-linear least squares problem", S.I.A.M. Journal on Numerical Analysis 15, pp.977-992, 1978.

9. Hsia, T.C., "System Identification - Least Squares Methods", Lexington Books, 1977.

10. Hagglund, T., "New Estimation Techniques for Adaptive Control", PhD Thesis, Dept. of Automatic Control, Lund University, Sweden, Dec. 1983.

CHAPTER 5

Adaptive Control in the Frequency Domain

5.1 Introduction

This chapter investigates how the identification methods discussed previously can be applied to the area of adaptive control. In this introduction a brief outline of adaptive control is presented and a review is made of methods in which frequency domain techniques have been applied. The remainder of the chapter proposes some new adaptive control schemes and investigates their advantages and limitations. Results obtained from simulation of these systems are given in the following chapter.

As many surveys, review articles and tutorial papers have been written on the subject of adaptive control [Asher et al (1), Iserman(2)], only a very brief introduction necessary to outline a framework for frequency response methods will be given here.

Unlike fixed control systems, adaptive systems adapt their behaviour to the (changing) properties of the control processes and their signals. Restricting the discussion to feedback controllers, these can be broadly divided into two groups:

(i) Self-optimising adaptive controllers. These try to reach an optimal performance subject to a given controller type and the obtainable information on the process and its signals.

(ii) Model reference adaptive controllers. These try to reach, for a given input signal, a control behaviour close to a given reference signal.

These two groups can be shown in block diagram form as in Figs. 5.1(a) and 5.1(b).

Within these two broad groups a large number of individual schemes are possible, depending on:

(i) the form of the system model;

(ii) the method of identification;

(iii) the criterion for controller design;

(iv) the controller algorithm.

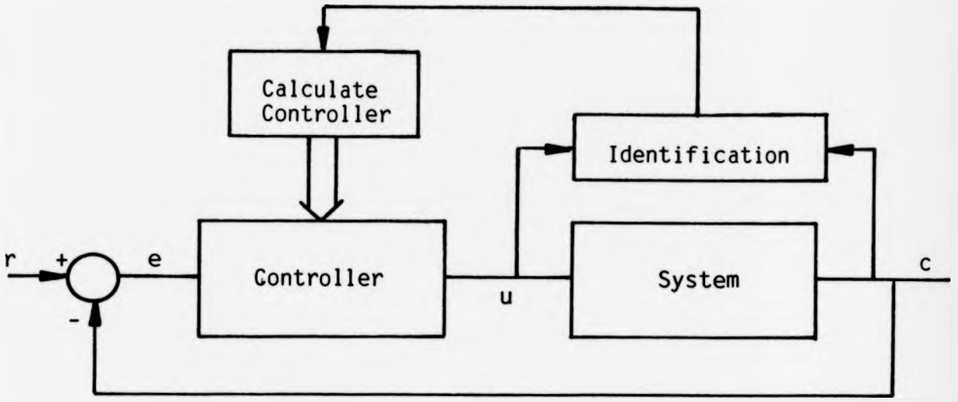


Fig. 5.1(a) Self-optimising adaptive controller.

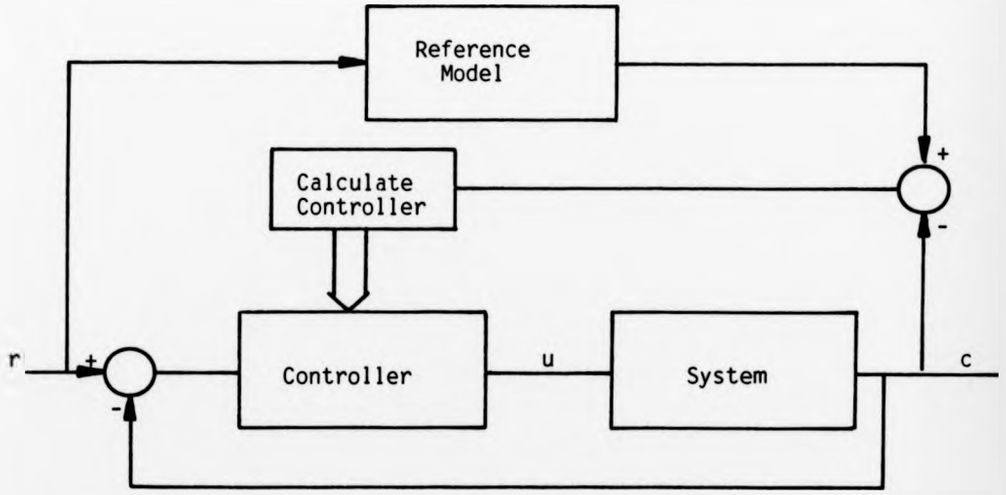


Fig. 5.1(b) Model reference adaptive controller

In this thesis the interest is in systems where the system model is given by its frequency response function and the performance criterion is based in the frequency domain. Although a great number

of adaptive control schemes have been proposed, the number using frequency response methods is small. Asher (1) lists over 700 papers on the subject of adaptive control; of these only six make reference to frequency domain methods. The possible reasons for this will be explained later.

Neglecting the use of sinusoids as perturbation signals in hill-climbing systems, the earliest use of frequency domain techniques involved identification at one frequency only. Smith (3) describes such a technique applied to adaptive control of an aircraft autopilot. This is shown in Fig. 5.2.

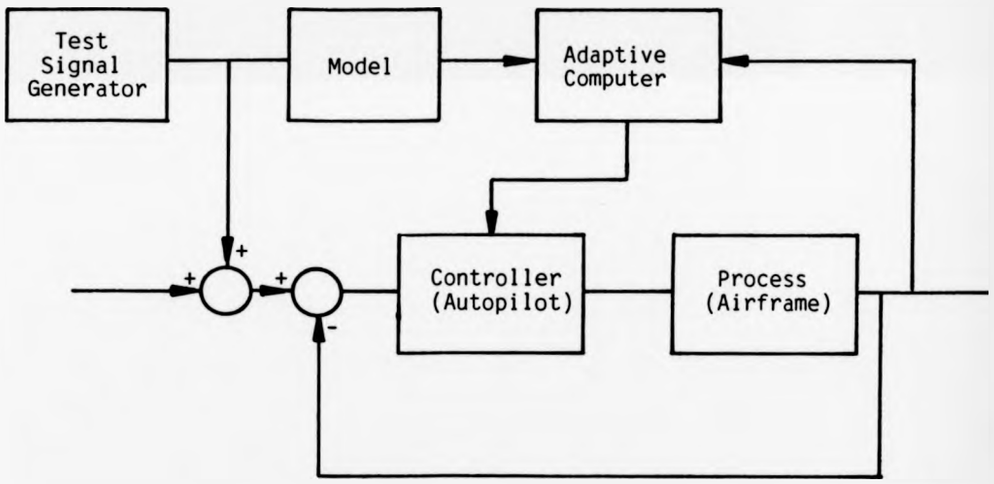


Fig. 5.2 Adaptive control for aircraft autopilot

The test signal is a single sinusoid and the adaptive computer measures the gain or the phase shift of the basic loop output at this frequency and compares it with that determined by the model. Any difference is used to control the gain of the autopilot. The paper compares the use of gain with that of phase as a performance index and discusses the factors determining the choice of test frequency.

This basic idea is extended in papers by Womack and Watt (4), Rao and Seshadri (5), Gururaja and Deekshatulu (6), Narayan and Seshadri (7). These papers describe systems using both gain and phase measurement to obtain improved adaptation for one parameter, extend the adaptation to two parameters and, by injecting a further test signal, extend to three parameters. The adaptive computer measures the gain and phase at the test frequencies and pre-programs the non-linear relationship between measured parameters and controller parameters.

A variation on this method is to provide the test signal not from an external source but by causing the system to be self oscillating. This is done via an additional feedback path containing a non-linearity (ideal relay) to limit the amplitude of the self oscillation. Such systems are described by Horowitz (8), Horowitz et al (9).

An alternative method of adaptation using frequency domain criteria is described by Staffin (10). The normal operating signal is used as input and the value of ω_m (frequency of the maximum value of closed-loop gain) used as a performance criterion. The system is shown in Fig. 5.3.

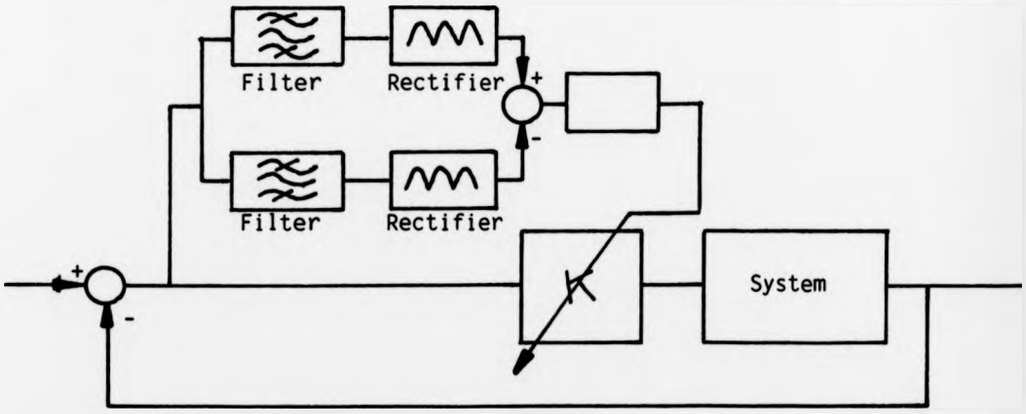


Fig. 5.3 Adaptation using ω_m criteria

Two band-pass filters tuned either side of ω_m are connected to the error signal, the difference between the rectified outputs of these filters is integrated to control the system gain. If ω_m is correct both filters give the same output and the gain is unaltered; any change in system gain causes a change in ω_m and a corresponding correction to the controller gain K.

In a recent paper, Bin and Chun-bo (11), propose a scheme where identification takes place at a number of frequencies. This is done by a series of narrow band filters at the input and output of the system and uses normal operating signals. Little specific detail on the implementation of such a scheme is given and the paper concentrates on whether convergence to a stable equilibrium point can be obtained.

The majority of papers outlined in the previous pages were written before 1974 and none of them used modern ideas of frequency response estimation. The probable reason for the lack of interest in frequency domain methods was the growth that has occurred in time domain methods. These gave recursive methods of identification well suited to implementation on small digital computers and micro-processors. This, together with implicit methods of adaptation, gave rise to the interest in self-tuning algorithms.

Although there was a lack of frequency response methods applied directly to adaptive control problems, modern methods of spectrum estimation have been applied to related fields. One of these areas is the production of a transducer signal (displacement, velocity, acceleration) with a prescribed frequency spectrum on a test specimen where the action causing the transducer response is remote from the transducer and hence modified by the specimen and test fixing.

One specific area is the reproduction of road response data on vehicles (Dodds (12), Styles & Dodds (13), Cryer & Nawrocki (14). For test purposes the vehicle is placed on wheel pans attached to vertical actuators which must be excited so as to produce signals corresponding to road conditions at specific points on the vehicle. Direct recordings of road noise cannot be fed to the actuators because:

(i) it is difficult to record road displacements as would be experienced by the tyre.

(ii) the characteristics of the tyre are different in the rolling and non-rolling conditions.

The problem is complicated by the correlation that exists between the signals applied to each wheel.

Another area of application is in wave making for model testing in a wave tank (Borgman (15)). Observed waves at a given point in the tank must be reproduced. This is in spite of the fact that the actuator producing the waves may be some distance from the point and reflections from the tank wall can modify the wave pattern.

A third area of application is the simulation of the effect of earthquakes on model structures (Carvalhal et al (16)). Reproduction of a given vibration spectrum at some location on the model is required. This spectrum must be maintained throughout the duration of the test even though the dynamic characteristics of test rig and model vary due to failure modes.

The method of solution in all these areas is shown in Fig. 5.4.

The spectrum as measured by the transducer is compared, component by component, with the required spectrum. The differences form an error spectrum which is inverted to form the drive signal to the actuator.

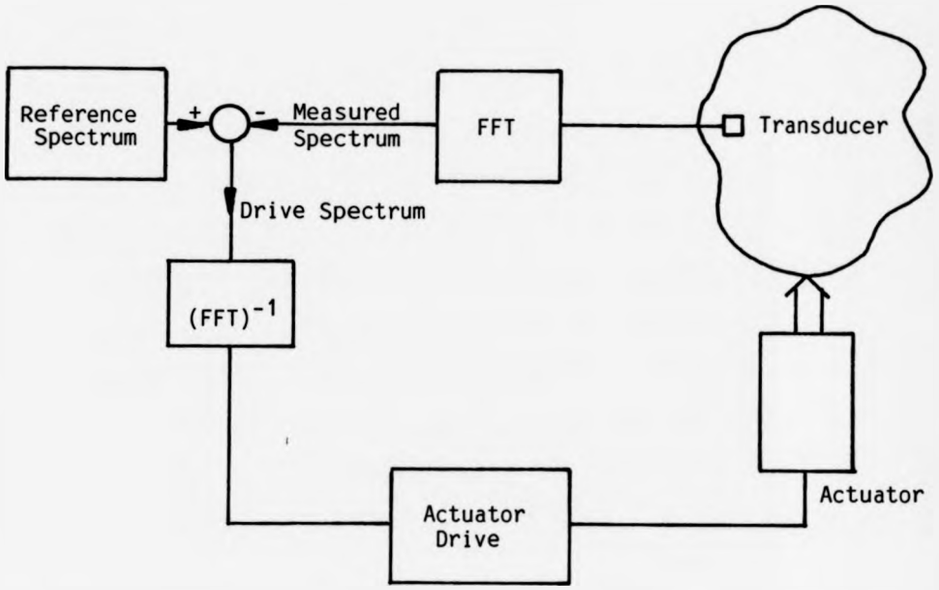


Fig. 5.4 Adaptive System to produce given spectrum

Another allied field where spectral methods have been used is that of adaptive filtering. Adaptive filtering has been applied to problems in speech encoding, interference cancellation, analysis of biological signals and channel equalisation. The usual adaptive filter used is a transversal filter where the outputs of a tapped delay line are weighted and summed under the control of a recursive algorithm to form the filter output. Dentino et al (17), Bershad and Feintuch (18), describe a frequency domain adaptive filter which is shown in fig. 5.5.

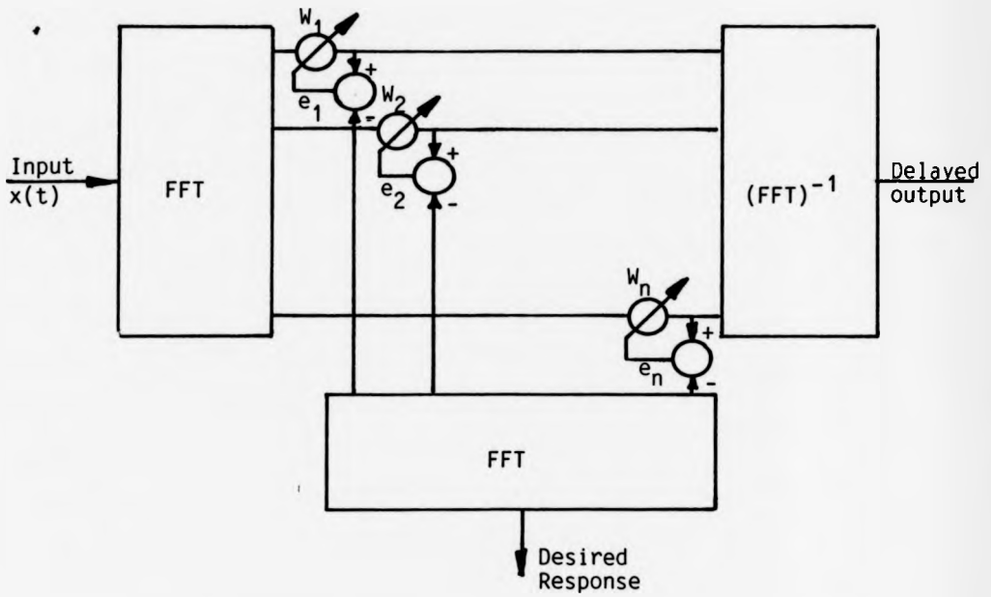


Fig. 5.5 Adaptive filter in frequency domain

Each component of the FFT of the input signal is weighted (weights W_1, W_2 .....W_n) and the inverse transform taken to give the output signal. Initially, in the first data block, these weights are set at unity. The FFT of the desired response is compared, component by component, with the weighted FFT of the input. The complex error signals are used to update the weightings, according to a suitable algorithm, in the next data block. The authors claim a much reduced computing requirement for this type of filter even when the need for three FFT per data block is taken into account.

The main advantages of the application of modern frequency domain techniques to adaptive control are considered to be:

(i) For non-parametric identification no assumptions need to be made concerning the system transfer function. Such assumptions are the order of the system, the number of poles and zeros and the magnitude of any pure time delays.

(ii) The performance criteria can be specified in the frequency domain. As much design and tuning of systems is already done in

terms of parameters such as gain and phase margins, gain and phase cross-over frequencies, M_{max} and ω_{max} and bandwidth, this gives a practical method of specifying controller settings. It also offers the possibility of prescribing the complete frequency response of the system.

(iii) It gives the possibility of imposing constraints, that are frequency domain constraints, in the controller. For example, the high frequency content of the controller output signal could be restricted to avoid undue acceleration of following mechanical components.

5.2 General Adaptive Scheme

The general method of adaption can be understood by reference to Fig. 5.6.

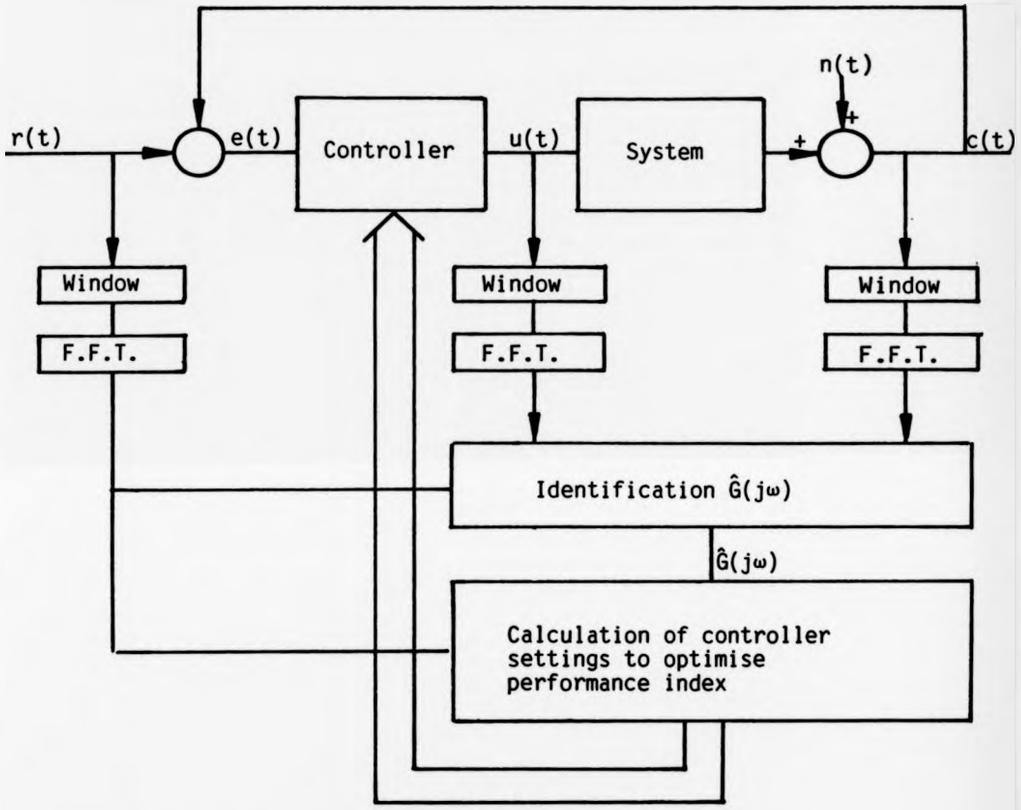


Fig. 5.6 General method for adaptation in frequency

The method bears similarities to adaptive schemes that are used in the time domain. The major difference however is that the control signal u(t) is obtained from a performance index that is based on the estimate of the system frequency response function G(jω). This in turn infers that complete data blocks (r(t), u(t), c(t)) must be obtained over an observation time T, before any change can be made to the controller parameters. This is in contrast to

the time domain when the control signal can be updated at every sample. In order to track changing parameters it is required to keep T small; this however increases the bias and variance of the estimates as discussed in previous chapters. Some form of updating algorithm is required to combine data in the present block with that from previous blocks in order to give an improved estimate.

The general method shown in Fig. 5.6 can lead to a number of individual schemes depending on,

(i) the form of controller,

(ii) whether the identification is parametric or non-parametric,

(iii) the nature of the performance index chosen.

As one advantage of the frequency domain approach is the use of common and well understood design parameters, it was decided to restrict the controller to a P.I.D. controller. This controller is remarkably effective in regulating a wide range of processes and many algorithms exist for setting its parameters. The continuous form of its frequency response function will be used although in practice, in adaptive form, it would probably be implemented digitally. Hence:

$$D(j\omega) = \frac{U(j\omega)}{E(j\omega)} = K + \frac{K_I}{j\omega} + j\omega T_D$$

where K, K_I and T_D represent proportional gain, integral gain and derivative time constant respectively.

The effect of parametric or non-parametric identification alters the details of the adaptive scheme and both methods will be investigated.

The choice of performance index has the greatest influence on the details of the method. Two performance indices will be investigated:-

the time domain when the control signal can be updated at every sample. In order to track changing parameters it is required to keep T small; this however increases the bias and variance of the estimates as discussed in previous chapters. Some form of updating algorithm is required to combine data in the present block with that from previous blocks in order to give an improved estimate.

The general method shown in Fig. 5.6 can lead to a number of individual schemes depending on,

(i) the form of controller,

(ii) whether the identification is parametric or non-parametric,

(iii) the nature of the performance index chosen.

As one advantage of the frequency domain approach is the use of common and well understood design parameters, it was decided to restrict the controller to a P.I.D. controller. This controller is remarkably effective in regulating a wide range of processes and many algorithms exist for setting its parameters. The continuous form of its frequency response function will be used although in practice, in adaptive form, it would probably be implemented digitally. Hence:

$$D(j\omega) = \frac{U(j\omega)}{E(j\omega)} = K + \frac{K_I}{j\omega} + j\omega T_D$$

where K, K_I and T_D represent proportional gain, integral gain and derivative time constant respectively.

The effect of parametric or non-parametric identification alters the details of the adaptive scheme and both methods will be investigated.

The choice of performance index has the greatest influence on the details of the method. Two performance indices will be investigated:-

(i) minimisation of error power,

(ii) specification of desired closed-loop performance.

These two methods can be compared to minimum variance and pole-placement methods used in the time domain.

5.3 Minimisation of Error Power

The total error power is the integral of the error spectral density (this is also the error variance).

$$V = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Phi_{ee}(j\omega)d\omega \qquad (5.1)$$

The estimated error spectral density $\hat{\Phi}_{ee}(j\omega)$ can be expressed in terms of the estimated input spectral density $\hat{\Phi}_{rr}(j\omega)$ the controller $D(j\omega)$ and the estimated transfer function $\hat{G}(j\omega)$.

$$\Phi_{ee}(j\omega) = \left|\frac{1}{1 + D(j\omega)\,\hat{G}(j\omega)}\right|^2 \hat{\Phi}_{rr}(j\omega) \qquad (5.2)$$

and

$$V = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left|\frac{1}{1 + D(j\omega)\,\hat{G}(j\omega)}\right|^2 \hat{\Phi}_{rr}(j\omega)d\omega \qquad (5.3)$$

This expression for V must be minimised with respect to the controller parameters K, K_I, T_D. The parameter values for minimum V are then implemented in the controller. Details of the scheme are shown in 5.7 and the steps in the adaptive routine are listed below before being considered in more detail.

1. Short term Fourier Transforms are taken of the signals r(t), u(t), c(t), suitable windowing being used dependent upon the identification scheme.
2. From these transforms an estimate of the system transfer function $\hat{G}(j\omega)$ is made - this may be parametric or non-parametric in form.
3. From the input Fourier Transform an estimate of the input spectral density $\hat{\Phi}_{rr}(j\omega)$ is made.
4. Using the initial controller settings, the error transfer function is obtained.
5. From the estimated error transfer function and input spectral density, the estimated error power V is obtained.

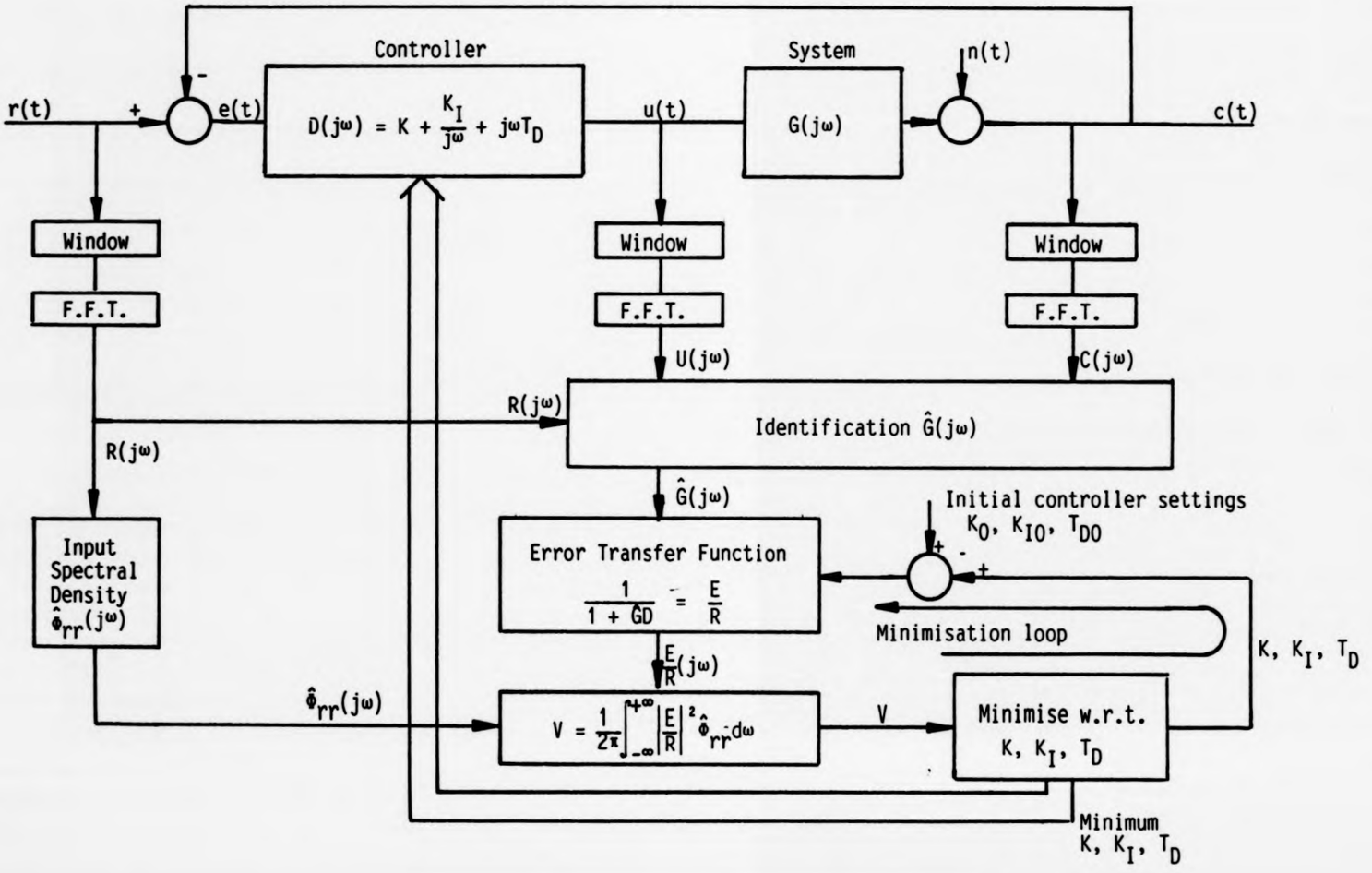


Fig. 5.7 Adaptation by minimisation of error power

6. By forming a minimisation loop around steps 4 and 5, the controller settings that minimise V are obtained.

7. The controller settings obtained in 6 are implemented.

8. The process is repeated for the next data block, FFT data being combined with previous data to obtain improved estimates.

5.3.1 Estimation of Input Spectral Density

The estimation techniques discussed in the previous chapters have concentrated upon the estimation of frequency response functions. The estimation of power density spectra is very well documented (Kay & Marple (19)) and the model can be parametric or non-parametric.

For a non-parametric model the estimator

$$\bar{\Phi}_{rr}(j\omega) = \frac{1}{L} \sum_{1}^{L} R(j\omega)R^*(j\omega)$$

is used, where L is the number of blocks. To account for time varying input spectra a forgetting factor λ can be introduced, giving

$$\Phi_{rr}(j\omega)\Big|_L = (1 - \lambda)\bar{\Phi}_{rr}(j\omega)\Big|_{L-1} + \lambda\Phi_{rr}(j\omega)\Big|_L$$

The use of parametric models for power spectra estimation has been the subject of much research over recent years (Cadzow (20)). A general approach is to assume that the spectrum to be estimated has been obtained by passing white noise through an appropriate transfer function type model; the estimation routine then identifies the parameters in the model. It should be noted that the methods described in previous chapters cannot be used to take the finite observation time into account as the white noise input is assumed unobservable. One method of taking the effects of finite observation

time into account is to overparametrise the model (Cadzow (20)) although methods such as maximum entropy can be applied (Jaynes (21)).

5.3.2 Estimation of Error Power

The total error power has already been given in equations (5.1) and (5.2). Repeating these equations for convenience

$$V = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Phi_{ee}(j\omega)d\omega \qquad \text{(5.1)}$$

$$\Phi_{ee}(j\omega) = \frac{1}{1 + D(j\omega)\,\hat{G}(j\omega)}\,\hat{\Phi}_{rr}(j\omega) \qquad \text{(5.2)}$$

If the estimates for $\hat{G}(j\omega)$ and $\hat{\Phi}_{rr}(j\omega)$ are obtained in parametric form then the expression for $\Phi_{ee}(j\omega)$ can be written

$$\Phi_{ee}(j\omega) = \frac{c(j\omega)\,c^*(j\omega)}{d(k\omega)\,d^*(j\omega)} \qquad \text{(5.4)}$$

where $c(j\omega) = \sum_{k=0}^{n-1} c_k(j\omega)^k$

and $d(j\omega) = \sum_{k=0}^{n} d_k(j\omega)^k$

where the coefficients c_k and d_k are known.

Hence from equations 5.1 and 5.4 the total error power is given by

$$V = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{c(j\omega)\,c^*(j\omega)}{d(j\omega)\,d^*(j\omega)}\,d\omega$$

This integral has been evaluated and is available in tabulated form, V being expressed for given n in terms of the parameters c_k, d_k. This form however is not very suitable in this application. A more suitable form has been developed by Astrom (22) who uses a recursive algorithm that can be used for any order system.

In addition this program also gives information about the stability of the system (poles in right hand half plane). This information is required as will be shown in section 5.5.

If non-parametric estimation is used then $\Phi_{ee}(j\omega)$ is available as a set of magnitudes corresponding to discrete frequency points. In this case the integral of equation (5.3) is replaced by a summation.

$$V = \frac{1}{2\pi} \sum_{i=1}^{M} \left| \frac{1}{1 + D(j\omega_i)\, \hat{G}(j\omega_i)} \right|^2 \Phi_{rr}(j\omega_i)\, \Delta\omega$$

5.3.3 Minimisation of Error Power

The calculated value of V depends upon the parameters K, K_I, T_D in the controller $D(j\omega)$. Values of these parameters that minimise V are required. Analytical minimisation of such an expression would lead to a set of non-linear equations which cannot be solved directly. Some form of iterative technique is required. At this stage it is also required to put non-linear constraints on the values of the parameters.

(i) To prevent them taking negative values.

(ii) To prevent them taking values that would produce instability if implemented (see section 5.5).

(iii) To prevent them taking values that would produce an undesirable spectrum for the drive to the system $U(j\omega)$.

Because of these constraints, it was decided to use a minimisation technique of the "direct search type". Several such techniques exist, but, with only three parameters in the search, there is little to choose between them and the technique used was the Hooke & Jeeves method (Hooke & Jeeves (23)).

5.3.4 Updating of estimates

The method used depends on whether parametric or non-parametric estimates are used. At this stage "forgetting factors" are introduced to cater for time-varying parameters or input. The techniques used have been treated in detail in the previous chapters.

5.4 Specification of Closed-Loop Performance

This method is a model reference system where the required closed-loop frequency response function is compared to that obtained from the estimated $\hat{G}(j\omega)$. The controller settings are chosen to minimise a performance criterion based on the difference between these responses.

The estimated closed-loop frequency response function can be written in terms of $\hat{G}(j\omega)$ and $D(j\omega)$ as

$$\left(\frac{\hat{C}}{R}(j\omega)\right) = \frac{\hat{G}(j\omega)\ D(j\omega)}{1 + \hat{G}(j\omega)\ D(j\omega)}$$

If the desired closed-loop frequency response function is $M(j\omega)$ then a suitable function for the performance criterion could be

$$S = \int_{-\infty}^{+\infty} \left| M(j\omega) - \left(\frac{\hat{C}}{R}(j\omega)\right) \right|^2 d\omega$$

Such a criterion would entail specifying the gain and phase of $M(j\omega)$. Alternatively, by specifying just the magnitude $|M(j\omega)|$ then the function to be minimised is

$$S = \int_{-\infty}^{+\infty} \left(|M(j\omega)| - \left|\frac{C}{R}(j\omega)\right| \right)^2 d\omega$$

The function $|M(j\omega)|$ could be specified by numerical values at harmonic frequencies. Alternatively it could be specified in terms of a "standard frequency response function." This function would depend upon the order of the system and could be specified in terms of design parameters such as M_{max}, ω_{max}, bandwidth, steady-state error coefficients.

As with the minimum error power method, estimation could produce a parametric or non-parametric model. This in turn would

alter the details of the minimisation routine. If a parametric model is obtained the possibility exists of specifying the controller settings in terms of the performance index and the required design parameters. This would give an implicit method of adaption (there is no requirement of minimising a function) that resembles the pole placement method in the time domain.

5.5 Stability

Particular difficulties can arise when the mean-square error is being evaluated numerically, since unstable systems lead to finite values for this quantity.

In the time domain recursive estimation can produce a control signal u(t), which, due to noise, or during the initial transient period, can give an unstable system. However, as the estimate is updated one sample later, this is unlikely to cause problems. In the frequency domain schemes suggested, once the controller parameters are implemented, they remain at that value for one complete data block. If parameters that caused instability are implemented then the oscillation would grow over the time of tne data block and this could prove disastrous in practice. Without some form of constraints being placed on the controller parameters this situation is quite likely to occur.

Consider the Nyquist plot for the function $\hat{G}(j\omega)\ D(j\omega)$ as shown in Fig. 5.8.

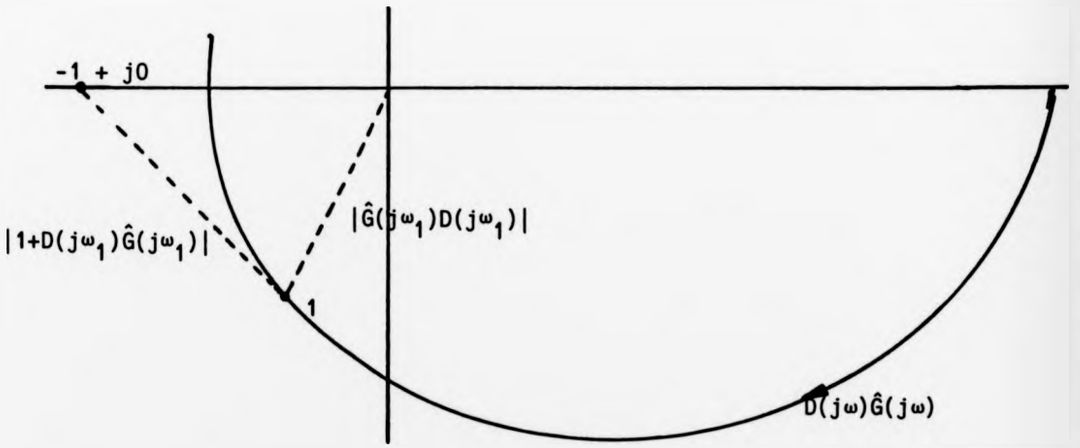
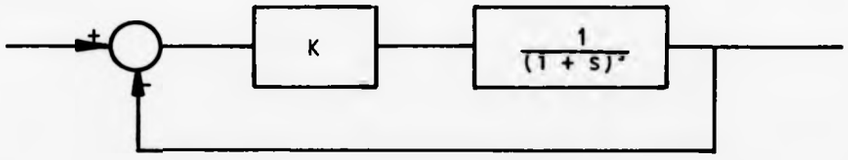


Fig. 5.8 Graphical representation of error frequency response

The error spectral density Φ_{ee} is given by

$$\Phi_{ee}(j\omega) = \frac{1}{|1 + D(j\omega)\ \hat{G}(j\omega)|^2}\ \Phi_{rr}(j\omega)$$

At a given frequency ω_1, $|1 + D(j\omega_1)\ \hat{G}(\omega_1)|$ represents the distance from the -1 point to the point $\hat{G}(j\omega_1)\ D(j\omega_1)$. The reciprocal of this distance squared represents the weighting given to $\Phi_{rr}(j\omega)$ to produce the contribution to $\Phi_{ee}(j\omega)$. Should variation of one of the parameters in $D(j\omega)$ cause the point $\hat{G}(j\omega_1)\ D(j\omega_1)$ to move closer to the -1 point this weighting will increase, rising to infinity at the -1 point - instability occurs. Further change of the parameters will cause the weighting to fall even though the system is now unstable. This is true for all frequency points; the total error power will be infinite at instability; it will then fall and may exhibit minima in the unstable region. This situation is illustrated for a specific system with a simple gain controller in Fig. 5.9.



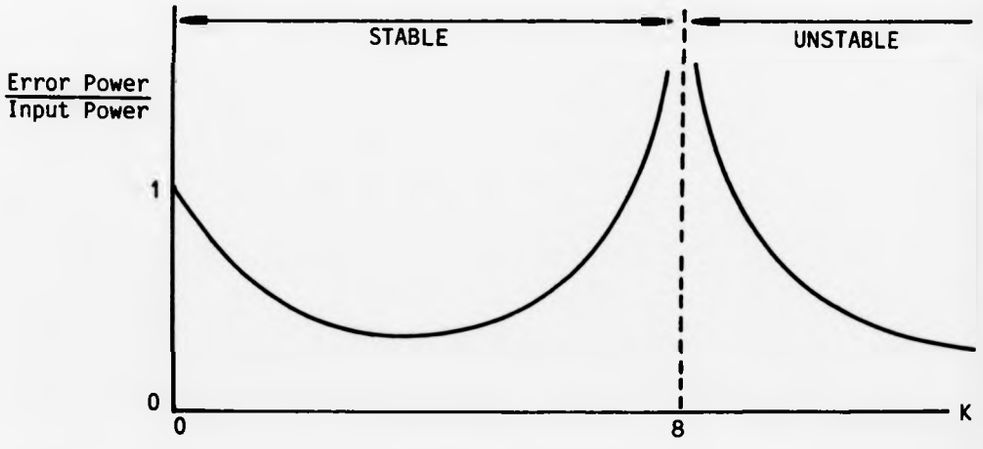


Fig. 5.9 Variation of calculated error power with gain

The search method used has the characteristic that the step length in the search is increased whenever a successful iteration is performed. The possibility exists of a step moving the search into the unstable region and continuing towards a minimum in this region.

Some form of constraint must be incorporated into the search routine to prevent this happening. Two conditions are checked at each calculation of V in the search routine; if either are violated that point is treated as a constraint on the search.

(i) The distance between the -1 point and the point $\hat{G}(j\omega)\ D(j\omega)$ is calculated. The condition imposed is that this should not be less than a certain value. As the centres of M circles on the Nyquist diagram approach the -1 point as M increases this is equivalent to imposing a constraint of a certain M_{max} value.

(ii) A check is made on the stability of the system. With the parametric form of $\hat{G}(j\omega)\ D(j\omega)$ the algorithm of Astrom is used to calculate the error power also checks that the coefficients satisfy the conditions for stability. With a non-parametric model a new numerical algorithm for testing Nyquist's Stability Criterion was developed; this algorithm is given as appendix 1

The reasons that two constraints are imposed in the search routine are as follows. Constraint (i) is very easy to check computationally and does not slow down the search algorithm. It will provide an effective constraint in most cases. It could fail, however, if the required minimum is very close to the stability boundary. Constraint (ii) takes more time to execute, especially in the non-parametric form, but will detect any situations where the search passes into the unstable region but is not detected by (i).

References - Chapter 5

1. Asher, "Bibliography on Adaptive Control Systems", Proc. IEEE, Vol.64, No.8, pp.1226-1240, Aug. 1976.

2. Isermann, R., "Parameter Adaptive Control Algorithms - A Tutorial", Automatica Vol.18, No.5, pp.513-528, 1982.

3. Smith, K.C., "Adaptive Control through Sinusoidal Response", Trans. IRE, AC-6, pp.129-137, 1962.

4. Womack, B.F. and Watt, J.T., "A Two Parameter Adaptive Control System using Sinusoidal Test Signals", Trans. IRE, AC-10, pp.194- 1965.

5. Rao, D.V.R. and Seshadri, V., "Application of Phase Equivalent Reduction in Multiple Parameter Identification using Sinusoidal Test Signals", Int.J.Control IV, 6, pp.567-584, 1966.

6. Gururaja, H.V. and Deekshatulu, "Process Identification and Adaptation by High Frequency Sinusoidal Test Signals", Int.J.Control, Vol.3, No.6, pp.567-586, 1966.

7. Narayan, D.V.L. and Seshadri, V., "Design of an Adaptive Control System using Frequency Domain Criteria", J.Inst. Electron. and Telecomm., Vol.23, part 5, pp.293-297, 1977.

8. Horowitz, I.M., "Comparison of Linear Feedback Systems with Self-oscillating Adaptive Systems", IEEE Trans. Aut. Cont., AC-9, pp.386-392, Oct. 1974.

9. Horowitz, I.M., Smay, J.W. and Shapiro, A., "A Synthesis Theory for Externally Excited Adaptive Systems", IEEE Trans. on Aut. Control, AC-9, pp.101-107, April 1974.

10. Staffin, R., "Executive Controlled Adaptive Systems", AIEE Trans. Applic. and Industry, pp.523-530, Jan. 1962.

11. Bin Lin and Chun-bo Feng, "Frequency Domain Approach for the Design of Adaptive Control and System Identification", I.F.A.C. Conference, York, UK, 1985.

12. Dodds, C.J., "The Laboratory Simulation of Vehicle Service Stress", J.Eng. for Industry (ASME) 96(2), pp.391-398, 1974.

13. Styles, D.D. and Dodds, C.J., "Simulation of Random Environments for Structural Dynamics Testing", Experimental Mechanics, Vol.16, No.11, pp.416-424, Nov. 1976.

14. Cryer, B.W. and Nawrocki, P.E., "A Road Simulation System for Heavy Duty Vehicles", Soc. of Automotive Engineers, Automotive Engineering Congress and Exposition, Detroit, Michigan, Feb. 23-27 1976.

15. Borgman, L.E., "Ocean Wave Simulation for Engineering Design", J.of Waterways and Harbours Div. Proc. Am. Soc. Civil Eng., WW4, pp.557-582, 1969.

16. Carvalhal, F.J., Wellstead, P.E. and Pereira, J.J., "Identification and Adaptive Control in Earthquake and Vibration Testing of Large Structures", I.F.A.C. Symposium on Identification and Parameter Estimation, Georgia, USSR, Sept. 1976.

17. Dentino, M., McCool, J. and Widrow, B., "Adaptive Filtering in the Frequency Domain", Proc. IEEE, Vol.66, pp.1658-1660, Dec. 1978.

18. Bershad, N.J. and Feintuch, P.L., "Analysis of the Frequency Domain Adaptive Filter", Proc. IEEE, Vol.67, No.12, Dec. 1979.

19. Kay, S.M. and Marple, S.L., "Spectrum Analysis - A Modern Perspective", Proc. IEEE, Vol.69, No.11, pp.1380-1419, Nov. 1981.

20. Cadzow, J.A., "Spectral Estimation: An Overdetermined Rational Model Equation Approach", Proc. IEEE, Vol.70, No.9, Sept. 1982.

21. Jaynes, E.T., "On the Rationale of Maximum-Entropy Methods", Proc. IEEE, Vol.70, No.9, pp.939-952, Sept. 1982.

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

23. Hooke, R. and Jeeves, T.A., "Direct Search Solution of Numerical and Statistical Problems", Journal of the Assoc. for Computing Machinery, No.8, pp.212-229, 1961.

CHAPTER 6

Adaptive control - Simulation Studies on Specific Systems

6.1 Introduction

The previous chapter considered methods by which adaptive control can be applied using frequency domain techniques. This chapter considers some specific systems and obtains simulation results to illustrate these methods. Two basic systems are chosen and three methods of implementing adaptive control are considered. The two systems investigated are a third order system having transfer function

$$G(s) = \frac{1}{(1 + s)^3}$$

and a first order system together with a delay

$$G(s) = \frac{e^{-s}}{(1 + 5s)}$$

The methods investigated are

(i) Minimisation of error power, both parametric and non-parametric schemes are examined.

(ii) Adaptation to a prescribed closed-loop frequency response.

6.2 Minimisation of error power - system $1/(1 + s)^3$

The overall system is as shown in Fig. 6.1.

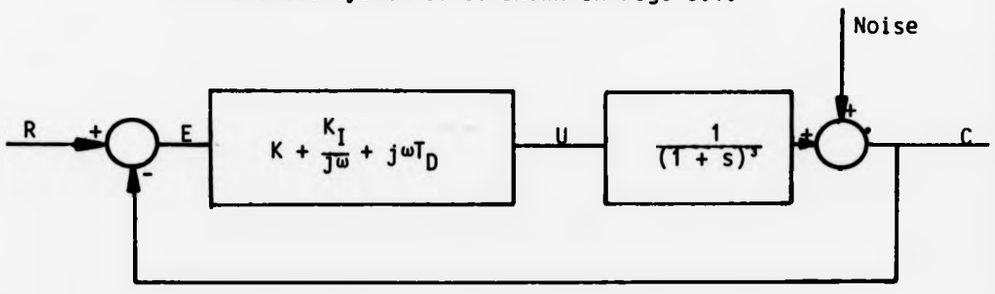


Fig. 6.1 Adaptive controller - system $1/(1 + s)^3$

The adaptive system was such as to set the parameters K, K_I, T_D in the three term controller to minimise the error power. The complete system is described in the previous chapter. The input signal was obtained by passing white noise through a first order lag having a time constant of 10 seconds and the noise disturbance was chosen to have an identical spectrum.

Theoretical plots of error power/parameter value are shown in Fig. 6.2. The error power has been normalised by expressing it as a ratio of input power and for each graph two of the parameters have been held constant. The error power when plotted against K and K_I shows a minimum, but when plotted against T_D there was no minimum, the graph falling monotonically as T_D was increased. In practice the effect of large values of T_D would be to increase the high frequency content of the error power spectrum. As this would be undesirable, a constraint was placed on the optimisation to prevent T_D from exceeding unity.

The graphs were obtained by numerical integration of the expression for error power spectral density, a method which enabled results to be obtained in the unstable region. As can be seen, if T_D is limited to unity the optimum settings are

$$K = 3.5, \quad K_I = 0.4, \quad T_D = 1.0$$

these settings producing a ratio of error power/input power of 0.15. In the simulation results that follow it has been assumed for convenience that the input spectrum has been identified satisfactorily. This, in practice, would correspond to the situation where the input spectrum remains unchanged, although investigation indicates that the optimum parameter settings are not particularly sensitive to the precise input spectrum.

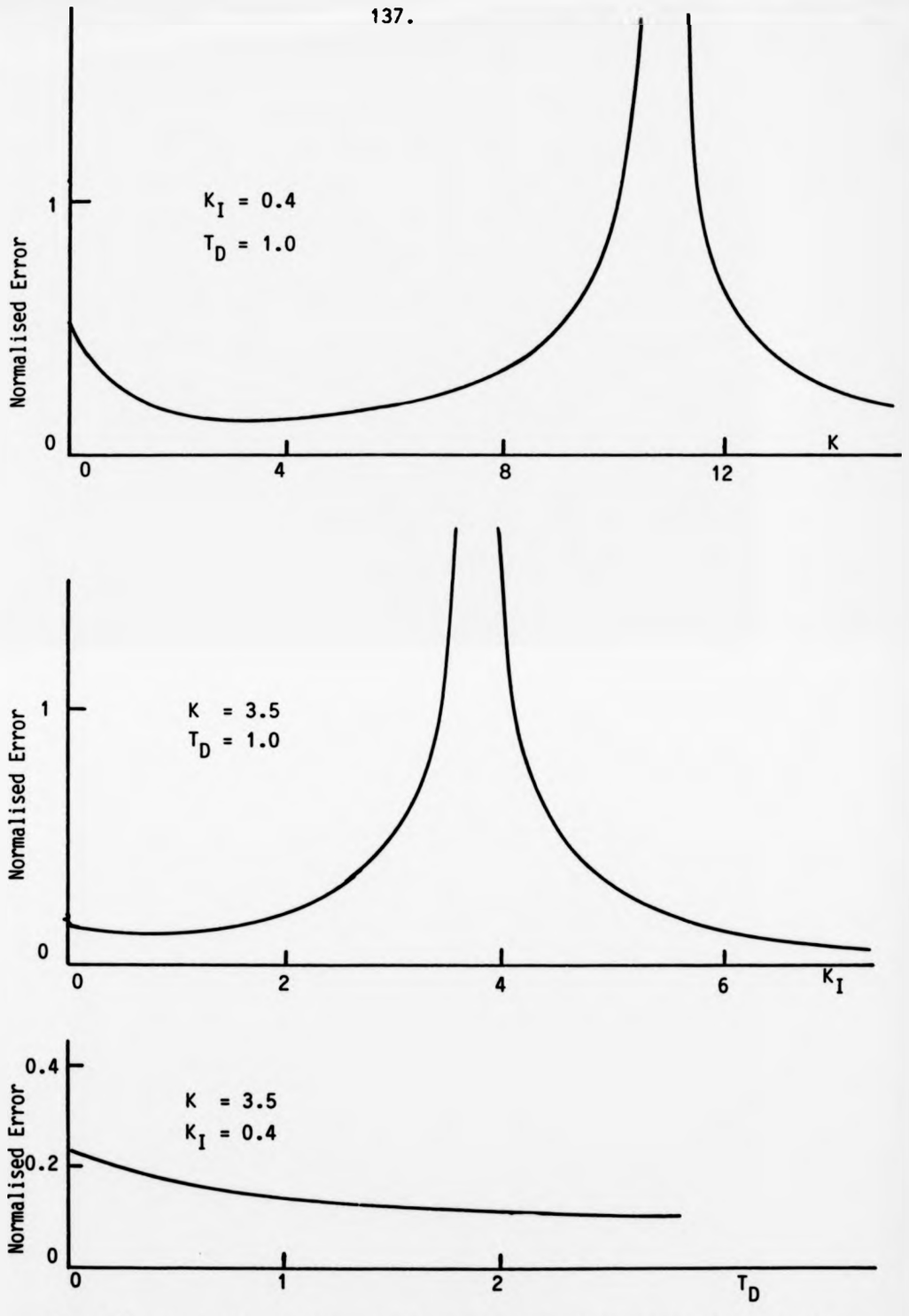


Fig. 6.2 Variation of error power with controller parameters - system $1/(1 + s)^3$

6.2.1 Non-parametric identification

As described in the previous chapter, the estimator used for non-parametric identification was

$$\hat{G} = \frac{\sum R^{*}C}{\sum R^{*}U}$$

The data blocks corresponding to R, U and C were modified by use of a Hanning window before taking the Fourier Transforms and experiments were carried out using block lengths ranging from 6.4 to 51.2 seconds. The shorter block lengths produce significant bias in the estimates and an approximation to this bias can be obtained by consideration of the forward path transfer function.

$$G(s) = \frac{1}{(1 + s)^3}$$

From chapter 2 the biased estimate $\overline{G}(j\omega)$ is given by

$$\overline{G}(j\omega) = G(j\omega) + \frac{1}{T}\frac{dG(s)}{ds}\Big|_{s = j\omega}$$

$$= \frac{1}{(1 + j\omega)^3} - \frac{1}{T} \cdot \frac{3}{(1 + j\omega)^4}$$

Fig. 6.3 shows plots of $G(j\omega)$ and $G(j\omega)$ for T = 6.4 and T = 51.2 seconds. As can be seen, there is significant bias for T = 6.4 seconds. Because of this bias, and because there are so few points to make significant contribution to the calculation of error power, one would expect poor results for this block length.

Initially the simulation results were obtained for the noise free case and the initial controller settings were,

$$K = 0.02, \quad K_I = 0.0, \quad T_D = 0.0$$

the system then being allowed to adapt from these settings. Results showing the ensuing variation in controller settings with

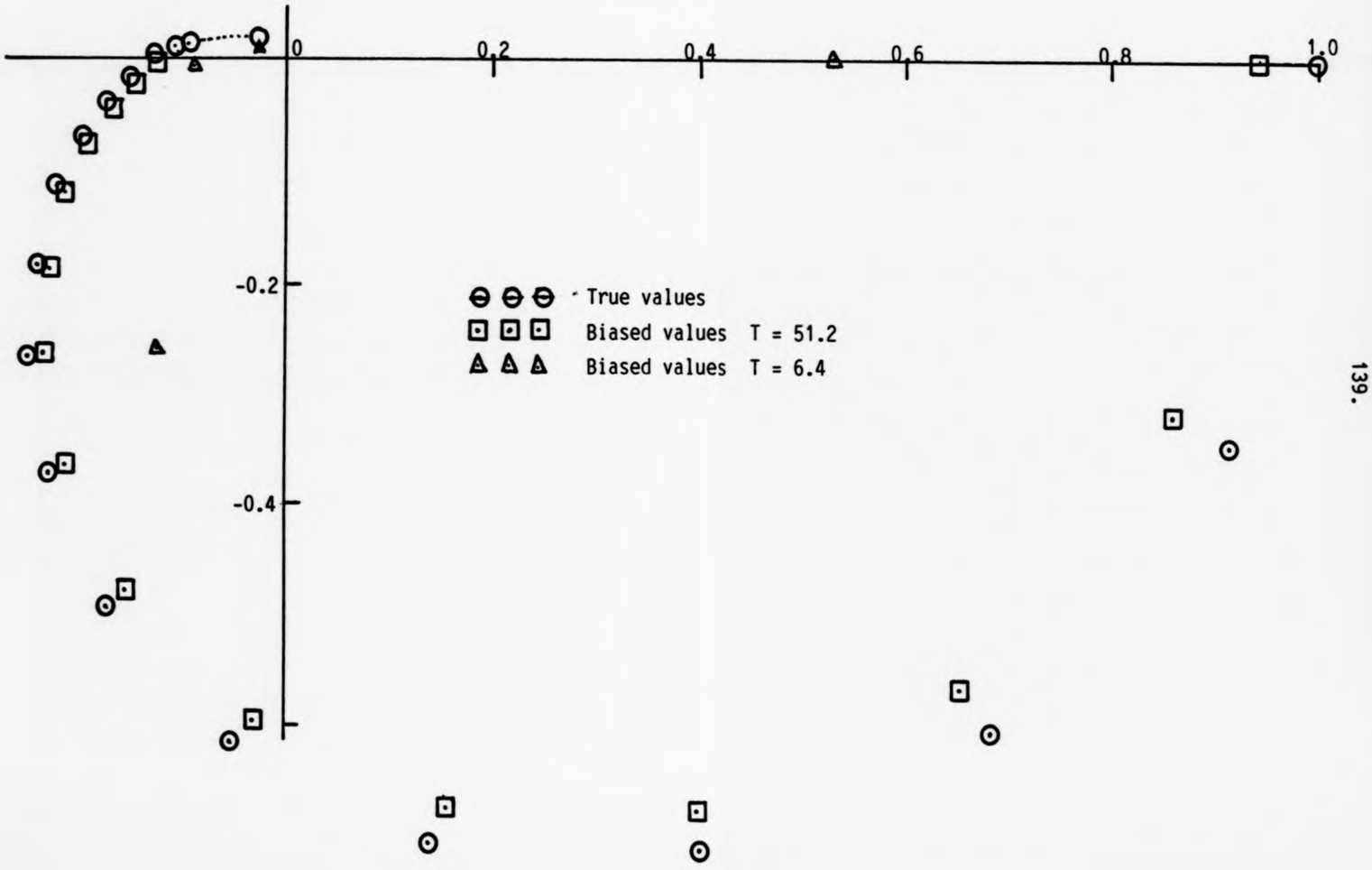
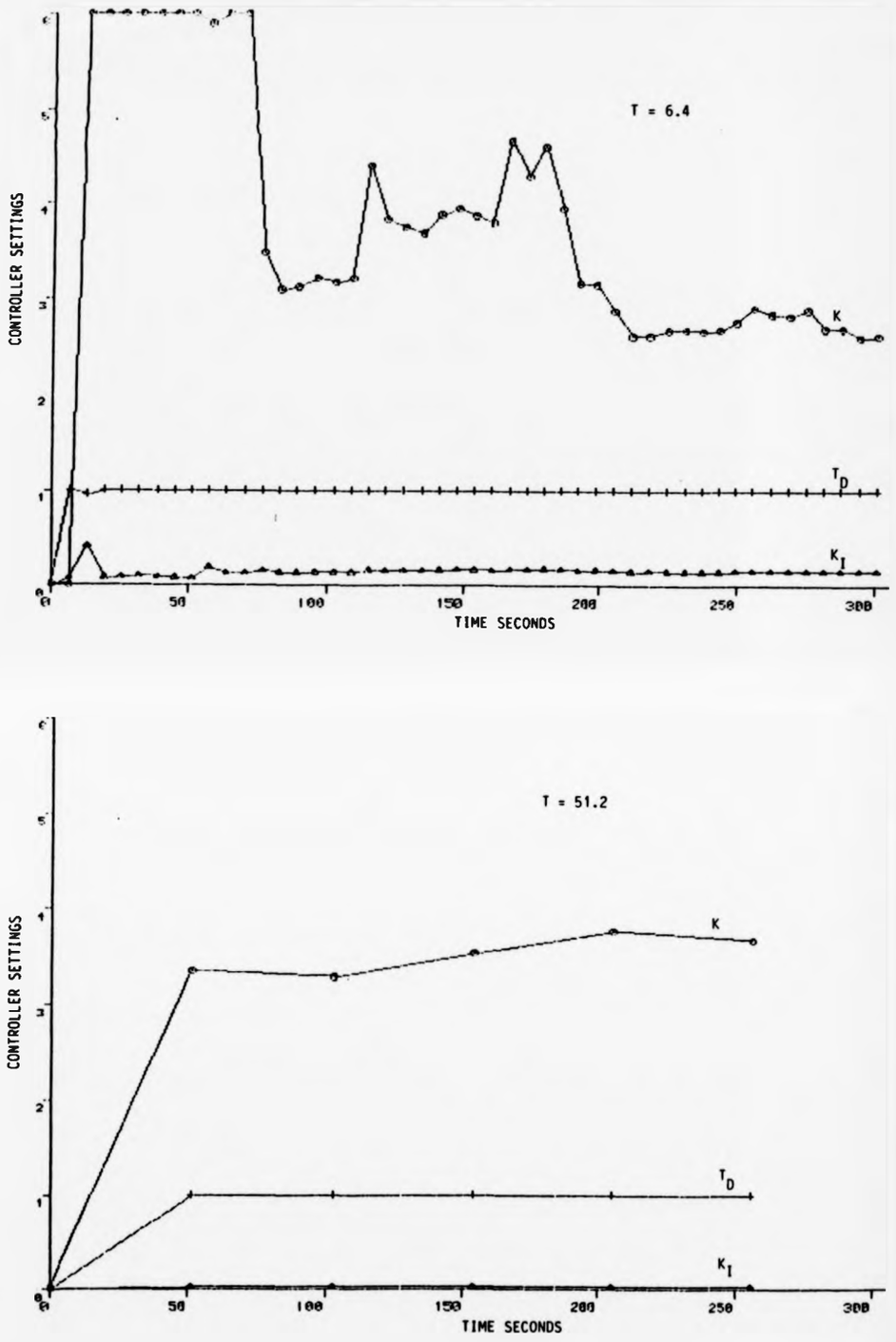


Fig. 6.3 Bias for system $1/(1 + s)^3$

time are shown in Fig. 6.4; these are given for block lengths of 6.4 seconds and 51.2 seconds. As can be seen, the controller settings converge towards a constant value after about 250 seconds (5 blocks of length 51.2 seconds). The final settings obtained agree with those predicted from Fig. 6.2 in the case of a block length of 51.2 seconds, but the final gain setting is not the predicted minimum when the block length was 6.4 seconds. However, referring to Fig. 6.2, this final setting would not cause significant increase in the error power due to the shallow minimum exhibited by the curve.

The variation of controller parameters with time have to be related to error power via the curves plotted in Fig. 6.2. This disadvantage can be overcome by portraying the variation in error power with time as the parameters adapt to their optimum values. Because of the random input signal, such curves would show a statistical variation; hence an experiment consisting of allowing the system to converge from initial controller settings was repeated a large number of times (200). Ensemble averages of cumulative error power against time are shown in Fig. 6.5; here the error power has been normalised to the mean input power for a block length T = 52.6 seconds. These results were obtained by direct measurement of the time error signal in the simulation.

Fig. 6.5 shows results both for the no noise situation and for the case where there was noise present. In the latter case the noise power/input power ratio was 0.25:1. This, due to the low initial gain setting, represents an initial figure of noise/signal power at the output of 5:1 and a ratio of 0.3:1 at the optimum setting. Fig. 6.5 shows that the shorter block length gives lower values of cumulative error despite the greater bias and variance associated with such short lengths. The slopes of the curves



6.4 Graphs showing convergence of controller settings with time.

System $G(s) = 1/(1 + s)^3$. Non-parametric estimation. Convergence is shown under no noise conditions for block lengths T = 6.4 and T = 51.2 seconds.

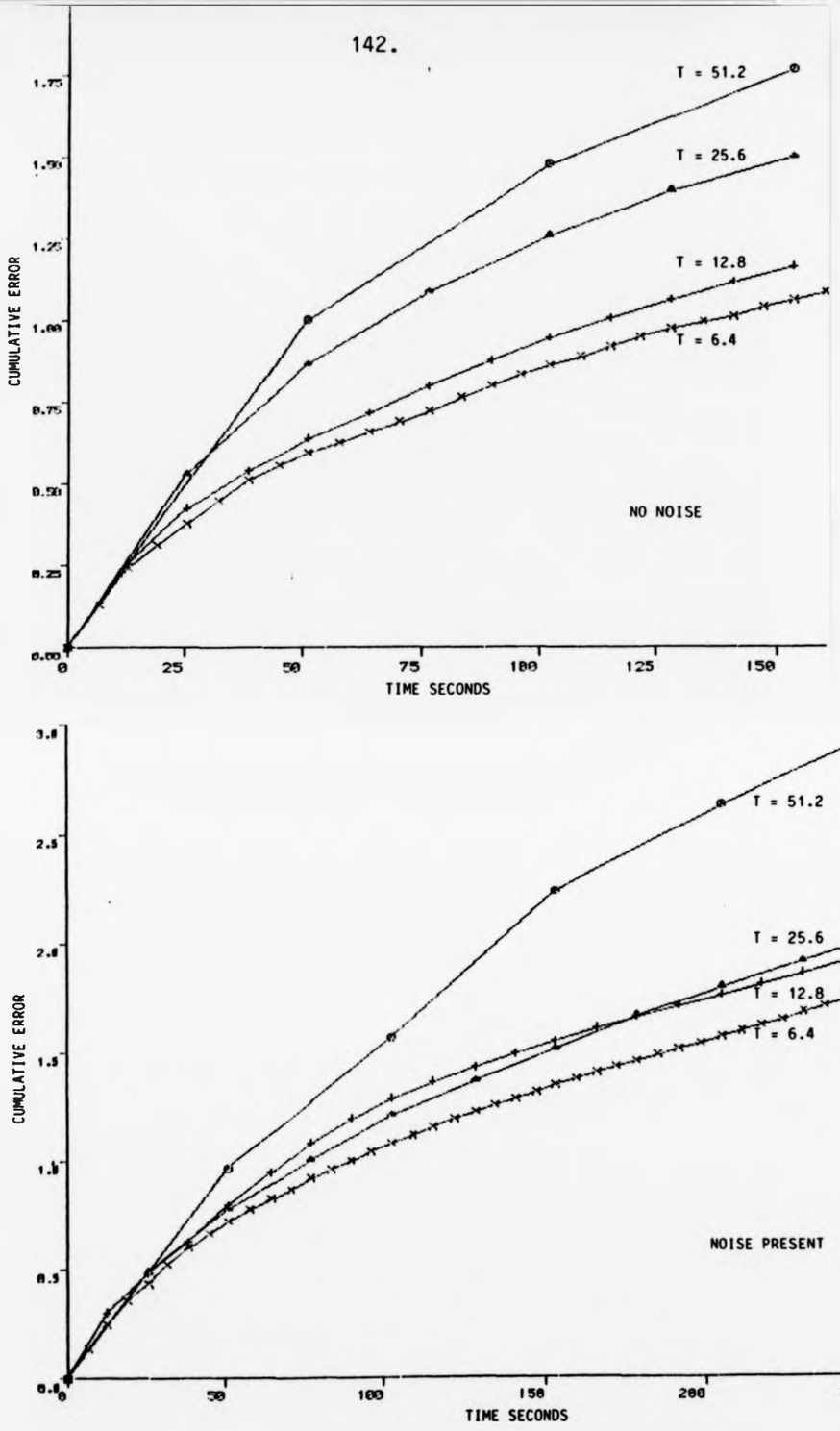


Fig. 6.5 Graphs showing cumulative error/time

System $G(s) = 1/(1 + s)^3$. Non-parametric identification. Graphs are shown with and without noise present.

converge to a value of 0.17 for the no noise case and to a value of 0.21 when noise was present; these compare favourably with theoretical figures of 0.15 and 0.19 respectively.

6.2.2 Parametric Identification

The same system, under the same conditions, was used with parametric identification techniques and the results are shown in Fig. 6.6. The results with no noise present show that adaptation is possible in one data block, even for a block length of 6.4 seconds. With noise present adaptation takes longer, but the results obtained are superior to the non-parametric case. However, these results were obtained under the assumption that the form and order of the parametric model were correct.

6.3 Minimisation of Error Power - System $e^{-s}/(1 + 5s)$

The overall system was as shown in Fig. 6.1, but the third order transfer function was replaced by

$$G(s) = \frac{e^{-s}}{(1 + 5s)}$$

The input and noise spectra were identical to the system in section 6.2.

Theoretical plots of normalised error power/controller parameters are shown in Fig. 6.7, these plots being similar in shape to those obtained for the previous system.

For non-parametric identification the finite block length again causes bias in the estimates. It is instructive to consider the bias in a system consisting of a pure time delay only.

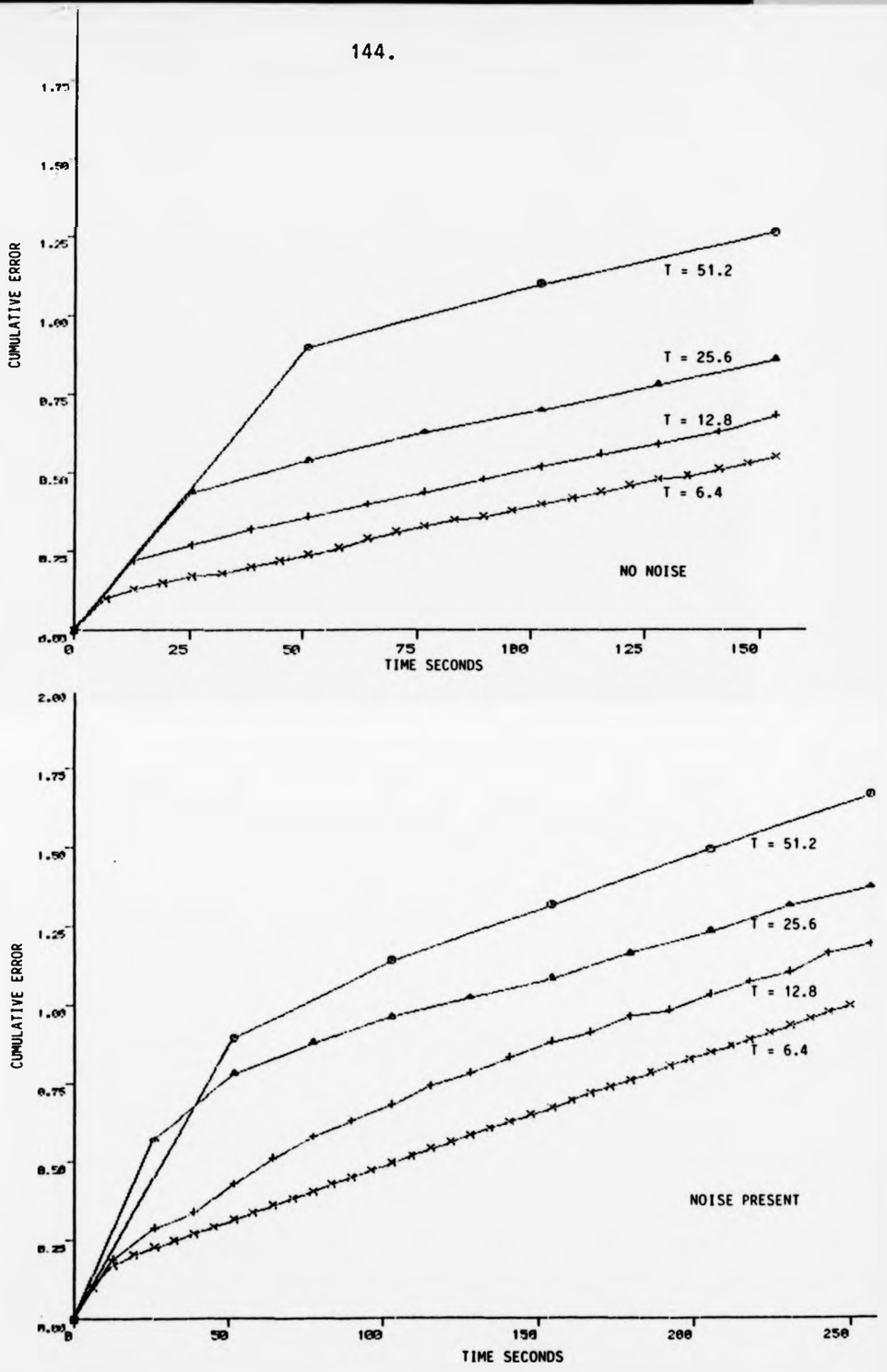


Fig. 6.6 Graphs showing cumulative error/time.

System $G(s) = 1/(1 + s)^3$. Parametric identification. Graphs are shown with and without noise present.

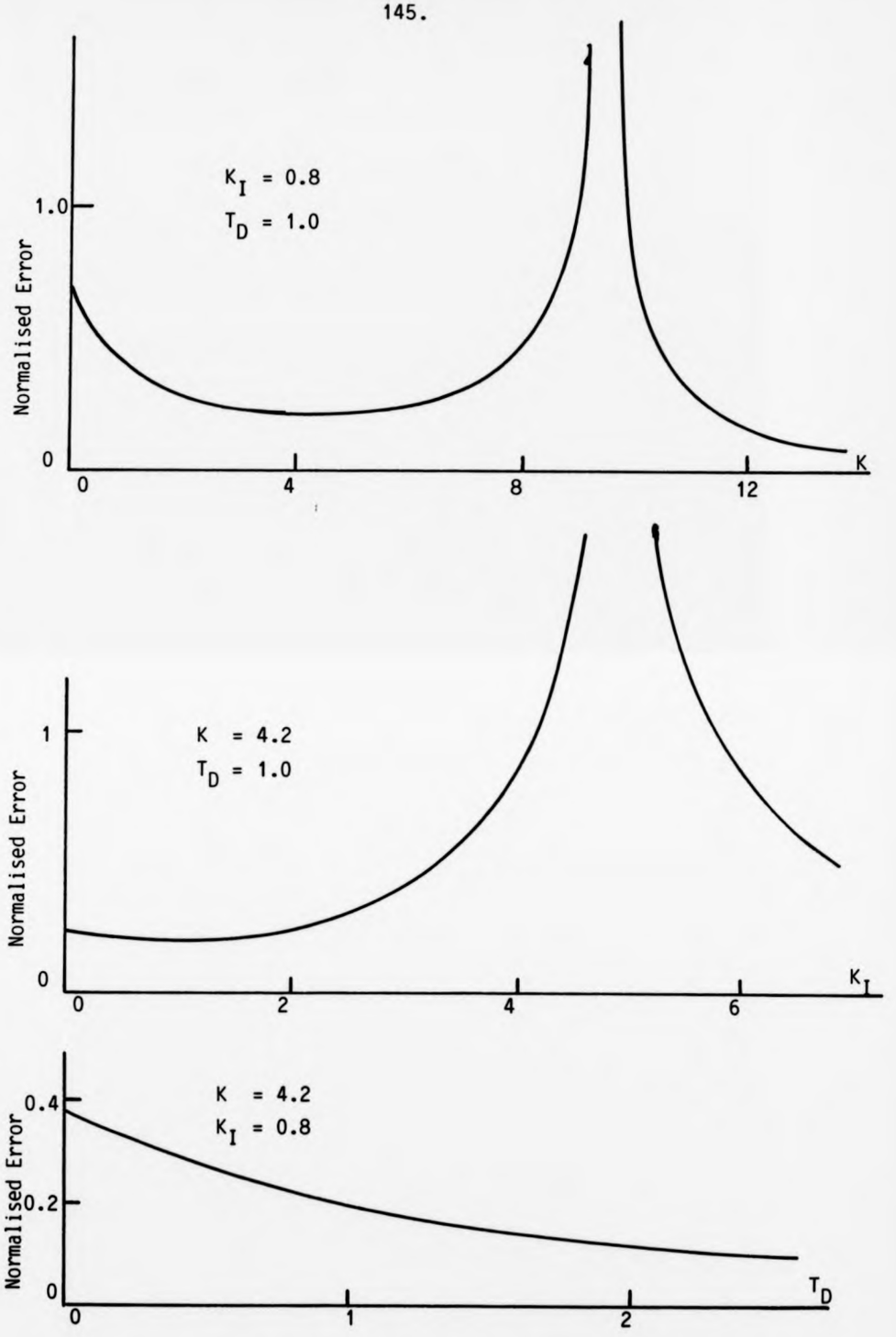


Fig. 6.7 Variation of error power with controller parameters - system $e^{-s}/(1 + 5s)$

$$G(s) = e^{-sT_1}$$

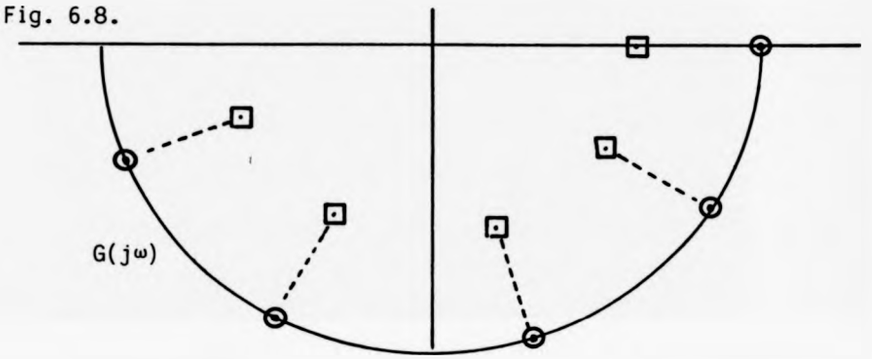
then

$$G(j\omega) = G(j\omega) + \frac{1}{T}\frac{dG(s)}{ds}\Big|_{s = j\omega}$$

giving

$$\bar{G}(j\omega) = (1 - \frac{T_1}{T})e^{-j\omega T_1}$$

This result can be interpreted graphically as shown in Fig. 6.8.



⊙ ⊙ ⊙ True points ⊡ ⊡ ⊡ Biased points – – – – Radial lines

Fig. 6.8 Bias for system $G(s) = e^{-sT_1}$

The plot of $G(j\omega)$ is a circle and the finite block length causes bias in the estimate by moving the true point radially towards the centre, a distance T_1/T. When $T = T_1$ this would give a mean estimate, at all frequencies, of $\bar{G}(j\omega) = 0$. This corresponds to the situation where there is a time delay of 1 block between input and output records. As explained in chapter 1, this case calls for a realignment of the records before taking the Fourier Transforms. Hence the results obtained used a shift in the records of 1 second and a Hanning window was used on the data records corresponding to R, U and C.

The system was allowed to adapt following initial controller settings $K = 0.02$, $K_I = 0.0$, $T_D = 0.0$ and plots of cumulative error power against time are shown in Fig. 6.9. Again, results are as predicted from Fig. 6.7 both for the case with no

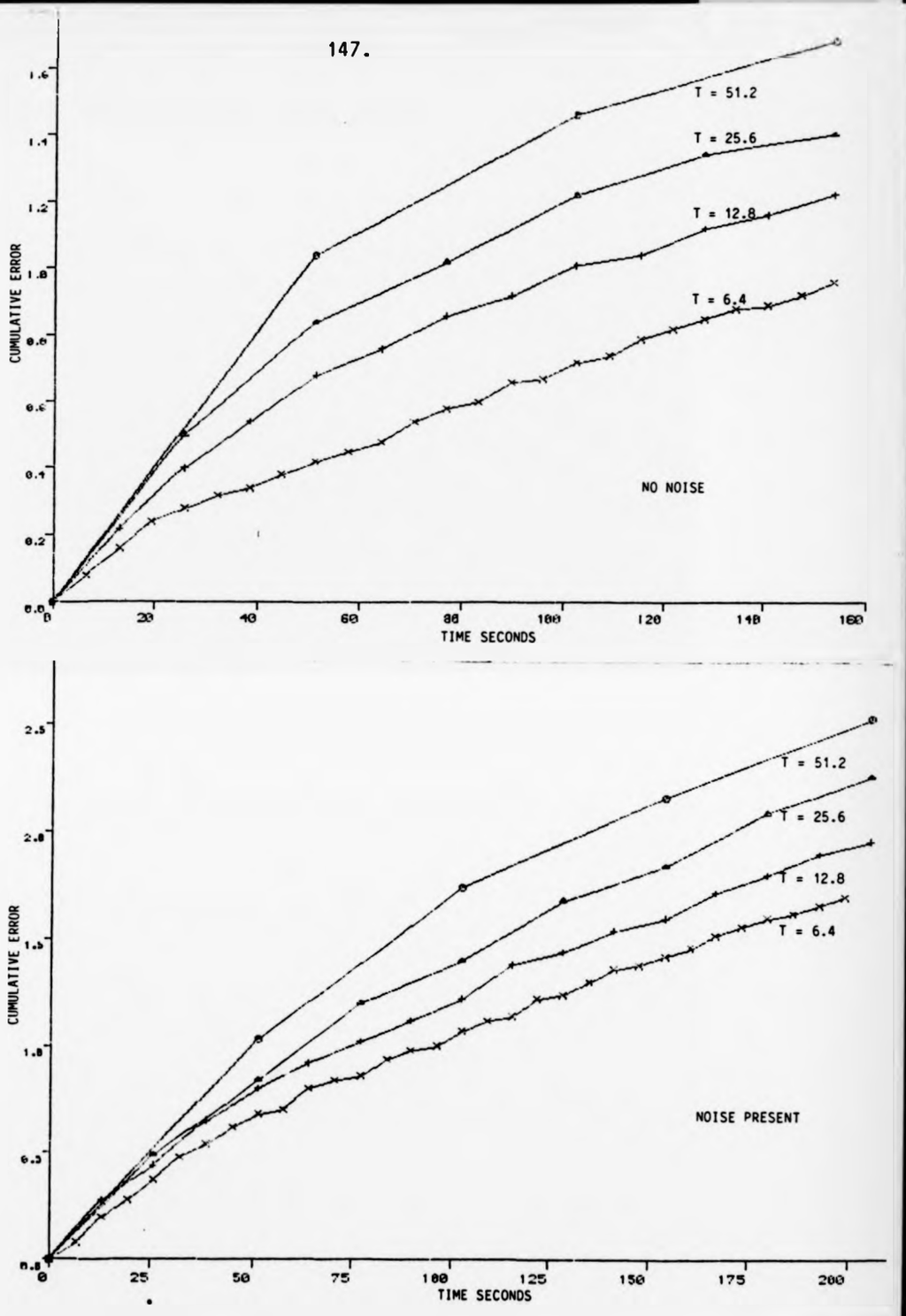


Fig. 6.9 Graphs showing cumulative error/time.

System $G(s) = e^{-s}/(1 + 5s)$. Non-parametric identification. Graphs are shown with and without noise present.

noise and when there is noise present.

No results have been included for parametric identification. The method was investigated qualitatively and appeared to give results similar to those obtained for the system in section 6.2.2. However, again the form of the model was assumed to be correct (first order) and the magnitude of the time delay was taken into account by re-alignment of the records.

6.4 Specification of Closed-Loop Performance - System $1/(1 + s)^3$

The adaptation of the system having forward path transfer function $1/(1 + s)^3$ was investigated using the principles described in section 5.4. The performance criteria was the integral of error squared between predicted and prescribed closed-loop frequency response.

The prescribed response was as follows: magnitude of closed-loop frequency response constant up to frequency ω_c and zero for frequencies beyond that. Results are given in Fig. 6.10 for four values of ω_c. These results are plotted on linear scales so that the $\omega = 0$ point can be included and to emphasise the equal weighting given to all frequency points. The results show the frequency points on the required response and the actual response achieved after the controller settings have stabilised; these settings are quoted on the plots. The open-loop response is also included on each plot.

Fig. 6.11 shows results for the same system, but the required closed-loop response is now specified in a more practical manner.

The response is second order.

The magnitude of the gain at $\omega = 0$ is unity.

The M_m and ω_m values are specified.

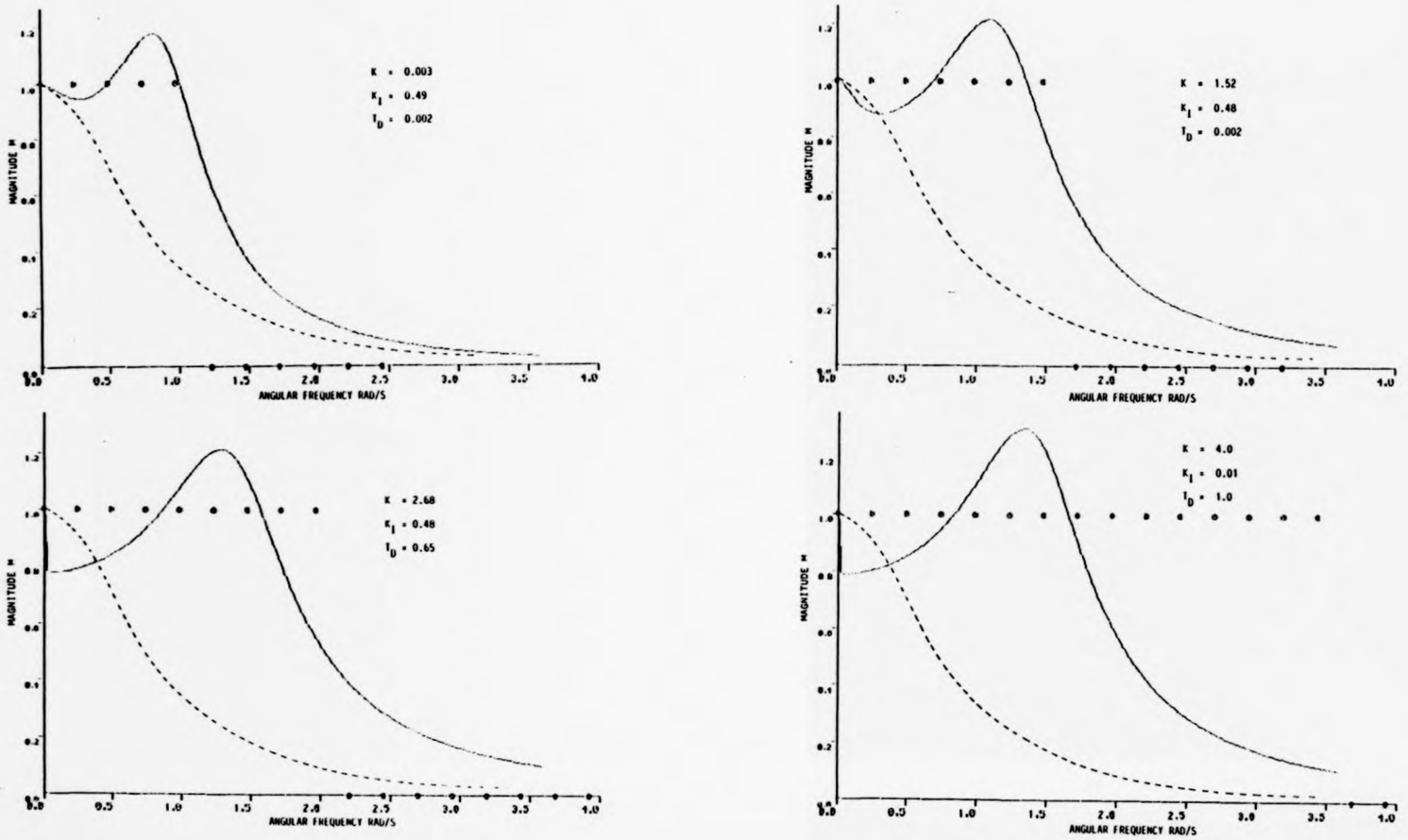


Fig. 6.10 Frequency response - System $G(s) = 1/(1 + s)^3$

⊙⊙⊙ Prescribed closed-loop response. ——— Closed-loop response obtained. - - - - - Open-loop response of system. Final controller settings are shown.

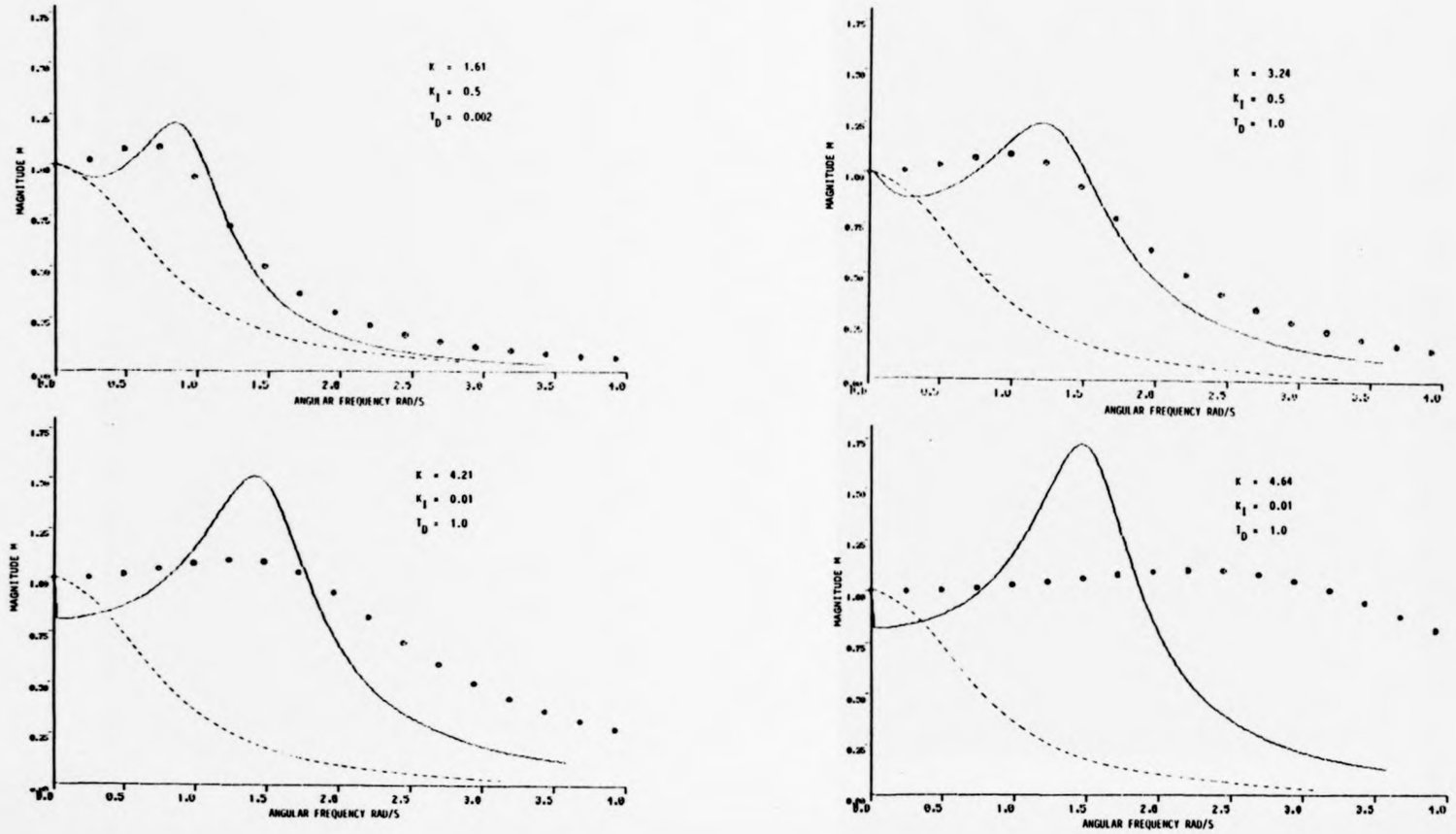


Fig. 6.11 Frequency response - System $G(s) = 1/(1 + s)^3$

⊙⊙⊙ Prescribed closed-loop response. ——— Closed-loop response obtained. - - - - - Open-loop response of system.
Final controller settings are shown.

The M_m value was specified as 1.1 and Fig. 6.11 shows the responses obtained for four values of ω_m. Again, the open-loop response has been included for comparison purposes.

Although the required closed-loop response has been specified by two different methods, the controller settings produced by both methods follow a similar pattern. When the required bandwidth is small the controller produces a high integral gain which gives small low frequency error at the expense of bandwidth. As the required bandwidth is increased the controller reduces the integral term but increases the differential term until this is limited by the constraints built into the controller. None of the responses obtained are particularly satisfactory with respect to the M_m values produced, these in all cases being higher than the specified values.

These results, although encouraging, can only be regarded as a preliminary investigation into the method. Improved results could probably be obtained by investigation on the following lines.

1. Only specifying the required response in the pass band and allowing the system to adapt to an arbitrary response outside this band.
2. Not giving equal weighting to the error at all frequency points. This would enable the controller to adapt to the natural logarithmic form of the response. It would also enable a balance to be obtained between steady state and transient requirements.
3. Weighting the error by means of the measured input or required error spectrum.

CHAPTER 7

Conclusions

This chapter draws conclusions based on results derived in the thesis and makes some suggestions for possible areas of future investigation.

Chapter 1 reviews work already done in the area of short term spectral analysis. From work published in this field, it is apparent that spectral analysis is a very important technique for the investigation of a wide range of systems. These include not only systems in engineering but in a variety of fields, e.g. economic, biomedical, geophysical. The problem of obtaining spectra from short record lengths has arisen in many diverse applications and there have been many approaches to the solution of the problem. In the area of frequency response estimation the methods of approach have been more restricted. This, to some extent, is due to the requirement that many applications in the control field need real time on line identification calling for very efficient computing algorithms. However, there are techniques, that have been outlined in chapter 1, such as use of combined linear and quadratic weighting and the use of the WOSA and STUSE algorithms, that do appear to offer improved estimates and are also computationally efficient. These methods do not appear to have been investigated from the viewpoint of identification in the control field and such investigation could form the basis of further work.

Chapter 2 introduces some of the major new ideas that are used throughout the thesis. The conventional method of considering errors in frequency response estimates due to finite record length is via the errors that occur in the estimation of

the auto and cross spectra. However, the approach introduced here is to attribute the errors to transient terms that occur at the start and end of the data block. Qualitative investigation of the transient terms gives an intuitive approach to the manner in which these terms cause bias and variance in the estimates and suggests methods by which these effects may be reduced. Quantitative investigation of the terms enables them to be split into a part correlated with the input and an uncorrelated term. It is shown that the correlated term leads to bias in the estimate and this is independent of the number of blocks used. Two methods of evaluating the bias are developed. One method is general and will apply for any input spectrum and with arbitrary data windows used; the other is more restrictive and applies to white noise input only, together with rectangular windows. The latter method, however, has the advantage of having a simple physical interpretation.

It is shown that the uncorrelated portion of the transient can be regarded as an equivalent external noise source. This is an important result as it enables the effect of finite record length to be analysed by similar methods to those used to investigate the effects of external uncorrelated noise.

The example used to illustrate the results obtained in chapter 2 shows that considerable labour is involved in the calculation of the bias and variance, even for a first order system. The calculations for higher order systems would, in general, be too complex to be feasible. A program to calculate the bias and variance for the general system is now available and future work could develop this into a sophisticated computer aided design package. One use of such a package would be to aid

in the determination of the optimum form of windowing for use in frequency response estimation. In chapter 2, it is suggested by qualitative reasoning that non-symmetrical windows which would be different at system output and input may give better results than the more conventional windows. A full quantitative investigation would be required to justify such an approach.

The analysis presented in chapter 2 applies to data taken from independent blocks. Further investigation could consider the effects on the estimates of non-independent and of overlapping blocks.

Chapter 3 extends the ideas introduced in chapter 2 to include frequency response estimation under closed-loop conditions. The effects of the transients are now more complex than in the open-loop configuration. It has been shown elsewhere [Douce (1)] that external noise, uncorrelated with the input, causes bias under closed-loop conditions. For this reason, it is not possible to split the transient into correlated and uncorrelated components, as is done in the open-loop analysis. However, an expression for bias is derived for this situation, but, unlike the open-loop bias, it does depend upon the number of data segments used in the estimate. For a large number of segments it is shown that the bias approaches that obtained under open-loop conditions. No expression for the variance of the estimates is obtained; however, it is shown that confidence intervals for these estimates can be constructed by transformation of the open-loop confidence intervals.

The analysis presented for the closed-loop system is restricted to unity feedback systems only. If the feedback is not unity, additional transients appear due to the non-zero initial states in the feedback transfer function. Although these

can be regarded as an additional equivalent noise source in the feedback path, there is correlation between this and the equivalent noise source in the forward path. Further investigation in this area could prove useful.

Chapter 4 considers parametric methods of identification in the frequency domain. The importance of this approach in the present context is that it enables the transient terms to be incorporated into a parametric model of the system. Using such an approach, it is shown that, in the noise free case, it is possible to eliminate the effects of finite observation time completely. This result does rely on the important assumption that the correct form and order for the mode has been chosen in the parametric representation of the system.

The method of obtaining the parameters in the model follows methods similar to those used for parametric identification in the time domain, i.e. regression analysis, recursive methods and use of extended least squares. A problem that arises in the frequency domain, that is not generally present in the time domain, is the abrupt change in certain parametric values (the ones introduced to describe the transient terms) at the end of each data block. Methods are suggested to overcome this problem and simulation results appear encouraging. However, it does appear that the parameter in the frequency domain model cannot be obtained in as short a record length as those in the corresponding time domain model. There is scope for future work in this area.

One disadvantage of the parametric method compared to the non-parametric is that assumptions have to be made regarding the form and order of the system model. Another disadvantage is that the algorithm used in the parametric method cannot approach

the computational efficiency of the Fast Fourier Transform algorithm. Future work could investigate the modification of efficient recursive algorithms developed for auto spectral estimation.

Chapter 5 considers the application of frequency domain methods to the field of adaptive control. A review of such methods shows that recent advances in adaptive control have favoured time domain methods, especially with the advent of self-tuning controllers. Possible advantages of using the frequency domain are, (i) it may be non-parametric method where few prior assumptions need be made regarding the system, (ii) the performance index can be based on frequency domain criteria which are widely used for design purposes. Possible adaptive schemes are suggested in chapter 5 and these are investigated using simulated systems in chapter 6.

All the methods investigated produced adaptive schemes that gave controller settings that minimised their respective performance indices. Under no noise conditions methods using parametric identification gave most rapid adaption. Unfortunately such methods nullify one of the main advantages of using frequency response methods, i.e. little prior knowledge is required of the system.

When noise was added there was no clear advantage to non-parametric methods. Although no direct comparisons were made with time domain methods, it does appear that adaptation in the time domain would be significantly faster. Hence, in the use of adaptive systems for tracking time varying parameters or for systems having a time varying input spectrum, there would appear to be no advantage in using frequency domain methods.

A field that is receiving current attention is that of "auto-tuning" and the determination of the initial controller settings for a system. It is felt that frequency domain methods of providing self-tuning for these controllers could be of importance in this area. This method provides the ability to specify the required system response in the frequency domain in terms of performance parameters widely used and accepted in control engineering. Flexibility is available in weighting the response in relation to the input or noise spectrum. Simulation results using such methods are encouraging, but it is felt the area is open to much wider investigation.

Throughout this thesis the systems considered have been single input single output systems. Many of the ideas put forward could be extended to multi-input multi-output systems and this could provide an area for future research.

Reference

1. Douce, J.L., "Bias of frequency-response estimates in closed-loop systems", Proc. IEE, Vol.127, Pt.D, No.4, pp.149-152, July 1980.

APPENDIX 1

A Numerical Method for applying Nyquist's Stability Criterion

As described in section 5.5, a method is required to determine whether the contour $G(j\omega)$, as ω varies $0 \rightarrow \infty$ (plus its conjugate) encloses the $-1 + j0$ point for the case where $G(j\omega)$ is only available as a set of discrete points.

To do this a fundamental theorem of complex integration is used

$$\oint_C f(z)dz = 2\pi j \sum \text{(residues within } c\text{)}$$

Hence a function $f(z)$ is chosen that has a residue at $-1 + j0$ and this function is evaluated at the points on the Nyquist Plot. Approximating the integration by a summation, this will give a result that is only zero if there is no nett rotation about the -1 point. This is illustrated in Fig. A1.1.

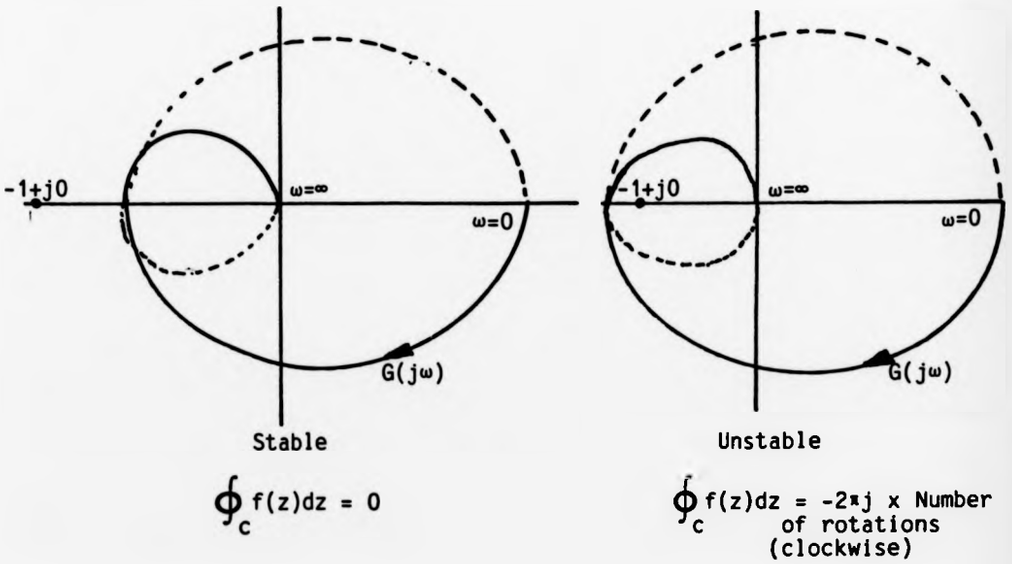


Fig. A1.1 Nyquist's stability criterion

Any function that has a residue at z = -1 + j0 could be chosen for the function f(z). The simplest is

$$f(z) = \frac{1}{z - z_1}$$

where z_1, the pole of the function, is -1 + j0. There is only one residue to the function which is at this point and has value 1.

The function f(z) can only be evaluated at a discrete number of points, hence the contour integration must be approximated by a summation. Consider two adjacent points on the plot (x_1, y_1) and (x_2, y_2) as shown in Fig. A1.2.

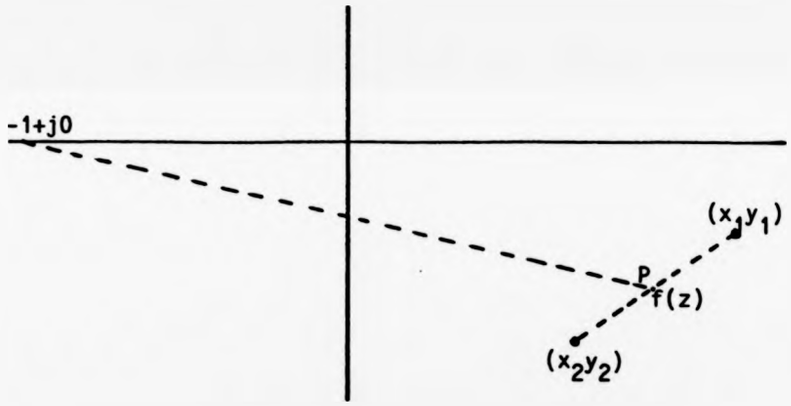


Fig. A1.2 Approximation to contour integration

The function f(z) is evaluated at point P, the arithmetic mean of the points (x_1, y_1), (x_2, y_2)

$$P = x + jy = \frac{(x_1 + x_2)}{2} + j\frac{(y_1 + y_2)}{2}$$

and the increment dz is taken as

$$dz = dx + jdy = (x_2 - x_1) + j(y_2 - y_1)$$

Writing f(z) in terms of its real and imaginary parts and as functions of x, y then

$$f(z)dz = (f_1(x,y) + jf_2(x,y))(dx + jdy)$$

$$= (f_1(x,y)dx - f_2(x,y)dy) + j(f_2(x,y)dx + f_1(x,y)dy)$$

The criterion for stability is on the imaginary part of this function.

Stable $\sum (f_2(x,y)dx + f_1(x,y)dy) = 0$

Unstable $\sum (f_2(x,y)dx + f_1(x,y)dy) = -2\pi$ (or multiple)

Writing

$$f(x,y) = \frac{1}{(x + jy) - (-1 + j0)}$$

then

$$f_1(x,y) = \frac{x + 1}{(x + 1)^2 + y^2}, \quad f_2(x,y) = \frac{-y}{(x + 1)^2 + y^2}$$

giving at point $x = \frac{x_1 + x_2}{2}$, $y = \frac{y_1 + y_2}{2}$

$$f_1(x,y) = \frac{2(x_1 + x_2 + 2)}{(x_1 + x_2 + 2)^2 + (y_1 + y_2)^2}$$

$$f_2(x,y) = \frac{-2(y_1 + y_2)}{(x_1 + x_2 + 2)^2 + (y_1 + y_2)^2}$$

Hence, the imaginary part of f(z)dz is given by

$$\frac{2[(x_2 + x_1 + 2)(y_2 - y_1) - (y_2 + y_1)(x_2 - x_1)]}{(x_1 + x_2 + 2)^2 + (y_2 + y_1)^2}$$

and

$$\oint_C f(z)dz = 2 \sum_{n=0}^{N} \frac{(x_{n+1} + x_n + 2)(y_{n+1} - y_n) - (y_{n+1} + y_n)(x_{n+1} - x_n)}{(x_{n+1} + x_n + 2)^2 + (y_{n+1} + y_n)^2} \quad \text{.....(A1.1)}$$

where N is the number of points. The point $(x_{N+1} + jy_{N+1})$ is taken as the conjugate of $(x_N + jy_N)$, i.e. $x_{N+1} = x_N$, $y_{N+1} = -y_N$. As the conjugate portion is not summed, instability is represented by $-\pi$ for one encirclement.

If the function G(s) is known to have a pole at the origin (as in the case of integral control) the contour must be modified. This is done by adding a circle of infinite radius to complete the contour from the point $\omega = 0^-$ to the point $\omega = 0^+$. It can easily be shown that the contribution of this portion of the contour to the total integral is $-\pi$ and the criterion must be modified accordingly.

It should be noted that the test described in this section tests the simplified form of Nyquist's criterion, i.e. it is assumed that the function G(s) has no poles in the right-hand half plane.

Examples

To illustrate the method, the integral given by eqn. (A1.1) has been evaluated around the contour defined by a given frequency response function. This has been done for a finite number of points and for a range of values of K. By plotting the integral against K, instability will be shown by rapid changes of its value (0 to 2π) at the value of K, indicating marginal stability.

The first example has forward path transfer function

$$G(s) = \frac{K}{(1 + s)^4}$$

When used with unity feedback this would give a system that is stable $0 < K < 8.0$. The integral determined by eqn.(A1.1) is evaluated over 30 points and plotted against K in fig. A1.3. (Note that points defining the complex conjugate have not been included, hence a value of $-\pi$ for the integral indicates instability). As can be seen, values of $K < 6.0$ are indicated as stable and values of $K > 8.6$ are indicated as unstable.

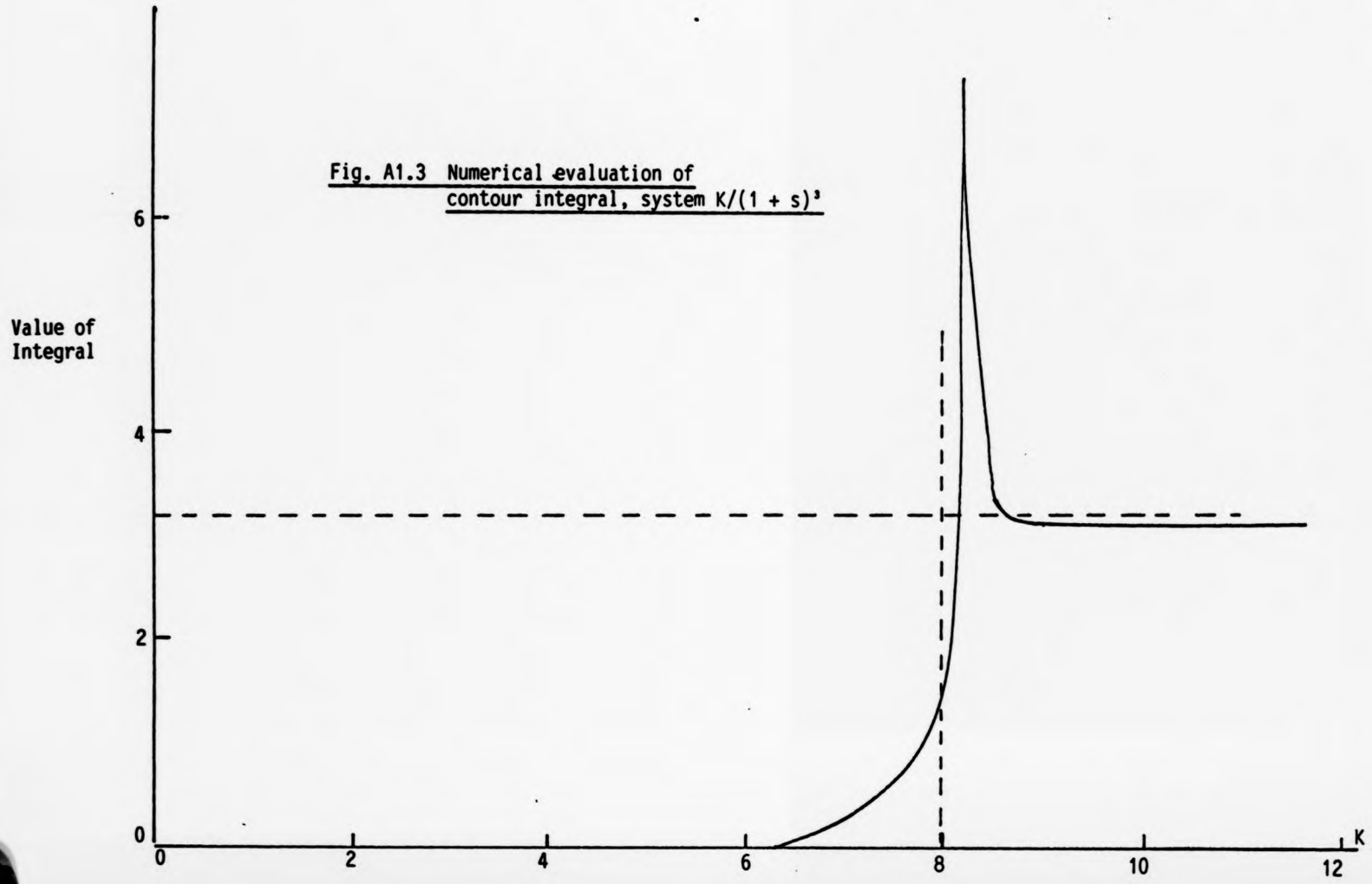


Fig. A1.3 Numerical evaluation of contour integral, system $K/(1 + s)^3$

Between these values lies a region of uncertainty and taking any value of the integral greater than unity as an indication of instability would give a satisfactory test.

The second example is a unity feedback system having forward path transfer function

$$G(s) = \frac{K(1 + 0.1s)^2}{s(1 + s)^2}$$

This is a type 1 system that exhibits conditional stability. The system is, stable

$$0 < K < 3.43$$

unstable

$$3.43 < K < 291.3$$

stable

$$291.3 < K < \infty$$

Fig. A1.4 shows the integral evaluated over 50 points and plotted against K. Note that the scale has been broken to give more definition to the areas of interest and the result given has been calculated for the complete contour (-2π indicating instability). Again, although regions of uncertainty exist, taking any value of the integral greater than unity as an indication of instability gives a satisfactory test.

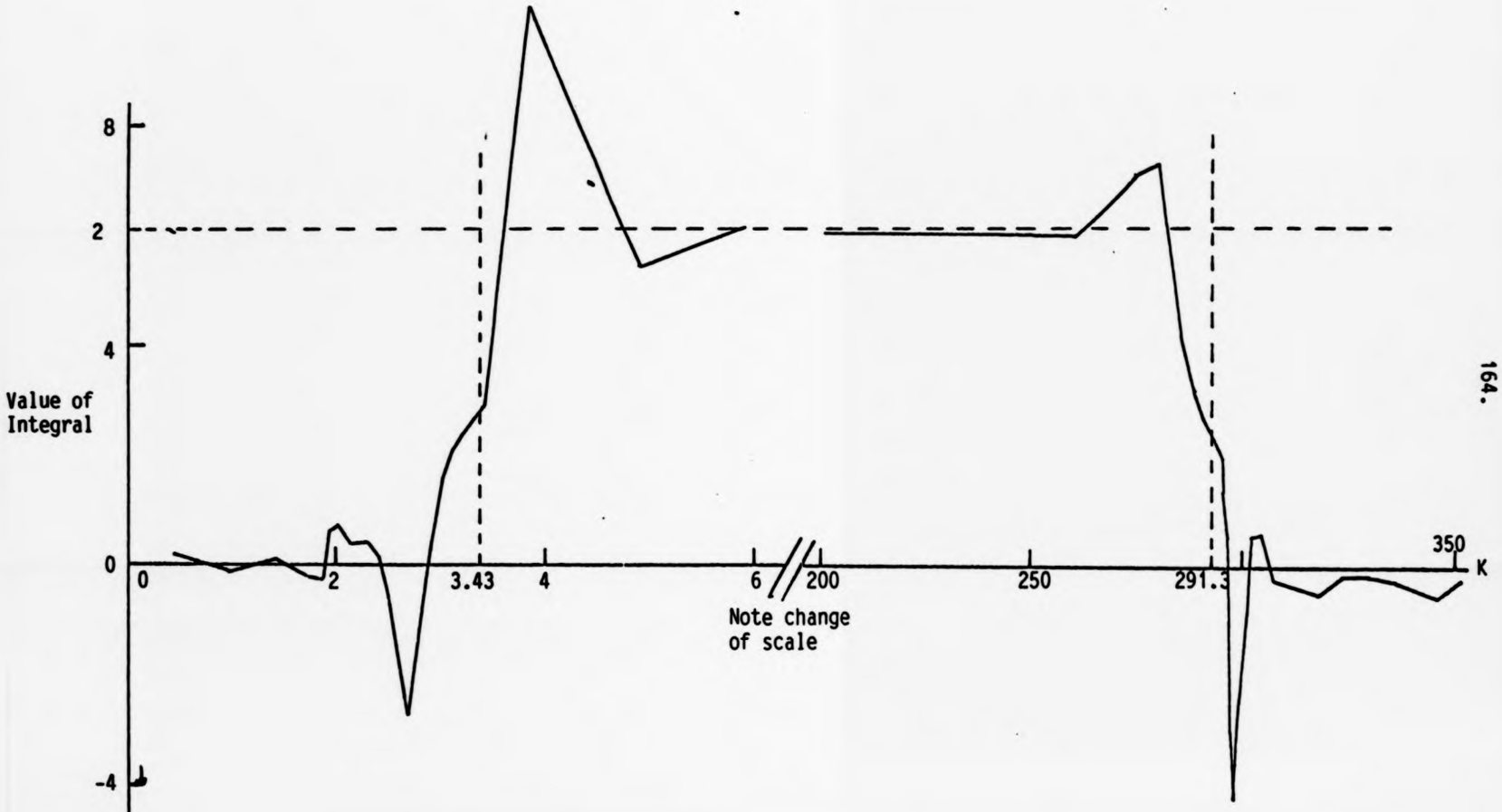


Fig. A1.4 Numerical evaluation of contour integral, system $K(1 + 0.1s)^2/s(1 + s)^2$

APPENDIX A2

Notes on the Computing Methods Used

The computer used for the work in this thesis was the SERC/Engineering Department, Prime Computer installation at the University of Warwick. The computer was used at Warwick University and by remote terminal at Coventry (Lanchester) Polytechnic. As many of the results were of a statistical nature involving the averaging of many simulation runs, extensive use was made of the "job facility", which enabled runs of up to 20 hours in length to be made overnight and at weekends. The program was written in Fortran 77 programming language.

A listing of the principal program developed to simulate the adaptive controller is given in this appendix. The program can be understood by the comments in this listing, the flowcharts of Figs. A2.1 and A2.2 and the notes that follow. The listing and Fig. A2.1 refer to non-parametric identification, Fig. A2.2 shows the modification to the flowchart for parametric identification and the additional subroutines required are also listed.

The representation of the system and controller in the program was by means of digital simulation. This was obtained by means of the transformation

$$s = \frac{2}{T_s} \frac{(1 - z^{-1})}{(1 + z^{-1})}$$

on the appropriate transfer function, T_s being the sampling time. The resulting expression in z was then interpreted as a difference equation and implemented digitally.

The following are brief notes on the various sections of the program.

MAIN PROGRAM Generates random data using NAG routine G05DDF. Filters this to form the input data and noise disturbance for the system. Links together the subroutines forming the remainder of the program.

SUBROUTINE WINDOW (N, WD) Produces the weighting WD(J), corresponding to a Hanning window, for N data points.

SUBROUTINE FT01A (IT, INV, TR, TI) This is a FFT routine to calculate the Fourier Transform of data having real and imaginary parts TR, TI. IT specifies the number of points in the transform and INV specifies whether the direct or inverse transform is required. The real and imaginary parts of the transform are returned as TR, TI.

SUBROUTINE DIF (N, NB, R, C, E, U, W, A, B, GN, GNI, TD, LFLAG) This subroutine simulates the system and controller. From the N data points forming the input signal R, and noise disturbance W, the subroutine generates output C, error E and control U. NB is the order of the system, A, B are the parameters of the system and GN, GNI and TD are the controller parameters. LFLAG initialises the system if this is required.

SUBROUTINE MINIM (BI, S) From an initial parameter vector BI this subroutine searches for the parameter values that give minimum cost function S. This subroutine calls the subroutines E( ), F( ).

SUBROUTINE E(B, H, N, W) This searches, from an initial parameter vector B, in N dimensions with step size H. The minimum value W found in the search is returned from the subroutine.

SUBROUTINE F(X, M, V, IND) This subroutine calculates the cost function V at parameter vector X (M dimensions). It uses values of $R(j\omega)$ and $G(j\omega)$ transferred via common statements. It checks for constraints on the parameters and on stability; any violation of these is indicated in the value of IND.

SUBROUTINE REGR (XC, NF) This subroutine identifies the parameters XC by regression in the frequency domain. The data used is passed to the routine via common statements. The subroutine makes use of the NAG routine FOICKF to perform matrix multiplication.

SUBROUTINE COLOSS (A, B, IERR, V, IN) This subroutine was developed by Astrom (see section 5.5). It calculates the loss function V for the ratio of two polynomials with coefficient vectors A, B. The value of IERR indicates if the denominator represents an unstable system.

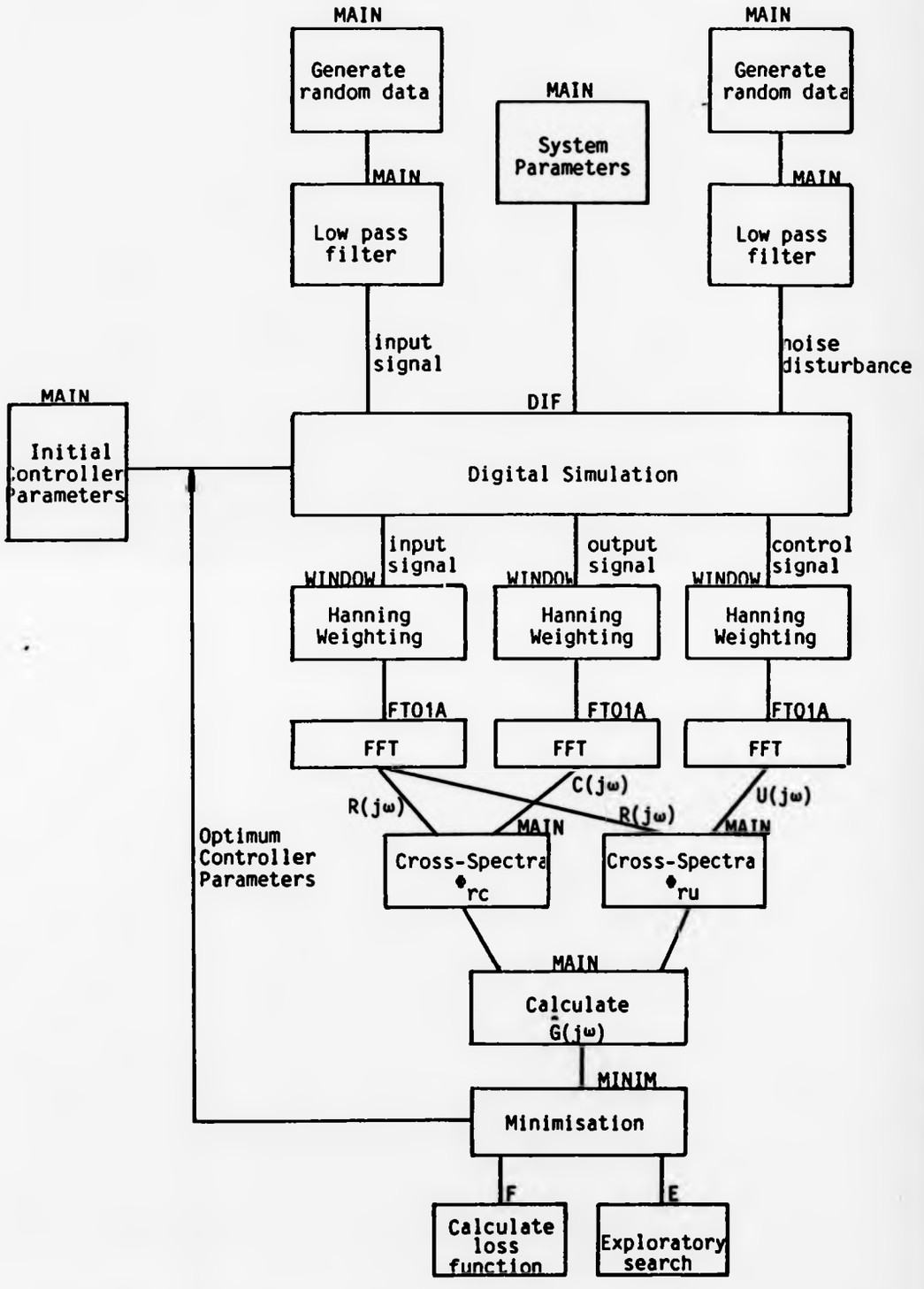


Fig. A2.1 Flow chart for non-parametric adaptation

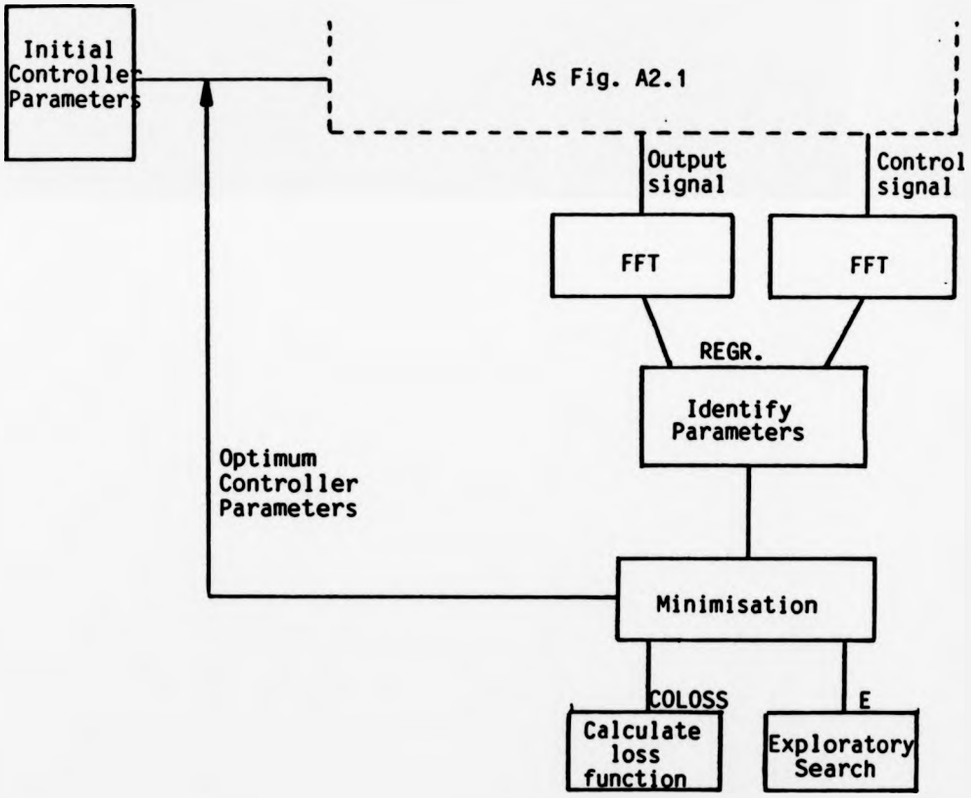


Fig. A2.2 Modification to flow chart for parametric adaptation

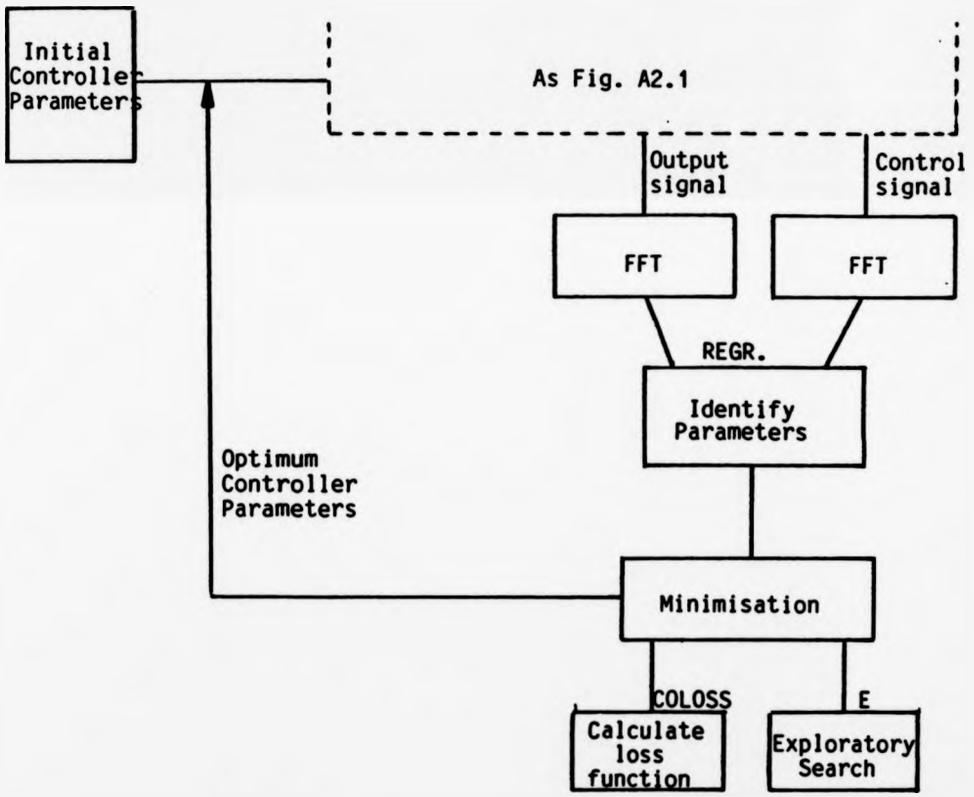


Fig. A2.2 Modification to flow chart for parametric adaptation

```
C           THIS PROGRAM SIMULATES ADAPTIVE SYSTEM USING
C           NON-PARAMETRIC IDENTIFICATION. SYSYEM ADAPTS
C           FROM INITIAL PARAMETER SETTINGS-PARAMETER
C           VALUES ARE PRINTED EVERY BLOCK
C
C
      DOUBLE PRECISION RD(1025),WND(1025),A(10),B(10),RMEAN,RVAR,G05DDF
      COMPLEX G(1025),PHRE(1025),PHRC(1025)
      DIMENSION R(1025),S(1025),C(1025),WN(1025),U(1025),WD(1025),BI(3)
      COMMON/COM1/N,ER(1024) ,EI(1024),CR(1024),CI(1024),RMR(1024)
     +                  ,RMI(1024)
      COMMON/COM2/AE,NE
      COMMON/COM3/NN,NR
      DATA RN,CN,WS,PS/0.0,0.0,0.0,0.0/
C
C           SET FLAGS TO RESET SIMULATION
      NFLAG=0
      MFLAG=0
C
C           SET CONSTANTS(TS=SAMPLING TIME,NR=SYSTEM ORDER)
      PI=3.14159
      TS=0.1
      NR=5
C           INITIALISE RANDOM INPUT (RMEAN=MEAN VALUE,RVAR=VARIANCE)
C           G05CBF IS NAG ROUTINE TO PROVIDE RANDOM DATA
      RMEAN=0.0
      RVAR=1.0
      CALL G05CBF(0)
C
C
C
C              INITIALISE PARAMETER ARRAYS
C
      DO 100 J=1,10
      A(J)=0.0
      B(J)=0.0
  100 CONTINUE
```

```
C           SET PARAMETERS IN SIMULATION
C
      A(1)=1.0
      A(2)=3.0
      A(3)=3.0
      A(4)=1.0
      B(1)=9261.0
      B(2)=-25137.0
      B(3)=22743.0
      B(4)=-6609.0
C
C           SET PARAMETERS AT RUN TIME
C           N=NO. OF POINTS IN BLOCK, NF=NUMBER OF FREQENCIES
C           NBK=NO. OF BLOCKS, AWN=NOISE FACTOR
      PRINT 210
  210 FORMAT('GIVE NO OF POINTS IN BLOCK')
      READ(1,*)N
      PRINT 220
  220 FORMAT('GIVE NO OF POINTS IN ESTIMATION')
      READ(1,*)NF
      PRINT 230
  230 FORMAT('GIVE NO OF BLOCKS')
      READ(1,*)NBK
      PRINT 240
  240 FORMAT('GIVE NOISE FACTOR')
      READ(1,*)AWN
C
C           SET NN=N FOR COMMON LINK
      NN=N
C
C           SET HANNING WINDOW
      CALL WINDOW(N,WD)
C           SET CROSS SPECTRA TO ZERO
      DO 250 J=1,N
      PHRE(J)=(0.0,0.0)
  250 PHRC(J)=(0.0,0.0)
C
C
C           SET INITIAL ESTIMATOR PARAMETERS
      GN=0.0
```

```
      GN1=0.001
      TD=0.002
C
C           START ADAPTIVE CONTROL LOOP
      DO 900 LCNT=1,NRK
C           CHECK FLAG IF ZERO INITIALISE SYSTEM
C
      IF(NFLAG.EQ.1)GO TO 444
C
C             LOW PASS FILTER
C             DIGITAL LOW PASS FILTER,COEFFICIENTS RN,RN1,CN,CN1
C
      DO 300 J=1,N
      RN1=RN
      RD(J)=GOSDDF(RMEAN,RVAR)
      RN=RD(J)
      CN1=CN
      CN=(RN+RN1+199.0*CN1)/201.0
      R(J)=CN
      WS1=WS
      WND(J)=GOSDDF(RMEAN,RVAR)
      WS=WND(J)
      PS1=PS
      PS=(WS+WS1+199.0*PS1)/201.0
      WN(J)=AMIN1(PS
  300 CONTINUE
C
C           SUBROUTINE DIF SIMULATES SYSTEM AND CONTROLLER
C           R=INPUT,C=OUTPUT,E=ERROR,U=CONTROL,A,B=SYSTEM
C           COEFFICIENTS,LFLAG INITIALISES SYSTEM
      CALL DIF(N,NB,R,C,E,U,WN,A,B,GN,GN1,TD,LFLAG)
      NFLAG=1
  800 CONTINUE
C
C             SYSTEM SIMULATION
C
      DO 350 J=1,N
      RN1=RN
      RD(J)=GOSDDF(RMEAN,RVAR)
      RN=RD(J)
      WS=0.0
```

```
      CN1=CN
      CN=(RN+RN1+199.0*CN1)/201.0
      R(J)=CN
      WS1=WS
      WND(J)=GOSDDF(RMEAN,RVAR)
      WS=WND(J)
      PS1=PS
      PS=(WS+WS1+199.0*PS1)/201.0
  350 WN(J)=AWN*PS
      CALL DIF(N,NB,R,C,E,U,WN,A,B,GN,GNI,TD,LFLAG)
      LFLAG=0
C               WINDOW DATA
C
      DO 400 J=1,N
      R(J)=WD(J)*R(J)
      U(J)=WD(J)*U(J)
      E(J)=WD(J)*E(J)
      C(J)=WD(J)*C(J)
  400 CONTINUE
C               TAKE FFT
C
      DO 450 K=1,N
      CR(K)=C(K)
      RMR(K)=R(K)
      ER(K)=U(K)
      RMI(K)=0.0
      EI(K)=0.0
  450 CI(K)=0.0
C
C         FTOIA  FFT SUBROUTINE RMR,RMI=REAL AND IMAG
C         PARTS OF DATA RETURNED AS REAL AND IMAG PARTS
C         OF TRANSFORM
      CALL FTOIA(N,1,RMR,RMI)
      CALL FTOIA(N,1,CR,CI)
      CALL FTOIA(N,1,ER,EI)
C
C           CALCULATE CROSS SPECTRA
C
      DO 500 J=1,NB
      PHRE(J)=PHRE(J)+(CMPLX(RMR(J),-RMI(J)))*(CMPLX(ER(J),EI(J)))
      PHRC(J)=PHRC(J)+(CMPLX(RMR(J),-RMI(J)))*(CMPLX(CR(J),CI(J)))
  500 G(J)=PHRC(J)/PHRE(J)
```

```
C
C
C              MINIMISATION
C
C          SET INITIAL POINT FOR MINIMISATION
      BI(1)=GN
      BI(2)=GNI
      BI(3)=TD
C
C          SUBROUTINE MINIM MINIMISES COST FUNCTION
C          WITH RESPECT TO PARAMETERS BI
      CALL MINIM(BI,S)
C
C          SET CONTROLLER PARAMETERS AT MINIMUM
      GN=BI(1)
      GNI=BI(2)
      TD=BI(3)
C          PRINT VALUES OF PARAMETERS

      PRINT *,GN,GNI,TD
  900  CONTINUE
      STOP
      END
C
      SUBROUTINE DIF(N,NB,R,C,E,U,W,A,B,GN,GNI,TD,LFLAG)
C          THIS SUBROUTINE SOLVES THE DIFFERENCE EQUATION
C          RELATING OUTPUT C AND ERROR E TO INPUT R AND
C          NOISE W.
C
C          SYSTEM  G(Z)=A(1)+A(2)*Z**-1+.........
C                       ----------------------------
C                       B(1)+B(2)*Z**-1+.........
C
C          CONTROLLER   D(Z)=AL(1)+AM(1)*Z**-1+AN(1)*Z**-2
C                           ---------------------------
C                           BD(1)+BD(2)*Z**-1+BD(3)*Z**-2
C
C          NB=ORDER OF B(Z)
C
```

```
      DIMENSION R(N),C(N),E(N),U(N),W(N)
      DOUBLE PRECISION A(10),B(10),AD(10),RM(10),EM(10),CM(10),WM(10)
      DOUBLE PRECISION BD(10),BBD(10),UM(10),AL(10),AM(10),AN(10)
      DOUBLE PRECISION ESUM,CSUM,USUM
      DATA RM,EM,CM,WM,UM/10*0.0,10*0.0,10*0.0,10*0.0,10*0.0/
      DATA AL,AM,AN/10*0.0,10*0.0,10*0.0/
C
C         SET CONSTANTS
      NB1=NB+1
      TS=0.1
C
C           SETS ALL INITIAL CONDITIONS TO ZERO IF LFLAG=1
      IF(LFLAG.EQ.0)GO TO 10
      DO 20 J=1,10
      EM(J)=0.0
      CM(J)=0.0
      UM(J)=0.0
      RM(J)=0.0
      WM(J)=0.0
   20 CONTINUE
   10 CONTINUE
C
C             SET DENOMINATOR OF CONTROLLER
      BD(1)=1.0
      BD(2)=-1.0
      BD(3)=0.0
C
C            FORM B(Z)*BD(Z)
      BBD(1)=B(1)*BD(1)
      BBD(2)=B(1)*BD(2)+B(2)*BD(1)
      BBD(3)=B(1)*BD(3)+B(2)*BD(2)+B(3)*BD(1)
      BBD(4)=           B(2)*BD(3)+B(3)*BD(2)+B(4)*BD(1)
      BBD(5)=                     B(3)*BD(3)+B(4)*BD(2)
      BBD(6)=                               B(4)*BD(3)
      BBD(7)=0.0
C
C           FORMS AL(1),AM(1),AN(1) IN TERMS OF CONTROLLER
C           PARAMETERS
      AL(1)=GN+TD/TS+TS*GNT
```

```
      AM(1)=-(GN+TD/TS+TD/TS)
      AN(1)=TD/TS
C
C             AD(Z)=A(7)*(AL+AM*Z**-1+AN*Z**-2)
C
      AD(1)=AL(1)*A(1)
      AD(2)=AN(1)*A(1)+AL(1)*A(2)
      AD(3)=AN(1)*A(1)+AM(1)*A(2)+AL(1)*A(3)
      AD(4)=           AN(1)*A(2)+AM(1)*A(3)+AL(1)*A(4)
      AD(5)=                     AN(1)*A(3)+AM(1)*A(4)
      AD(6)=                               AN(1)*A(4)
      AD(7)=0.0
C
C          SOLVE DIFFERENCE EQUATION
C
      DO 50 K=1,N
      RM(1)=R(K)
      WM(1)=W(K)
C
      ESUM1=0.0
      CSUM=0.0
      USUM=0.0
      DO 75 J=1,NB1
      ESUM1=ESUM1+(RM(J)-WM(J))*BBD(J)-EM(J+1)*(BBD(J+1)+AD(J+1))
 75   CONTINUE
C               DIVIDE BY COEFF. OF C(N),E(N)
C
      EM(1)=(ESUM1)/(BB)(1)+AL(1)*A(1))
      USUM=USUM+AL(1)*EM(1)+AM(1)*EM(2)+AN(1)*EM(3)+UM(2)
      UM(1)=USUM
C
      DO 81 J=1,4
      CSUM=CSUM+A(J)*UM(J)-CM(J+1)*B(J+1)+WM(J)*B(J)
 81   CONTINUE
      CM(1)=CSUM/B(1)
C               SET CURRENT VALUE OF C(N),E(N)
C
      C(K)=CM(1)
      E(K)=EM(1)
      U(K)=UM(1)
```

```
C             UPDATE POSITION OF PREVIOUS DATA IN STORE
C
      DO 100 J=1,NB1
      RM(NB1+2-J)=RM(NB1+1-J)
      CM(NB1+2-J)=CM(NB1+1-J)
      WM(NB1+2-J)=WM(NB1+1-J)
      EM(NB1+2-J)=EM(NB1+1-J)
      UM(NB1+2-J)=UM(NB1+1-J)
      AL(NB1+2-J)=AL(NB1+1-J)
      AM(NB1+2-J)=AM(NB1+1-J)
      AN(NB1+2-J)=AN(NB1+1-J)
  100 CONTINUE
   50 CONTINUE
      RETURN
      END
C
C
C
C
      SUBROUTINE FTO1A (IT,INV,TR,TI)

C###### 15/05/70 LAST LIBRARY UPDATE
C          THIS ROUTINE CALCULATES THE FOURIER TRANSFORM OF EQUALLY SPACE
C          F(N)  N=0,1,...,IT-1
C          THE DATA IS TAKEN TO BE PERIODIC IE.    F(N+IT) = F(N)
C          +++++++ ARGUMENTS SET BY THE CALLING PROGRAM ++++++
C          IT IS THE PROBLEM SIZE AND MUST BE A POWER OF 2
C          INV = 2 FOR DIRECT TRANSFORM IE.
C          G(M) = SUM OVER N=0,1,...,IT-1 OF F(N)*EXP(2PI*SQRT(-1)*N*M/IT)
C          FOR N=0,1,...,IT-1
C          INV = 1 FOR INVERSE TRANSFORM IE.
C       F(N) = (1./IT)*(SUM OVER M=0,1, .,IT-1 OF G(M)*EXP(-2PI*SQRT(-1)*
C          FOR N =0,1,...,IT-1
C          TR(I)     I=1,2,..,IT MUST CONTAIN REAL PART OF DATA
C          TI(I)     I=1,2,..,IT MUST CONTAIN THE IMAGINARY PART OF DATA
C          ++++++ ARGUMENTS SET BY ROUTINE ++++++
C          IF IT IS NOT A POWER OF 2 INV IS SET TO -1 FOR ERROR RETURN
C         TR(I)     I=1,2,..,IT IS SET TO REAL PART OF TRANSFORM
C          TI(I)     I=1,2,..,IT IS SET TO THE IMAGINARY PART OF TRANSFORM
C          THE METHOD USED IN THIS ROUTINE IS DISCRIBED IN
C          (GENTLEMAN AND SANDE, PROC. FALL JOINT COMPUTER CONFER. 1966)
```

```
      DIMENSION TR(1025),TI(1025),UR(15),UI(15)
      DATA KJUMP/1/
      GO TO(100,200),KJUMP
  100 UM=.5
      DO 50 I=1,15
      UM=.5*UM
      TH=6.283185307178*UM
      UR(I)=COS(TH)
   50 UI(I)=SIN(TH)
  200 UM=1.
      GO TO(1,2),INV
    1 UM=-1.
    2 IO=2
      DO 3 I=2,16
      IO=IO+IO
      IF(IO-IT)3,4,5
    3 CONTINUE
C     ERROR IN IT - SET INV=-1 AND RETURN
    5 INV=-1
      RETURN
C     IT= 2**I - INITIALISE OUTER LOOP
    4 IO=I
      II=IO
      I1=IT/2
      I3=1
C     START MIDDLE LOOP
   10 K=0
      I2=I1+I1
C     CALCULATE TWIDDLE FACTOR E(K/IP)
   11 WR=1.
      WI=0.
      KK=K
      JO=IO
   24 IF(KK)21,22,21
   21 JO=JO-1
      KK1=KK
      KK=KK/2
      IF(KK1-2*KK)23,21,23
   23 WS=WR*UR(JO)-WI*UI(JO)
      WI=WR*UI(JO)+WI*UR(JO)
```

```
      WR=WS
      GO TO 24
   22 WI=WI*UM
C     START INNER LOOP
      J=0
C     DO 2*2 TRANSFORM
   31 L=J*I2+K
      L1=L+I1
      ZR=TR(L+1)+TR(L1+1)
      ZI=TI(L+1)+TI(L1+1)
      Z=WR*(TR(L+1)-TR(L1+1))-WI*(TI(L+1)-TI(L1+1))
      TI(L1+1)=WR*(TI(L+1)-TI(L1+1))+WI*(TR(L+1)-TR(L1+1))
      TR(L+1)=ZR
      TR(L1+1)=Z
      TI(L+1)=ZI
C     INDEX J LOOP
      J=J+1
      IF(J-I3)31,12,12
C     INDEX K LOOP
   12 K=K+1
      IF(K-I1)11,6,6
C     INDEX OUTER LOOP
    6 I3=I3+I3
      I0=I0-1
      I1=I1/2
      IF(I1)51,51,10
C     UNSCRAMBLE
   51 J=1
      UM=1.
      GO TO(61,52),INV
   61 UM=1./FLOAT(I1)
   52 K=0
      J1=J
      DO 53 I=1,I1
      J2=J1/2
      K=2*(K-J2)+J1
   53 J1=J2
   54 IF(K-J)66,56,55
   56 TR(J+1)=TR(J+1)*UM
      TI(J+1)=TI(J+1)*UM
```

```
      GO TO 66
   55 ZR=TR(J+1)
      ZI=TI(J+1)
      TR(J+1)=TR(K+1)*UM
      TI(J+1)=TI(K+1)*UM
      TR(K+1)=ZR*UM
      TI(K+1)=ZI*UM
   66 J=J+1
      IF(J-IT+1)52,57,57
   57 TR(1)=TR(1)*UM
      TI(1)=TI(1)*UM
      TR(IT)=TR(IT)*UM
      TI(IT)=TI(IT)*UM
      RETURN
      END
C
C
C
      SUBROUTINE MININ(BI,S)
C
C           GIVEN AN INITIAL PARAMETER VECTOR BI
C           THIS SUBROUTINE SEARCHS FOR THE VECTOR
C           THAT MINIMISES S.
C
      DIMENSION BI(3),BR(3),H(3),B(20),BP(10)
C
C            SET END CRIERION EPSIL AND DIMENSION N
C            INITIAL SEARCH POINT B AND STEP LENGTH H
C
      N=3
      EPSIL=0.01
      DO 70 K=1,N
   70 B(K)=BI(K)
      DO 60 K=1,N
   60 H(K)=0.1
C
C          EVALUATE LOSS FUNCTION AT B GIVE IT VALUE S
      CALL F(B,N,S,IND)
   48  CONTINUE
C
```

```
C          PRESERVE CO-ORDS OF INITIAL POINT AND MAKE
C          EXPLORATORY MOVE USING E(     )
      DO 6 KA=1,N
 6    BR(KA)=B(KA)
      CALL E(BR,H,N,S,VB)
C         IF MOVE SUCESS GO TO 7
      IF(S.LT.VB)GO TO 7
      GO TO 8
7     DO 9 KA=1,N
9     BP(KA)=B(KA)
14    CONTINUE
C          RECORD INITIAL POINT AND BEST POINT OF
C          EXPLORATORY MOVE
      DO 29 KA=1,N
 29   B(KA)=BR(KA)
      VB=S
C                DO PATTERN MOVE
      DO 10 KA=1,N
 10   BR(KA)=BP(KA)+2.0*(B(KA)-BP(KA))
      CALL F(BR,N,S,IND)
      IF(IND.EQ.0)GO TO 48
      IF(S.GE.VB)GO TO 48
      GO TO 14
8     HA=0.0
C          MEASURE DISTANCE BETWEEN POINTS, IF LESS THAN
C          EPSIL TERMINATE, IF NOT HALVE EACH STEP
      DO 88 LAP=1,N
 88   HA=HA+H(LAP)*H(LAP)
      HA=SQRT(HA)
      IF(HA.LT.EPSIL)GO TO 12
      DO 13 KA=1,N
 13     H(KA)=0.5*H(KA)
      GO TO 48
 12   CONTINUE
      DO 19 K=1,N
 19   BI(K)=BR(K)
      RETURN
      END
```

```
      SUBROUTINE E(B,H,N,W,VB)
C
C        THIS SUBROUTINE CARRIES OUT EXPLORATORY MOVE
C
      DIMENSION B(N),H(N)
      CALL F(B,N,VB,IND)
      W=VB
      DO 2 KIP=1,N
      B(KIP)=B(KIP)+H(KIP)
      CALL F(B,N,V,IND)
      IF(IND.EQ.0)GO TO 4
      IF(V.LT.W)GO TO 3
 4    B(KIP)=B(KIP)-2.0*H(KIP)
      CALL F(B,N,V,IND)
      IF(IND.EQ.0)GO TO 5
      IF(V.LT.W)GO TO 3
 5    B(KIP)=B(KIP)+H(KIP)
      GO TO 2
 3    W=V
 2    CONTINUE
      RETURN
      END
C
C
C
C
C
      SUBROUTINE F(X,M,V,IND)
C
C        THIS SUBROUTINE FORMS THE FUNCTION TO BE MINIMISED
C
      DIMENSION X(M)
      COMPLEX RW(513),DW(513),EW(513),G(513),OMJ
      COMMON/COM3/N,NF,G
      IND=1
      PI2=6.283185
      N2=N/2
      TS=0.1
      V=0.0
```

```
C          CHECK CONSTRAINTS ON PARAMETERS
      IF(X(1).LT.0.0)GO TO 90
      IF(X(1).GT.6.0)GO TO 90
      IF(X(2).LT.0.0)GO TO 90
      IF(X(2).GT.0.5)GO TO 90
      IF(X(3).LT.0.0)GO TO 90
C          SET INITIAL PARAMETER VALUES
      IF(X(3).GT.1.0)GO TO 90
      GN=X(1)
      GNI=X(2)
      TD=X(3)
C        CALCULATE COST FUNCTION
      DO 10 J=1,NF
      OM=PI2*FLOAT(J-1)/(TS*FLOAT(N))
      IF(J.EQ.1)OM=0.001
      OMJ=CMPLX(0.0,OM)
      RW(J)=1.0/(1+10.0*OMJ)
      DW(J)=(GNI+GN*OMJ+TD*(OMJ**2))/OMJ
      EW(J)=1.0/(1.0+DW(J)*G(J))
      IF(CABS(EW(J)).GT.5.0)GO TO 90
      V=V+CABS(RW(J))*CABS(RW(J))*CABS(EW(J))*CABS(EW(J))
 10   CONTINUE
      RETURN
 90   IND=0.0
      RETURN
      END
C
C
C
C
      SUBROUTINE WINDOW(N,WD)
C
C          CALCULATES HANNING WEIGHTING
      DIMENSION WD(N)
      PI2=6.283185
      DO 20 J=1,N
      NW=J-1-N/2
 20   WD(J)=0.5*(1.0+COS(PI2*NW/N))
      RETURN
      END
```

```
C           THIS PROGRAM SHOWS THE MODIFICATIONS REQUIRED
C           FOR PARAMETRIC IDENTIFICATION THE SUBROTINE
C           F(     ) IS MODIFIED AND SUBROUTINES REGR(    )
C           AND COLOSS(     ) ARE ADDED.
C
C
C
C
      SUBROUTINE F(X,M,V,IND)
C
C            THIS SUBROUTINE FORMS THE FUNCTION TO BE
C            MINIMISED. IT CALLS THE SUBROUTINE COLOSS.
C
      DIMENSION X(M),AA(20),BB(20)
      COMMON/COM2/A(20),B(20)
      IN=20
      IND=1
C         CHECK CONSTRAINTS ON PARAMETERS
      IF(X(1).LT.0.0)GO TO 90
      IF(X(2).LT.0.0)GO TO 90
      IF(X(3).LT.0.0)GO TO 90
      IF(X(3).GT.1.0)GO TO 90
      GN=X(1)
      GNI=X(2)
      TD=X(3)
      T1=10.0
      NA=5
C
C           SET COEFFICIENTS IN CALCULATION OF LOSS FUNCTION
      AA(1)=T1
      AA(2)=1.0+B(3)*T1
      AA(3)=B(3)+B(2)*T1+A(1)*TD*T1
      AA(4)=B(2)+A(1)*TD+B(1)*T1+A(1)*GN*T1
      AA(5)=B(1)+A(1)*GN+A(1)*GNI*T1
      AA(6)=A(1)*GNI
      BB(1)=B(4)
      BB(2)=B(3)
      BB(3)=B(2)
      BB(4)=B(1)
      BB(5)=0.0
C
```

```
C            CALL SUBROUTINE TO CALCULATE LOSS
      CALL COLOSS(AA,BB,NA,IND,V,IN)
      RETURN
 90   IND=0.0
      RETURN
      END
C
C
      SUBROUTINE COLOSS(A,B,N,IERR,V,IN)
C
C            ASTROM'S PROGRAM FOR CALCULATION OF LOSS
      DIMENSION A(IN),B(IN)
      IERR=1
      V=0.0
      IF(A(1)) 70,70 ,10
 10   DO 20 K=1,N
      IF(A(K+1)) 70,70,30
 30   ALFA=A(K)/A(K+1)
      BETA=B(K)/A(K+1)
      V=V+BETA**2/ALFA
      K1=K+2
      IF(K1-N)50,50,90
 50   DO 60 I=K1,N,2
      A(I)=A(I)-ALFA*A(I+1)
 60   B(I)=B(I)-BETA*A(I+1)
 20   CONTINUE
      V=V/2
      RETURN
 70   IERR=0
      RETURN
      END
C
C
C
C

C
```

```
      SUBROUTINE REGR(XC,NF)
C
C          THIS PROGRAM USES REGRESSION IN THE FREQUECY
C          DOMAIN TO FIT A MODEL TO INPUT AND OUTPUT
C          FREQUENCY RESPONSES.
C
      DOUBLE PRECISION FR(7,1),HR(7,1),FRT(1,7),HRT(1,7)
      DOUBLE PRECISION ALFA(7,1),ALFA1(7,1),PR(7,7)
      DOUBLE PRECISION A(1,7),AA(1,1),BB(7,7),CC(7,7)
      DOUBLE PRECISION DD(1,1),EE(1,1),PR1(7,7),RKR(7,7)
      DOUBLE PRECISION CF(7,1),GH(7,1),GG(7,1),AM(7,1),V,Z(1)
      COMPLEX OMJ,BZ,U(7),UC(7),CWH
      COMPLEX CW(513),EW(513),RW(513)
      DIMENSION UT(7,7),UTM(7,7),UYM(7),XC(7),UY(7)
      COMMON/COM1/N,ER(513),EI(513),CR(513),CI(513),RR(513),RI(513)
      DATA ALFA1,PR1/7*0.0,100000.0,7*0.0,100000.0,7*0.0,100000.0
     +,7*0.0,100000.0,7*0.0,100000.0,7*0.0,100000.0,7*0.0,100000.0/
      TS=0.1
      PI2=6.283185
      IFAIL=0
      NP=7
      LS=7
      IA=NP
      NN=NP
      IAA=NP
      NU=NP
C
C          SET COMPLEX DATA ARRAYS
C
      DO 10 JF=1,NF
      RW(JF)=CMPLX(RR(JF),RI(JF))
      EW(JF)=CMPLX(ER(JF),EI(JF))
      CW(JF)=CMPLX(CR(JF),CI(JF))
 10    CONTINUE
C
C        INITIALISE ALPHA1,PR1
C
      DO 60 J=5,LS
      ALFA1(J,1)=0.0
 60   CONTINUE
```

```
      DO 80 J=5,LS
 80   PR1(J,J)=100000.0
C           INITIALISE ARRAYS UTD,UYD
C
C
C          LOOP DATA POINTS
C
      DO 21 LRPT=1,7
      DO 20 JF=1,NF
      OM=PI2*FLOAT(JF-1)/(TS*FLOAT(N))
      OMJ=CMPLX(0.0,OM)
      RM=CABS(RW(JF))*CABS(RW(JF))
      RM=1.0
C
C          SET U AND UC VECTORS
C
      U(1)=EW(JF)
      U(2)=-OMJ*CW(JF)
      U(3)=-((OMJ**2)*CW(JF))
      U(4)=-((OMJ**3)*CW(JF))
      U(5)=(1.0,0.0)
      U(6)=OMJ
      U(7)=OMJ**2
C
C         CALCULATE UPDATED ALFA1,PR1
C
C          CALCULATE UT AND ITS REAL & IMAG PARTS
C
      A2=SNGL(ALFA(2,1))
      A3=SNGL(ALFA(3,1))
      A4=SNGL(ALFA(4,1))
      BZ=1.0+A2*OMJ+A3*(OMJ**2)+A4*(OMJ)**3)
      DRS=(CABS(BZ))**2
C
      DO 200 J=1,LS
      FR(J,1)=DBLE(REAL(U(J)))
      FRT(1,J)=DBLE(REAL(U(J)))
      HR(J,1)=DBLE(AIMAG(U(J)))
      HRT(1,J)=DBLE(AIMAG(U(J)))
 200  CONTINUE
C
```

```
C            V=DRS/2*(FRT*PR1*FR)
C
      CALL F01CKF(A,FRT,PR1,1,LS,LS,Z,1,1,IFAIL)
      CALL F01CKF(AA,A,FR,1,1,LS,Z,1,1,IFAIL)
      V=DRS/2.0+AA(1,1)
C
C                CC=I-R*I-RT*PR1
C
      CALL F01CKF(BB,FR,FRT,LS,LS,1,Z,1,1,IFAIL)
      CALL F01CKF(CC,BB,PR1,LS,LS,LS,Z,1,1,IFAIL)
C
C           CC=1-(FR*FRT*PR1)/V
C
      DO 300 J=1,LS
      DO 310 K=1,LS
      CC(J,K)=-CC(J,K)/V
  310 CONTINUE
  300 CONTINUE
      DO 320 J=1,LS
      CC(J,J)=1.0+CC(J,J)
  320 CONTINUE
C
C              RKR=PR1*CC
C
      CALL F01CKF(RKR,PR1,CC,LS,LS,LS,Z,1,1,IFAIL)
C
C              V=DRS/2.0+HRT*RKR*HR
C
      CALL F01CKF(A,HRT,RKR,1,LS,LS,Z,1,1,IFAIL)
      CALL F01CKF(AA,A,HR,1,1,LS,Z,1,1,IFAIL)
      V=DRS/2.0+AA(1,1)
C
C                CC=HR-HRT*RKR
C
      CALL F01CKF(BB,HR,HRT,LS,LS,1,Z,1,1,IFAIL)
      CALL F01CKF(CC,BB,RKR,LS,LS,LS,Z,1,1,IFAIL)
C
C                DD=1-(HR*HRT*RKR)/V
C
```

```
      DO 340 J=1,LS
      DO 360 K=1,LS
      CC(J,K)=-CC(J,K)/V
  360 CONTINUE
  340 CONTINUE
      DO 380 J=1,LS
      CC(J,J)=1.0+CC(J,J)
  380 CONTINUE
C
C           PR=RKR*CC
C
      CALL F01CKF(PR,RKR,CC,LS,LS,LS,Z,1,1,IFAIL)
C
C             FR*(ER(I)-FRT*ALFA1)
C             HR*(IMAG(X)-HRT*ALFA1)
      CALL F01CKF(DD,FRT,ALFA1,1,1,LS,Z,1,1,IFAIL)
      CALL F01CKF(EE,HRT,ALFA1,1,1,LS,Z,1,1,IFAIL)
      DO 400 J=1,LS
      GF(J,1)=2.0*(CR(JF)-DD(1,1))*FR(J,1)/DRS
      GH(J,1)=2.0*(CI(JF)-EE(1,1))*HR(J,1)/DRS
      GG(J,1)=GF(J,1)+GH(J,1)
  400 CONTINUE
  222 FORMAT(2E20.10)
C
C             PR*GG
C
      CALL F01CKF(AM,PR,GG,LS,1,LS,Z,1,1,IFAIL)
      DO 420 J=1,LS
      ALFA(J,1)=ALFA1(J,1)+AM(J,1)
  420 CONTINUE
C
C              RESET ALFA1,PR1
C
      DO 440 J=1,LS
      ALFA1(J,1)=ALFA(J,1)
      DO 460 K=1,LS
      PR1(J,K)=PR(J,K)
  460 CONTINUE
  440 CONTINUE
   20 CONTINUE
```

```
C

 21   CONTINUE
      XC(1)=ALFA(1,1)
      XC(2)=ALFA(2,1)
      XC(3)=ALFA(3,1)
      XC(4)=ALFA(4,1)
      RETURN
      END
```

APPENDIX 3

Published Work

This appendix contains copies of the following papers that are related to the thesis and have been published, or presented at conferences.

1. Balmer, L. and Douce, J.L., "An application of Hill-climbing Techniques in Measurement", IEEE Trans. on Automatic Control, Vol. AC-27, pp.89-94, Feb. 1982.

2. Douce, J.L. and Balmer, L., "Bias of Frequency Response Estimates", 2nd International Conference of Systems Science", Wroclaw, Sept. 1981.
   Also published in "Systems Science", Vol.8, No.2-3, pp.95-102, April 1982.

3. Douce, J.L. and Balmer, L., "Frequency Response Estimation using Short Data Blocks", 3rd International Conference on Systems Engineering", Wright State University, Sept. 1984.

4. Douce, J.L. and Balmer, L., "Transient Effects in Spectrum Estimation", IEE Proceedings, Vol.132, Pt.D, No.1, pp.25-27, Jan. 1985.

5. Balmer, L. and Douce, J.L., "Adaptive Control in the Frequency Domain", 4th International Conference on Systems Engineering, Coventry (Lanchester) Polytechnic, England, Sept. 1985.

The following paper has been accepted for inclusion in the "International Conference on Systems Science", Wroclaw, Poland, Sept. 1986.

Douce, J.L. and Balmer, L., "Frequency-response Estimates using Short Records".

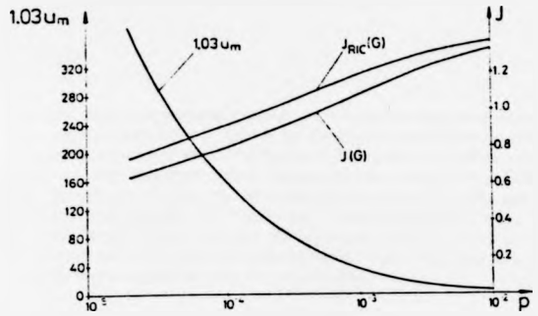


Fig. 4. Upper bound $1.03\,u_m$ and G-related performance indices $J_{RIC}(G)$ and $J(G)$ for the control systems (53) as a function of p in the case of the stable plant (44). The value of $u[G]$ is determined within the accuracy $\epsilon \le 0.03\,u[G]$ by $u_m \le u[G] \le 1.03\,u_m$.

sented in Fig. 4. The number of the required points $x^{(i)}(t_{i,j})$ was between 77 and 121 for all considered values p. Using a Fortran IV program and a PDP 11/10 computer, we did not need a computing time of more than 16 s for each value p. Reading Fig. 4 vice versa, we obtain (within the limits of ϵ) the smallest G-admissible value $p(u_{max})$ for any given u_{max}. The corresponding controller guarantees that the condition (47) is strictly met for all trajectories starting in G and that the admissible free play is actually used by the control input except for an amount of not greater than $0.03\,u_{max}$.

In Fig. 4, the curves $J_{RIC}(G)$ and $J(G)$ represent the mean values (with respect to all corners of G) of the integral (49), respectively, of the integral

$$J(x(0)) = \int_0^\infty x^T(t) c c^T x(t)\, dt \tag{54}$$

referring to the systems (53). It appears that for any given value of u_{max} the smallest G-admissible value $p(u_{max})$ yields that controller (50) which can be called "G-optimal" in the sense of the performance indices $J_{RIC}(G)$ and $J(G)$.

Remarks:

1) The choice of the positive definite matrix P determines the shape of the cutoff domains $H(r)$ and the value of the bound $v[H(r_2)]$. By using a suitably optimized matrix P instead of the identity matrix I we, therefore, can possibly reduce the number of required points $x^{(i)}(t_{i,j})$ even more. But since the optimization of P also needs numerical effort, it is open to further investigation whether the net computing time can be profitably reduced in this way.

2) If an unstable plant is considered, the results represented in Fig. 4 change as follows. The upper bound $(1+q)u_m$ now tends to a positive value u_{min} if p is increased at will. This is the minimum value of u_{max} which is required to stabilize the plant by a controller of the type (50).

3) If more than one constraint of the type (2) is given, for each of these constraints an individual condition (35) and individual radii r_0 and r_1 are obtained, which define corresponding cutoff domains (15). In this case, each time step $h(i, j)$ must be chosen so that each condition (35) is met as long as the corresponding cutoff domain is not yet reached.

VII. Discussion

The method described in this paper to calculate $v[G]$ or to establish G-admissibility for a given linear controller represents a new tool for analysis of linear control systems of the type (1). Moreover, it can be applied to arbitrary finite-dimensional continuous linear control systems containing a dynamical linear controller as these systems can also be represented in the form

$$\dot{\hat{x}}(t) = \hat{A}\hat{x}(t), \qquad u(t) = -K\hat{x}(t) \tag{55}$$

where the matrix $\hat{A}$ now is of higher dimension than the matrix A referring to the plant [5].

The above method can also serve as a design tool if it is introduced in any existing procedure for the design of continuous linear control systems. It makes it possible to ensure that all given inequality constraints of the type (2) are met without giving away admissible free play of the restricted variables. This is especially useful if only small admissible free play is left for certain variables or if other additional conditions strongly restrict the family of the admissible controllers. (For instance, additional conditions given in the K-space can be introduced by robustness requirements [6].)

The method proposed here can also be applied to discrete time systems of the type

$$x((k+1)T) = \Phi x(kT), \qquad k = 0, 1, 2, \cdots \tag{56}$$

if one is interested in the system performance only referring to the time instants kT. In this case the matrix R, which defines the cutoff domains, has to be chosen as solution of the equation

$$\Phi^T R \Phi - R = -P \tag{57}$$

where P is an arbitrary positive definite matrix. Furthermore, one can always work with constant time steps $h(i, j) = T$. Finally, there is no remaining error ϵ in this case.

As the image of an arbitrary ellipsoid referring to the map (8) again is an ellipsoid, the method proposed above can be suitably modified if the set G of the possible initial states can be better described by an ellipsoid G_0 than by a parallelepiped. In this case, instead of the points $x^{(i)}(t_{i,j})$ a sequence of ellipsoids

$$G_{i+1} = \bar{\Phi}_{t_i}(G_i), \qquad i = 0, 1, 2, \cdots \tag{58}$$

is considered where the time distances t_i are chosen on the analogy of Section IV. The procedure is stopped as soon as the last ellipsoid is included by the cutoff domain. Applications of this modified concept are described in the thesis [7], which is based on a preceding version of this paper.

Acknowledgment

All numerical calculations and simulations presented in this paper were executed by B. Adomeit. The computer programs used for this include subroutines written by H. Kaltenhäuser.

References

[1] H. Kiendl, "Eine suboptimale Regelstrategie auf der Basis des Theorems von Cayley–Hamilton zur Synthese von Abtastsystemen mit beschränkter Stellgröße," *Regelungstechnik*, vol. 28, pp. 250–258, Aug. 1980.
[2] R. W. Brockett, *Finite Dimensional Linear Systems*. New York: Wiley, 1971, pp. 60–62.
[3] F. R. Gantmacher, *Matrix Theory*, vol. I. New York: Chelsea, 1960, pp. 33–41.
[4] D. G. Schultz and J. L. Melsa, *State Functions and Linear Control Systems*. New York: McGraw-Hill, 1967, pp. 264–275.
[5] T. L. Johnson and M. Athans, "On the design of optimal constrained dynamic compensators for linear constant systems," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 658–660, Dec. 1970.
[6] J. Ackermann, "Parameter space design of robust control systems," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 1058–1072, Dec. 1980.
[7] M. Kammler, "Entwurf gebietsoptimaler kontinuierlicher Regler mit vollständiger und unvollständiger Zustandsrückführung unter Berücksichtigung von Stellgrößenbeschränkungen," Ph.D. dissertation, University Dortmund, Germany, Feb. 1980.

An Application of Hill Climbing Techniques in Measurement

L. BALMER AND J. L. DOUCE

***Abstract*—This paper examines the dynamic performance of an automated optical gauging instrument. A mathematical model of the system is developed which demonstrates the intrinsically nonlinear characteristics**

Manuscript received April 17, 1979; revised October 22, 1979. Paper recommended by H. A. Spang, III, Past Chairman of the Applications, Systems Evaluation, and Components Committee.

L. Balmer is with the Department of Electrical Engineering, Lanchester Polytechnic, Coventry, England.
J. L. Douce is with the Department of Engineering, University of Warwick, Coventry, England.

0018-9286/82/0200-0089$00.75 ©1982 IEEE

which provide coupling between two nominally independent control loops. Approximate stability bounds for the closed-loop system are established using a matrix of describing functions. The inverse Nyquist diagram is used to design appropriate compensation networks, taking into account design objectives including the avoidance of various limit cycles and a rapid transient response. The final design is implemented on the instrument and its behavior compared with the simulated system. It is shown that the response time is reduced to approximately 11 percent of that of the original system in agreement with design predictions.

I. Introduction

The precision gauging of small mechanical components is expensive because of the equipment requirements and the time of skilled operatives. However, [1] describes a device that offers low cost accurate gauging of circular holes in flat plates. This paper gives a brief outline of the operation of the instrument but its main concern is with the improvement of its dynamic performance.

The instrument is designed to measure automatically the misalignments in the positions of circular holes in a flat plate. Fig. 1(a) gives an outline diagram showing the basic operation while Fig. 1(b) is a more complete block diagram of the system. The system is a null system requiring a reference plate with the holes in the correct position. This is located with its face parallel to that of the gauged plate and both plates are located by a suspension system [not shown in Fig. 1(a)] that allows lateral motion. Electromechanical actuators can cause X and Y displacements of both plates independently. The surface of gauged plate is illuminated by normal parallel light which passes through the holes in both plates and falls on photocells placed underneath each hole pair.

By applying a small sinusoidally varying displacement to the X axis and a similar displacement, phase shifted by 90°, to the Y axis of the reference plate a circular perturbation is produced. This causes a corresponding variation in the output voltage of the photocells. By using the signal from one of the cells as the inputs to a pair of phase sensitive rectifiers (the reference signals being the X and Y perturbation signals) voltages representing the X and Y misalignments of the hole associated with that cell are obtained. These signals are used as error signals for X and Y position control systems which move the plates into alignment by driving linear actuators attached to the nonoscillating plate. The displacements finally required by these control systems then give an accurate measure of the misalignment of the hole under examination.

Detailed discussion of the system is given below; at this point it may be noted that the two loops are coupled since the light transmitted through two overlapping circular holes is a function of the radial misalignment R.

The position control loop for each axis may be modeled to a first approximation by a simple hill-climbing system. Fig. 2 shows the X control loop as such a system where $h(t)$ represents the impulse response of the position control dynamics.

Note that the perturbation signal is modeled by an input after the control system dynamics, since the constant probing signal is applied to the master plate independent of the alignment motions applied to the gauged plate. This feature permits the use of a perturbation frequency significantly higher (by a factor of two in practice) than the natural frequency of the system.

A typical component that could be gauged by such a system would be a bearing plate for a wrist watch. Such a plate could contain some twenty circular holes ranging from 0.17–1.7 mm in diameter. Each hole is gauged in turn and the time taken to check the complete plate depends on the response time for each hole. This is the time it takes for the instrument to align a hole pair following an initial misalignment. In the original instrument this was in the order of one second and any attempt to reduce it by increase of system gain caused instability.

The object of the following analysis is to investigate the dynamics of the system in order to design a controller to reduce the response time while still maintaining adequate stability.

II. System Analysis

A. The Mathematical Model

Referring to Fig. 1(b) x_r, y_r denote X, Y position of a hole centre in the reference plate and x, y the centre of the corresponding hole in the gauged plate.

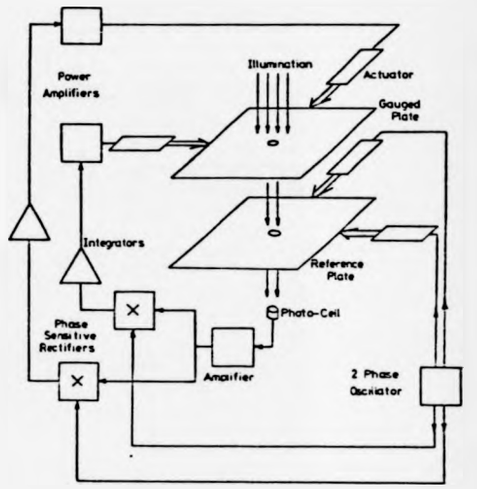


Fig. 1. (a) Basic form of instrument

The function $F_R(R)$ determines the area of overlap A, for a given radial misalignment R. It is given by

$$A = F_R(R) = \frac{D^2}{2}\left[\cos^{-1}\left(\frac{R}{D}\right) - \left(\frac{R}{D}\right)\sqrt{1-\left(\frac{R}{D}\right)^2}\right] \qquad (1)$$

where D is the diameter of the hole considered. This area determines the amount of light falling on a photocell which in turn determines, via the nonlinear function $F_A(A)$, the photocell output voltage. The form of the function $F_A(A)$ depends on the photodiode load resistor. For low values of load the relationship is linear, but for higher values of load resistor it takes the form

$$V = F_A(A) = k_1 \ln(1 + k_2 A) \qquad (2)$$

where k_1, k_2 are constants. The operating regime was deliberately chosen to correspond to this nonlinear function, as the overall control loop is less affected by different hole diameters than with the linear element.

The overall static characteristic, relating the mean value of the correcting signal, as measured at the output of a phase-sensitive rectifier, to hole misalignment has been calculated for a range of hole diameters D, for the particular perturbation amplitude $h = 0.127$ mm.

The table dynamics were modeled by a second-order transfer function. Initially no additional damping was incorporated, and the values of ζ and ω_n were of the order of 0.01 and $2\pi 35$ rad/s, respectively.

B. System Performance

Having chosen the constants of the mathematical model by simulation it was verified that the behavior of the model and real system were very close. The most satisfactory gain setting for K gave a settling time of the order of 1 s.

Increasing the gain of the forward path lead to instability of the form typified by Fig. 4.

C. Stability Analysis

The stability of the system is assessed by noting that it can be represented by two linear systems coupled by a nonlinear term. The outputs of the linear systems are the x and y misalignments of the centres of the reference and gauged holes. The nonlinearity arises since the light received by the photocell detector is proportional to the area of overlap of the two circular holes, which is in turn, a function of radial misalignment.

Hence, the system can be represented by the block diagram of Fig. 5. The inputs to the nonlinearity are linear misalignments and the outputs are the corresponding outputs of the phase sensitive detector. $N_{11}, N_{12}, N_{21}, N_{22}$ represent the describing functions of the nonlinearity.

The method of assessing stability is given in [2] and is based upon the position of the eigenvalues of the matrix $(N + G^{-1})$ where

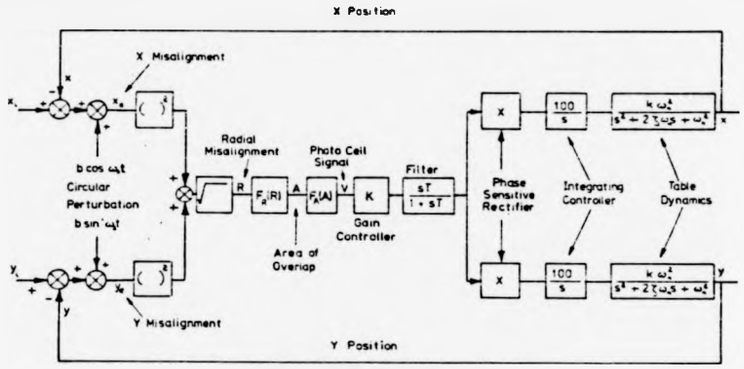


Fig. 1 (b) Block diagram of system.

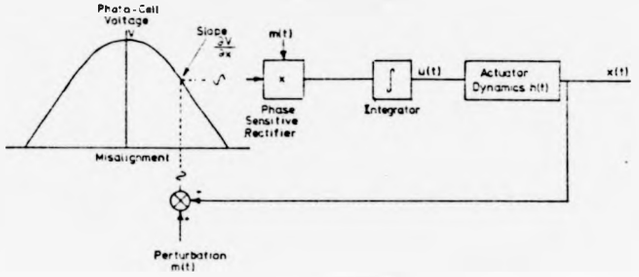


Fig. 2 One axis of the hill-climbing system.

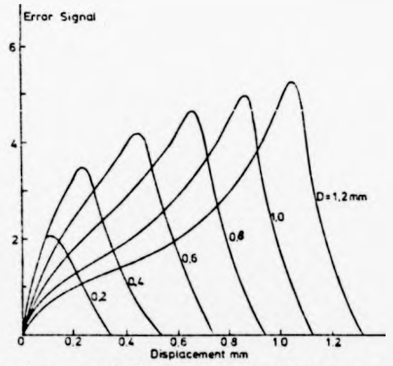


Fig. 3. Error signal/displacement for a range of hole diameters.

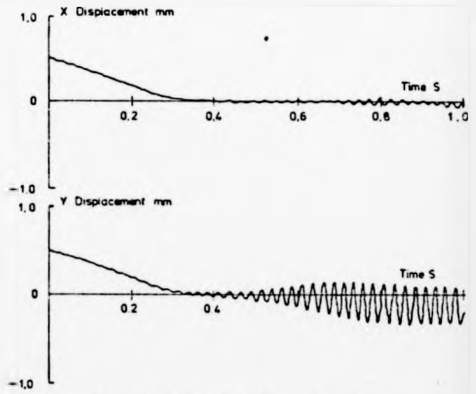


Fig. 4. Transient response showing instability.

$$N = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \quad G = \begin{bmatrix} G_{11} & 0 \\ 0 & G_{22} \end{bmatrix}.$$

D. System Stability

Using Gershgorin's theorem, the stability criterion is

$$|N_{kk}(a_k) + \tilde{G}_{kk}(j\omega)| > \sum_{j \neq k} |N_{jk}(a_k)| + \sum_{j \neq k} |\tilde{G}_{jk}(j\omega)|$$

where $\tilde{G}_{jk}$ represents the jk element in the matrix G^{-1} and a_k is the amplitude of its input signal.

Instability is here defined as a steady-state harmonic solution leading to a stable limit cycle.

For the case under consideration there is no interaction in the linear portion of the system and the nondiagonal terms in the G matrix are zero. This simplifies the stability criterion as for a diagonal matrix $\tilde{G}_{kk} = 1/G_{kk}$ and the criterion can be interpreted graphically as follows.

The describing function $-N_{11}$ is drawn and surrounded by a band of half width $|N_{21}|$, a Gershgorin band. For stability there must be no intersection between this band and the inverse Nyquist plot $1/G_{11}$. This criteria must also hold for the corresponding band $|N_{12}|$ around the plot of $-N_{22}$ and the plot of $1/G_{22}$.

In order to apply this criterion it is necessary to obtain the matrix of describing functions N.

The nonlinearity relating photocell output voltage to radial misalignment has been given in (1) and (2). The analytical description for this nonlinearity is too complex to use as a basis for the derivation of the describing function. However, some simplification can be obtained by approximating the relationship by a power series. As the relationship is even symmetrical around $R = 0$ the series will consist of even powers of R only.

$$v = c_0 + c_2 R^2 + c_4 R^4 + \cdots.$$

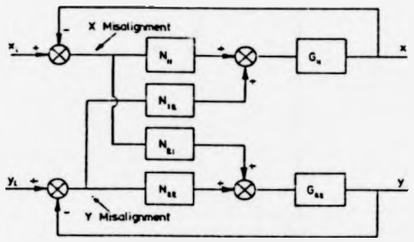


Fig. 5. System in multivariable form.

Expressing R as a ratio of its maximum value and normalizing to give $c_0 = 1$, the first four terms of a Chebyshev series for the 1 mm diameter hole are

$$v = 1 - 1.04R^2 + 0.39R^4 - 0.37R^6 + \cdots .$$

As the misalignments involved are less than half the hole diameter, the terms decrease rapidly in size and a reasonable approximation can be obtained by taking the first two terms only. The describing function between misalignment in one axis and phase sensitive rectifier output can be obtained by reference to Fig. 6. The relationship between $v(t)$, $x(t)$, $y(t)$ is given by

$$v(t) = c_0 - c_2\left[x^2(t) + 2bx(t)\cos\omega_s t + y^2(t) + 2by(t)\sin\omega_s t + b^2\right].$$

Note there are no terms involving both $x(t)$ and $y(t)$ in the above expression. As passing the signal into the phase sensitive rectifier will not introduce such terms, the nonlinearity has the form shown in Fig. 5. To obtain the describing function for a sinusoidal input the X and Y terms can be considered independently.

To obtain N_{11} consider $x(t) = x\cos(\omega t + \phi)$ and $y(t) = 0$. The action of the phase sensitive rectifier in the X channel can be described as a multiplication by the signal

$$m_x(t) = \frac{4}{\pi}\left[\cos\omega_s t - \frac{1}{3}\cos 3\omega_s t + \frac{1}{5}\cos 5\omega_s t \cdots\right]$$

where ω_s is the angular perturbation frequency, the perturbation frequency being 80 Hz.

The resultant output contains terms in ω, ω_s, $n\omega$, $m\omega_s$, $(n\omega \simeq m\omega_s)$ where m and n are integers. As is usual with the describing function method the harmonics of the input are ignored. However, the result is similar to that obtained by the dual-input describing function [3], the input ω must be considered to have an arbitrary phase relationship ϕ_x to the signal ω_s and the stability investigated for all values of ϕ_x. Special cases arise when ω and ω_s are commensurate, difference terms can then become equal to submultiples of the input frequency and produce sustained oscillation.

The term N_{12} is obtained by using the phase sensitive detector in the Y channel multiplying by the signal,

$$m_y(t) = \frac{4}{\pi}\left[\sin\omega_s t + \frac{1}{3}\sin 3\omega_s t + \frac{1}{5}\sin 5\omega_s t + \cdots\right].$$

A similar procedure with $x(t) = 0$, $y(t) = Y\cos(\omega t + \phi)$ gives expressions for the terms N_{21}, N_{22}.

As mentioned earlier stability should be investigated for all values of ω and ω_s that are commensurate. This involves obtaining the matrix of describing functions for each case and examining for any intersection of the Gershgorin band and the inverse Nyquist for the frequency point ω concerned. If this is done it is found that the cases where instability is most likely to occur are i) at an unrelated frequency (where the band must not enclose any part of the plot) and ii) when $\omega = \omega_s$ (where the 80 Hz point must not be enclosed).

i) At unrelated frequency

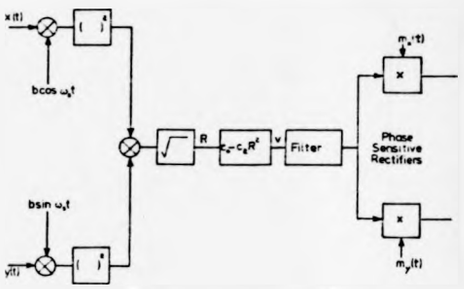


Fig. 6. Form of nonlinearity used to obtain the describing function matrix.

$$N = -c_2 \cdot \frac{2b}{\pi}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

ii) at the frequency $\omega = \omega_s$

$$N = -c_2 \cdot \frac{2b}{\pi} \cdot \frac{1}{6}\begin{vmatrix} (A + jB) & (B - jA) \\ (B - jA) & (A + jB) \end{vmatrix}$$

where

$$A = 6 + 3\frac{Z}{b}\cos\phi - 2\cos 2\phi - \frac{Z}{b}\cos 3\phi$$

$$B = 3\frac{Z}{b}\sin\phi + 2\sin 2\phi + \frac{Z}{b}\sin 3\phi$$

and Z is the amplitude of the appropriate input signal. The G matrix is given by

$$G = \begin{bmatrix} \dfrac{k_x\omega_{nx}^2}{S\left(S^2 + 2\zeta_x\omega_{nx}S + \omega_{nx}^2\right)} & 0 \\ 0 & \dfrac{k_y\omega_{ny}^2}{S\left(S^2 + 2\zeta_y\omega_{ny}S + \omega_{ny}^2\right)} \end{bmatrix}$$

The plot of $1/G_{11}$ is shown in Fig. 7. As the describing function matrix at unrelated frequency is diagonal with the elements constant then the $-N_{11}$ plot is a point on the real axis with no surrounding band.

The describing function $-N_{11}$ for $\omega = \omega_s$ is also drawn and also circles of radius $|N_{21}|$ for a range of values of 2, 0–0.3 mm, this being a more convenient method than drawing continuous bands. As can be seen the bands are well away from the ω_s point (80 Hz) on the plot. Stability is restricted by the enclosure of the point for unrelated frequency. The transient response for this condition has been given in Fig. 4, instability occuring at a frequency of 31.5 Hz as the gain is increased, since N is purely real for unrelated frequencies.

III. System Compensation

A. Addition of Local Velocity Feedback

As the limit to stability is caused by the peak in the frequency respons. a first step in compensation is to reduce this peak. This can be done by increasing the damping factor of the second-order system associated with the suspension system. This was done by adding pick-up coils to the platform carrying the gauged plates. These moved between the poles of permanent magnets and the EMF induced into them fed back into the appropriate inputs of the amplifiers supplying the deflection coils. This velocity feedback increased the damping factor of the suspension system and the resulting inverse Nyquist plot for $1/G_{11}$ shown in Fig. 8, in which gain has been increased to give the same intersection of the real axis as in Fig. 7. The resulting transient response is shown in Fig. 9 and the settling

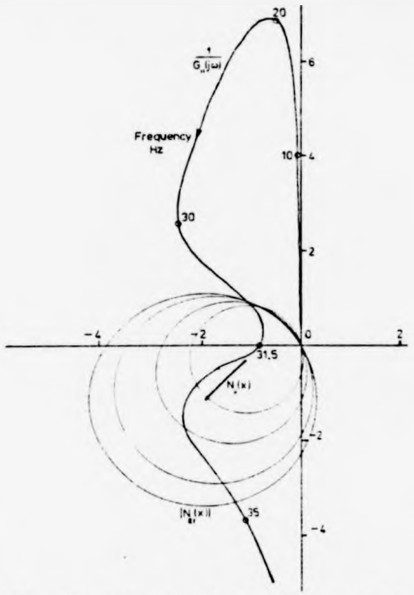


Fig. 7. Inverse Nyquist plot and Gershgorin bands for uncompensated system.

time has been reduced to the order of 0.1 second (responses are shown for $\zeta = 1$). However, as is shown in the plot of Fig. 8, the 80 Hz point is now much closer to the bands for $|N_{21}|$ and a small amount of 80 Hz component is now present in the transient of Fig. 9. If the gain is increased until instability occurs a definite 80 Hz component now appears in the resulting response.

B. Other Compensation Techniques

Two techniques have been investigated for introducing further improvement to the dynamic performance of the system.

Inspection of Fig. 8 suggests that overall velocity feedback, (from derivative of system output to the input of the integrator) may be advantageous, by raising the Nyquist diagram vertically, and subsequently permitting a higher loop gain. This is not the case, however, since the instability at 80 Hz becomes predominant in determining system performance.

Secondly, an attempt to reduce the loop gain to zero at 80 Hz has been made. This modification can readily be incorporated by adding a sample-and-hold unit, operating at perturbation frequency [4]. In this particular system, the additional phase lag due to the modification has a sufficiently detrimental effect to give no overall improvement.

IV. Conclusions

The results obtained by measurement of transients on the original instrument were in close agreement with those predicted by simulation. The predicted effect of velocity feedback was found to be substantiated in practice a significant improvement in system performance being obtained.

Although the addition of sample and hold circuits gave no improvement in the simulation a small improvement was obtained in practice. This was probably due to the fact that a component at perturbation frequency was present in the signal paths due to factors such as unbalance in the phase sensitive detectors and pick up from the fields of the deflection coils providing the perturbation. These factors were not included in the simulation and the suppression of their effects would have a practical significance.

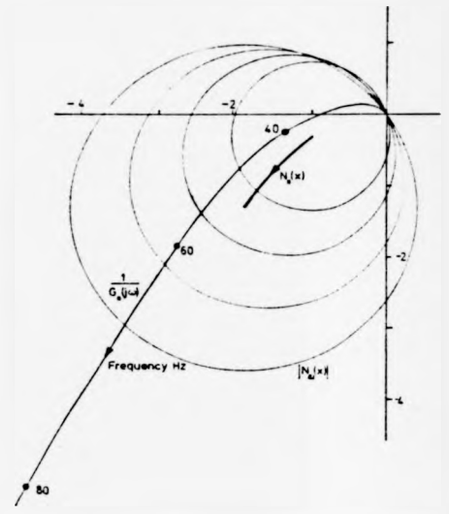


Fig. 8. Inverse Nyquist plot and Gershgorin bands for compensated system.

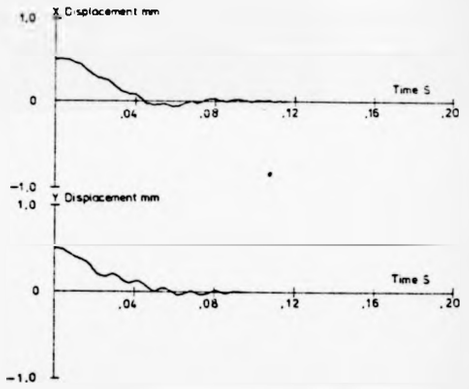


Fig. 9. Transient response of compensated system.

The overall effect of the compensation described was to reduce the response time from 0.9 s to just under 0.1 s. As explained in the introduction this is a very important reduction. For the watch bearing plate containing twenty holes this would reduce the total gauge in time per plate from 18 s to 2 s. This would not be the reduction obtained in total job time as the time to change plates remains unaltered. Even so it was felt that the instrument would be well worth while modifying to incorporate this form of controller.

References

[1] L. Balmer, A. H. Falkner, and W. N. Tabemer, "Novel electro-optical techniques in metrology," *Electron. Power*, vol. 22, pp. 27–31, 1976.

[2] A. Mees, "Describing functions, circle criteria and multiple loop feedback systems," *Proc. IEE*, vol. 120, pp. 126–130, 1973.

[3] J. C. West, J. L. Douce, and R. K. Livesley, "The dual input describing function and its use in the analysis of non-linear feedback systems," *Proc. IEE*, vol. B-103, pp. 463–474, 1956.

[4] J. L. Douce and A. D. Bond, "The development and performance of a self-optimising system," *Proc. IEE*, vol. 110, 1963.

Systems Science

Vol. 8 1982 No. 2–3

JOHN L. DOUCE*, LESLIE BALMER**

BIAS OF FREQUENCY RESPONSE ESTIMATES

This paper considers the bias of frequency domain measurements when the time of observation is limited. It is shown that estimates of auto- and cross-spectral densities are subjected to bias, so that estimates of frequency response functions will exhibit, in general, this systematic error. Simulation results confirm analytic predictions for particular system.

LIST OF PRINCIPAL SYMBOLS

$x(t), y(t)$ – Input and output time functions;
$X(j\omega)$ – Fourier coefficients of $x(t)$;
$\tilde{G}(j\omega)$ – Estimate of $G(j\omega)$ based on one block;
$\bar{G}(j\omega)$ – Estimate based on several blocks;
$R_{xy}(\tau)$ – Cross correlation function;
$\varphi_{xy}(\omega)$ – Cross spectral density;

I. INTRODUCTION

There are many situations in which only short-term records of system variables are available for analysis. For example data collection under closely controlled experimental conditions may be difficult to achieve over long periods. Another particular case being investigated by the authors is rapid adaptive control using frequency domain techniques. Here, a potentially time-varying system is to be identified and controlled, with small settling time.

In general, this limited period of observation leads to estimation errors. The variance of the estimate increases in general as the period of observation

* Warwick University, Great Britain.
** Coventry Polytechnic, Great Britain.

decreases. This paper considers the bias of the estimates, since this has been not widely studied and it places a fundamental limit on the accuracy achievable when the number of fixed-length blocks increases indefinitely.

II. EVALUATION OF THE BIAS

We consider the situation in which a linear system is subjected to a statistically stationary random signal, and records of input and output are available over only a limited time interval $t = 0$ to T.

The records of input and output data $x(t)$ and $y(t)$ $0 \leqslant t \leqslant T$ respectively have Fourier components

$$X(j\omega) \text{ and } Y(j\omega), \quad \omega = (2\pi n/T),\ n = 0, \infty.$$

The estimate of the frequency response function based on one block is

$$\hat{G}(j\omega) = \hat{\varphi}_{xy}/\hat{\varphi}_{xx} = Y(j\omega)/X(j\omega)$$

and if independent blocks are available by

$$\bar{G}(j\omega) = \bar{\varphi}_{xy}(j\omega)/\bar{\varphi}_{xx}(j\omega).$$

Independent disturbances present in the measured output $y(t)$ increase the variance of the above estimates, but do not introduce additional bias, hence they will be ignored in this subsequent analysis.

For tutorial purposes, we first present a simplified analysis, valid when the input is white noise. Secondly a result is derived for the general case when the number of blocks analysed is large, so that the expected value of the estimate based on

$$\bar{G}(j\omega) = \bar{\varphi}_{xy}/\bar{\varphi}_{xx}$$

is given by

$$E\{\bar{G}(j\omega)\} = E\{\bar{\varphi}_{xy}\}/E\{\bar{\varphi}_{xx}\}.$$

Consider the case of a single block where the frequency component of interest in the input is of frequency ω, and over the period $t = 0$ to T is given by

$$x(t) = A\sin(\omega t - \theta).$$

A is a random variable, of unspecified amplitude distribution, and θ is uniformly distributed over the range $0 \leqslant \theta < 2\pi$.

The response of the linear system of impulse response $g(t)$ is given by the convolution integral:

$$y(t) = \int_0^t g(\tau)\cdot x(t-\tau)\,d\tau.$$

We ignore other components in the input signal, assuming they are independent of the component considered. Similarly, the response of the system for $t > 0$ to the input signal for $t < 0$ is ignored, since this response is independent of, and hence uncorrelated with, the input signal considered, for 'white' excitation.

Fourier analysis of the response over the time interval 0 to T, with respect to $\sin(\omega t + \theta)$ and $\cos(\omega t + \theta)$ gives in-phase and quadrature components R and I.

$$R = \frac{A}{T}\int_0^T (T-\tau)g(\tau)\cos\omega\tau d\tau + \frac{A}{\omega T}\cos 2\theta \int_0^T g(\tau)\sin\omega\tau d\tau,$$

$$I = -\frac{A}{\omega T}\int_0^T (T-\tau)g(\tau)\sin\omega\tau d\tau + \frac{A}{\omega T}\cos 2\theta \int_0^T g(\tau)\cos\omega\tau d\tau.$$

Dividing these expressions by the input amplitude A gives the estimates of the real and imaginary components of the gain $G(j\omega)$.

Taking ensemble averages with respect to θ gives the expected value of these estimates and in each case the contribution of the final term in the above expressions is zero.

We assume that the observation time T significantly exceeds the settling time of the system, so that the upper limit of integration can be taken as infinity.

Since

$$\int_0^\infty tf(t)\varepsilon^{-st}dt = \frac{d}{ds}F(s)$$

where $F(s)$ is the Laplace transform of $f(t)$, the expected value of the estimate is

$$E\{Re(\hat{G})\} = Re(G) + \frac{1}{T}Re\left[\frac{dG(s)}{ds}\right]_{s=j\omega},$$

$$E\{I_m(\hat{G})\} = I_m(G) + \frac{1}{T}I_m\left[\frac{dG(s)}{ds}\right]_{s=j\omega}.$$

Letting $G'(j\omega) = \left[\frac{dG(s)}{ds}\right]_{s=j\omega}$

the above expressions may be written

$$E\{\hat{G}(j\omega)\} = G(j\omega) + \frac{1}{T}G'(j\omega).$$

The more general case, in which the input has an unknown power density spectrum, can be analysed for the case defined earlier, in which the number of blocks is large. Specifically, the number of blocks must be large enough so that the variance of estimates of the spectral density functions is sufficiently small so that the expected value of the estimate, given by

$$E\{\bar{G}(j\omega)\} = E\{\bar{\varphi}_{xy}(j\omega)/\bar{\varphi}_{xx}(j\omega)\}$$

can be replaced by

$$E\{\bar{G}(j\omega)\} = E\{\bar{\varphi}_{xy}(j\omega)\}/E\{\bar{\varphi}_{xx}(j\omega)\}.$$

The estimates of spectral densities based on blocks of duration T are biased. The bias may be evaluated either by elementary Fourier analysis or by noting that the estimate is based on the original time function multiplied by the rectangular time window of duration T.

In the time domain, this window modifies the estimated correlation functions according to

$$E\{\bar{R}(\tau)\} = \left(1 - \left|\frac{\tau}{T}\right|\right) R(\tau) \text{ for } |\tau| < T$$

$$= 0 \text{ for } |\tau| > T.$$

Taking the Fourier transform of the right hand side gives the well-known result that the windowing operation corresponds to convolution in the frequency domain with the function

$$\left(\frac{\sin \omega T/2}{\omega T/2}\right)^2.$$

Returning to the expression above for the expected value of the correlation function, the effect of the window is, in the frequency domain,

$$E\{\bar{\varphi}(\omega)\} = \int_{-T}^{T} \left(1 - \frac{|\tau|}{T}\right) R(\tau)\, \varepsilon^{-j\omega\tau} d\tau.$$

In all normal experimental conditions, the block length T is much longer than the settling time of the system. This allows the limits of integration to be replaced by $\pm\infty$. The first term,

$$\int_{-\infty}^{\infty} R(\tau)\, \varepsilon^{-j\omega\tau} d\tau,$$

gives the true value of the power spectral density, whilst the second term

$$-\frac{1}{T}\int_{-\infty}^{\infty}|\tau|R(\tau)\varepsilon^{-j\omega\tau}d\tau$$

represents the bias component in the estimate.

A useful form of the above equation is obtained by splitting $R(\tau)$ into even and odd parts denoted $R_e(\tau)$ and $R_0(\tau)$ defined by

$$R_e(\tau) = R(\tau)+R(-\tau),$$

$$R_0(\tau) = R(\tau)-R(-\tau).$$

With this notation, the power spectral density is given by

$$\varphi(\omega) = \int_0^{\infty} R_e(\tau)\cos\omega\tau d\tau - j\int_0^{\infty} R_0(\tau)\sin\omega\tau d\tau.$$

We define $\bar{\Phi}(\omega)$ by

$$\bar{\Phi}(\omega) = \bar{\Phi}_R(\omega)+j\bar{\Phi}_i(\omega) = \int_0^{\infty} R_0(\tau)\cos\omega\tau d\tau - j\int_0^{\infty} R_e(\tau)\sin\omega\tau d\tau$$

and elementary analysis shows that the bias term is given by

$$-\frac{1}{T}\int_{-\infty}^{\infty}|\tau|\varepsilon^{-j\omega\tau}d\tau = \frac{-j}{T}\frac{d\bar{\Phi}(\omega)}{d\omega}.$$

In the special cases such that $R(\tau) = 0$ for $\tau < 0$, $\bar{\Phi}(\omega) = \varphi(\omega)$, and the bias is as given in the previous section.

III. EXAMPLES

Three examples illustrate the previous analysis.

(a) As a trivial case, consider a pure time delay τ subjected to 'white' noise, and the input and output records analysed over an observation time $0 < t < T$.

Evidently the output over the time $0 < t < \tau$ is uncorrelated with the measured input signal. The estimate of the gain is hence $(1-\tau/T)$ rather than of unity modulus, and the expected value of the frequency response function is $(1-\tau/T)\varepsilon^{-j\omega\tau}$. This gives a circle on the complex plane, of radius $(1-\tau/T)$.

(b) We next consider a system of transfer function $1/1+sT_s$ subjected to

'white' noise, with the frequency response estimated using a block length of duration T.

The bias in the estimate is equal to

$$\frac{1}{T}\left[\frac{dG(s)}{ds}\right]_{s=j\omega} \text{ or } -\frac{T_s}{T}\frac{1}{(1+j\omega T_s)^2}.$$

In this special case, this can be expressed as

$$\frac{-T_s}{T}(G(j\omega))^2.$$

Thus the phase angle of the bias term is minus twice the phase angle of the true response function. Since the true frequency response locus on the complex plane is a semi circle, the bias term is, by elementary trigonometry, always directed towards the centre of the circle, as in Figure 1.

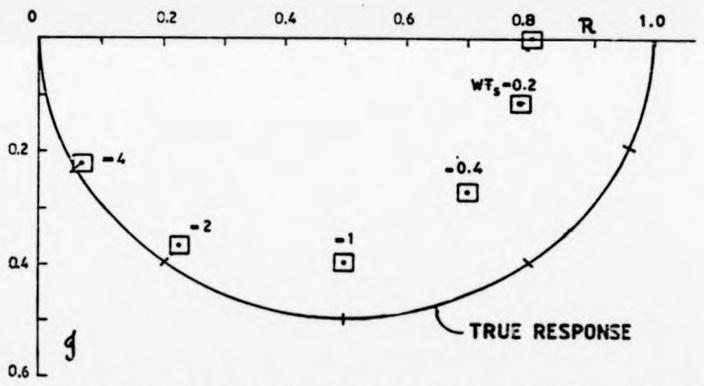


Fig. 1. Response of first-order system. Continuous curve is true response. Points are measurements based on $T = 5T_s$

(c) Finally, consider the case in which the system input signal has the power spectral density

$$\varphi_{xx}(\omega) = \frac{1}{1+\omega^2 T_s^2}$$

and the transfer function of the system is $G(jw) = \frac{1}{1+sT_s}$.

In this case, since the short-term estimate of the input spectrum is biased, we can obtain an analytic expression only for the bias of frequency response estimates based on a large number of blocks, such that

$$E\{\bar{G}(j\omega)\} = E\{\bar{\varphi}_{xy}\}/E\{\bar{\varphi}_{xx}\}.$$

Analysis shows that the terms on the right-hand side are given by

$$E\{\bar{\varphi}_{xx}\} = \varphi_{xx} - \frac{T_s}{T_0}\frac{(1-\omega^2T_s^2)}{(1+\omega^2T_s^2)^2}$$

and

$$E\{\bar{\varphi}_{xy}\} = \varphi_{xy} - \frac{T_s}{2T_0}\left\{\frac{1}{(1+\omega^2T_s^2)^3}\right\}\{(3-6\omega^2T_s^2-\omega^4T_s^4) - j2\omega T_s(3-\omega^2T_s^2)\}.$$

Taking the ratio of these two expressions using a binomial expansion for the denominator gives

$$E\{\bar{G}(j\omega)\} \approx G(j\omega) + \frac{T_s}{T_0(1+\omega^2T_s^2)^2}\{(-0.5+2\omega^2T_s^2+\frac{1}{2}\omega^4T_s^4) + j\omega T_s\cdot 2\}.$$

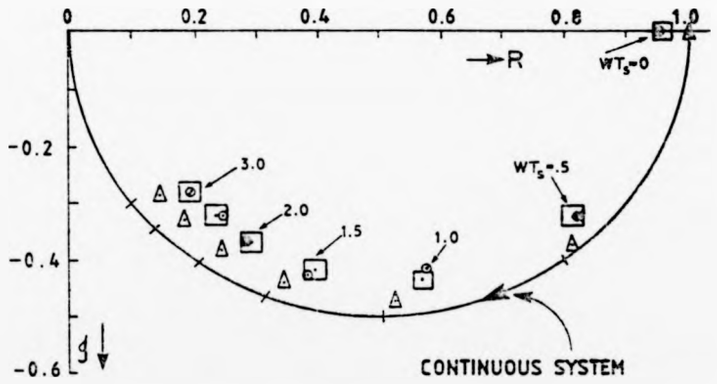


Fig. 2. System with non-white exitation
△ true response of sampled-data system
▷ measured response
□ predicted measurements

Figure 2 compares theoretical predictions shown as squares on the diagram with computer simulations. The simulated results were obtained as the average of 30,000 distinct blocks. In the case considered $T_0 = 12.8\ T_s$.

It is noteworthy that, for this case, the bias is of equal magnitude for $\omega \to 0$ and $\omega \to \infty$, and is given by $\pm T_s/2T_0$.

IV. CONCLUSIONS

Analysis, supported by some simulation results, shows that short-term estimates of auto- and cross-spectral densities contain a bias term which in turn causes resulting estimates of frequency response functions to be biased.

Several further research problems are suggested by this work. The effect of time windows is known to reduce leakage and the effect of various windows on the bias should be investigated. It is not obvious that, in this application, the time window should be symmetrical, and the effect of a differential time shift on input and output records should be investigated. The bias of estimates of elements within a feedback system also is worthy of attention. Finally, the estimates may themselves be used to estimate the bias present, so that a better approximation to the required statistics may be calculated.

REFERENCES

[1] Akaike H., Yamanouchi Y., *On the statistical estimation of frequency response functions*, An. Inst. Statistical Mechanics, Vol. 14, No. 1, 1962, p. 23.

[2] Wellstead P. E., *Real-time spectral analysis*, Ph. D. thesis, University of Warwick, 1971.

[3] Douce J. L., *Bias of frequency-response estimates in closed-loop systems*, Proc. IEE, Vol. 127, D, No. 4, 1980, p. 149.

Received July 7, 1981

ОШИБКА В ОЦЕНКЕ РЕАКЦИИ В ОБЛАСТИ ЧАСТОТЫ

В работе рассматривается измерительное отклонение в области частоты при ограниченном времени наблюдения. Показано, что авто- и взаимная спектральная плотность зависит от отклонения, следовательно, оценка реакции в области частоты покажет, в общем случае, симметричную ошибку. Моделирующие исследования подтверждают аналитические результаты для типичных систем.

FREQUENCY RESPONSE ESTIMATION USING SHORT DATA BLOCKS

Douce, John L.*
Balmer, Leslie†

*Warwick University, UK.
†Coventry Polytechnic, UK.

A method of obtaining system frequency response estimates from short records of input and output data is examined. This method makes allowance for the transient effects due to the short record length and examples given show an improvement on more conventional methods of frequency response estimation.

INTRODUCTION

In a previous paper [1] the authors examined the errors which arise in the estimation of the frequency response function of systems using short observation periods. It has been shown that conventional frequency domain techniques produce estimates of the system frequency response which are biased and subject to significant variance, even in the noise free case. The magnitude of these errors can be evaluated in terms of the transient effects at the ends of the record.

In this paper a method of reducing these errors is examined. This involves parametrising the system frequency response function, $G(j\omega)$, and relating these parameters to the short term spectral estimates of the input and output signals. End effects are allowed for by the introduction of additional parameters. The system parameters and those associated with end conditions can be evaluated using as data the complex spectral estimates of input and output signals. Other authors [2,3] have applied regression analysis to frequency response data but they have assumed the observation time is such as to make end effects negligible.

The technique proposed here is related to established methods for simultaneous state and parameter estimation in the time domain.

LIST OF SYMBOLS

$x(t)$, $y(t)$, $u(t)$. System input, output and noise signals $X(j\omega)$, $Y(j\omega)$, $N(j\omega)$. Short term Fourier Transforms of the above signals, defined by

$$X(j\omega) = \int_0^T x(t)e^{-j\omega t}dt$$

where T is the observation time. Similarly for $Y(j\omega)$, $N(j\omega)$

$g(t)$ System impulse response.

$G(s)$ System transfer function.

$B(s)$, $A(s)$ Numerator and denominator of $G(s)$

$$B(s) = b_o + b_1 s + b_2 s^2 + \text{————————} b_M s^M$$

$$A(s) = a_o + a_1 s + a_2 s^2 \text{————————} a_N s^N$$

where $M < N$

FREQUENCY RESPONSE FUNCTION FROM SHORT TERM ESTIMATES

Consider the system shown in Fig. 1. Here $x(t)$ is the statistically stationary input to the system measured over observation time $t = o$ to T. $x_R(t)$ is a periodic signal, equal to $x(t)$ over the observation time with

$$x_R(nT+t) = x(t) \quad n \text{ integer}, \quad 0 < t < T$$

Since $x_R(t)$ is periodic, the response $y_R(t)$ to this input is also periodic and the short-term Fourier Transforms of these signals at harmonic frequencies ω_i ($\omega_i = i\,2\pi/T$) are related by

$$X_R(j\omega_i) = X_R(j\omega_i) \cdot G(j\omega_i) + N(j\omega_i) \qquad (1)$$

Over the time of observation the response $y(t)$ differs from $y_R(t)$ due to the different initial states at $t = o$. Assuming for the moment that the roots λ_n of $A(s) = 0$ are all real and distinct, then the system response due to these initial states may be written

$$y(t)-y_R(t) = \sum_{n=1}^{N} K_n e^{\lambda_n t}$$

where K_n depends on the individual experiment.

Taking short-term Fourier Transforms

$$Y(j\omega_i)-Y_R(j\omega_i) = \sum_{n=1}^{N} K_n \int_o^T e^{(\lambda_n - j\omega_i)t} dt$$

$$= \sum_{n=1}^{N} \frac{K_n}{(\lambda_n - j\omega_i)} (e^{\lambda nT}-1)$$

$$\text{Writing } C_n = K_n(e^{\lambda_n T}-1)$$

$$Y(j\omega_i) - Y_R(j\omega_i) = \sum_{n=o}^{N} \frac{C_n}{(\lambda_n - j\omega_i)}$$

Since λ_n are roots of $A(s) = 0$

$$Y(j\omega_i) - Y_R(j\omega_i) = \frac{h_o + h_1(j\omega_i) + \text{————} h_{n-1}(j\omega_i)^{N-1}}{A(j\omega_i)}$$

$$= \frac{H(j\omega_i)}{A(j\omega_i)}$$

where h_o ———— h_{N-1} are constants.

Hence

$$Y(j\omega_i) = \frac{B(j\omega_i)}{A(j\omega_i)} \cdot X(j\omega_i) + \frac{H(j\omega_i)}{A(j\omega_i)} + N(j\omega_i) \qquad (2)$$

The result also holds for multiple and complex roots of $A(s) = 0$.

This expression forms the basis of the method of frequency response estimation. The additional N parameters in $H(j\omega)$ are introduced to allow for the transient effects.

REGRESSION ANALYSIS

The problem of obtaining the constants b_m, a_n, from the measured spectra $Y(j\omega_i)$ and $X(j\omega_i)$ can be treated as a problem in regression analysis [4]. The required parameters are those that minimise the expression

$$\sum_i \left| Y(j\omega_i) - \frac{B(j\omega_i)}{A(j\omega_i)} X(j\omega_i) - \frac{H(j\omega_i)}{A(j\omega_i)} \right|^2 \qquad (3)$$

where the summation is taken over the harmonic frequencies in the Fourier Transform. Since the form of the expression is such that differentiation with respect to the A parameters will not produce a linear set of equations, the problem is one of non-linear regression [5].

Equation (3) can be re-arranged to produce a linear regression problem by multiplying through the $A(j\omega_i)$. The expression to be minimised becomes

$$\sum \left| Y(j\omega_i)A(j\omega_i) - B(j\omega_i)X(j\omega_i) - H(j\omega_i) \right|^2$$

However the noise term is now $A(j\omega_i)N(j\omega_i)$ and this produces biased estimates [4]. The solution to this modified formulation is useful however to provide initial estimates for an iterative procedure for the non-linear regression.

Writing $$F = Y(j\omega)A(j\omega) - B(j\omega)X(j\omega) - H(j\omega)$$

then $$\frac{\partial |F|^2}{\partial a_n} = \frac{\partial (FF^*)}{\partial a_n}$$

$$= \frac{\partial F^*}{\partial a_n} F + \frac{\partial F}{\partial a_n} F^*$$

$$= 2 \text{ Re } \{ F \frac{\partial F^*}{\partial A^*(j\omega)} \cdot \frac{\partial A^*(j\omega)}{\partial a_n} \}$$

Similarly for h_n and b_n.

Setting this derivative to zero gives the following set of equations

$$\begin{array}{c|c|c|c|c|c} & q = o \text{ to } M & & q = o \text{ to } N-1 & & q = o \text{ to } N \\ \hline p = o \text{ to } M & \sum_i \text{Re } \{|X_i|^2 Z_i\} & & \sum_i \text{Re } \{X_i^* Z_i\} & \sum_i \text{Re } \{-Y_i X_i^* Z_i\} & \begin{bmatrix} b_p \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix} \\ \hline p = o \text{ to } N-1 & \sum_i \text{Re } \{X_i Z_i\} & & \sum_i \text{Re } \{Z_i\} & \sum_i \text{Re } \{-Y_i Z_i\} & \begin{bmatrix} h_p \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix} \\ \hline p = o \text{ to } N & \sum_i \text{Re } \{-X_i Y_i^* Z_i\} & & \sum_i \text{Re } \{-Y_i^* Z_i\} & \sum_i \text{Re } \{|Y_i|^2 Z_i\} & \begin{bmatrix} a_p \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix} \end{array}$$

where $Z_i = (-1)^p (j\omega_i)^{p+q}$

To solve these equations the value of one of the parameters is taken as unity. The column containing this parameter is taken to the right-hand side of the expression and the row corresponding to differentiation with respect to this parameter is eliminated.

To solve the non-linear regression problem some form of hill-climbing routine is required, [5] contains a review of methods suitable for regression problems and also contains an extensive list of further references. However the expression given by eqn. (3) does lead to a linear set of equations when the differentiated with respect to parameters b_m, h. Hence any iterative procedure need only be applied to the parameters a_n. For a given set of a_n parameters the linear equations can be solved for b_m and h and the expression in eqn. (3) calculated. This has the advantages that the initial estimates for the starting point are only required for the reduced number of parameters and also convergence appears to be faster and more stable because of the reduced dimension of the problem [6]. One disadvantage is that it is impractical to obtain analytical derivatives in the iteration routine and these must be replaced by finite differences. The equations to be solved for a given set a are now given by

$$\begin{array}{c|c|c|c} & q = o \text{ to } M & q = o \text{ to } N-1 & q = o \text{ to } N \\ \hline p = o \text{ to } M & \sum_i \text{Re } \left\{ \left| \frac{X_i}{A_i} \right|^2 Z_i \right\} & \sum_i \text{Re } \left\{ \frac{X_i^*}{|A_i|^2} Z_i \right\} & \begin{bmatrix} b_p \end{bmatrix} = \begin{bmatrix} \sum_i \text{Re } \left\{ \frac{X_i^* Z_i}{A_i^*} \right\} \end{bmatrix} \\ \hline p = o \text{ to } N-1 & \sum_i \text{Re } \left\{ \frac{X_i}{|A_i|^2} Z_i \right\} & \sum_i \text{Re } \left\{ \frac{1}{|A_i|^2} Z_i \right\} & \begin{bmatrix} h_p \end{bmatrix} = \begin{bmatrix} \sum_i \text{Re } \left\{ \frac{Y_i}{A_i^*} \right\} \end{bmatrix} \end{array}$$

If one of the parameters is taken as reference the appropriate row is deleted and the appropriate

column taken to the right-hand side. Alternatively one of the a parameters can be taken as reference reducing the number of dimensions in the hill-climb routine.

EXAMPLES

These examples illustrate the applications of the technique described and compare the resulting frequency response function with that obtained by conventional methods. In both cases the results have been obtained by digital simulation with a sampling interval sufficiently short to give insignificant difference between the frequency response of the continuous and sampled systems over the frequency range considered

(a) A noise-free system with the transfer function

$$G(s) = \frac{0.36}{s^2 + 0.024s + 0.36}$$

This represents a very lightly damped system with $\omega_n = 0.6$ and $\zeta = 0.02$. The decay time constant is 83.3 seconds and results are shown in Fig. 2 for an observation time of 102.4 seconds. This figure shows the measured response using the method of this paper and using the ratio output/input frequency spectra with a Hanning window. As can be seen, the latter method gives a spread of points about the correct curve. This spread could be reduced by taking more than one block and combining the spectra of the windowed estimates X_H, Y_H according to the estimator

$$\hat{G}(j\omega) = \frac{\sum X_H^* Y_H}{\sum X_H^* X_H}$$

The resulting estimates would however still show bias as discussed in [1].

(b) A third order system with a transfer function G(s) having multiple roots

$$\hat{G}(s) = \frac{1}{(s+1)^3} = \frac{1}{s^3 + 3s^2 + 3s + 1}$$

The input signal was band limited and a noise signal having a uniform spectrum was added at the output. An observation time of 25.6 seconds was used and the parametrised model had the form

$$G(s) = \frac{1}{a_3 s^3 + a_2 s^2 + a_1 s + a_o}$$

Fig. 3 shows ensemble averages for the parameter estimates as the ratio noise power/signal at the output is increased. Results are shown for linear and non-linear regression methods.

CONCLUSIONS

It has been shown that by taking into account the transients due to end conditions it is possible to obtain frequency response estimates superior to those obtained by conventional techniques, even when the observation time is much shorter than the settling time of the system. The disadvantage of the method is that it requires a parametric model which is not required in conventional frequency response estimation. However it appears that even if the order of the model is incorrect results for short term estimates are still superior to those obtained by conventional methods.

The results presented have applied only to open-loop systems but work is in progress to extend the analysis to include closed loop systems.

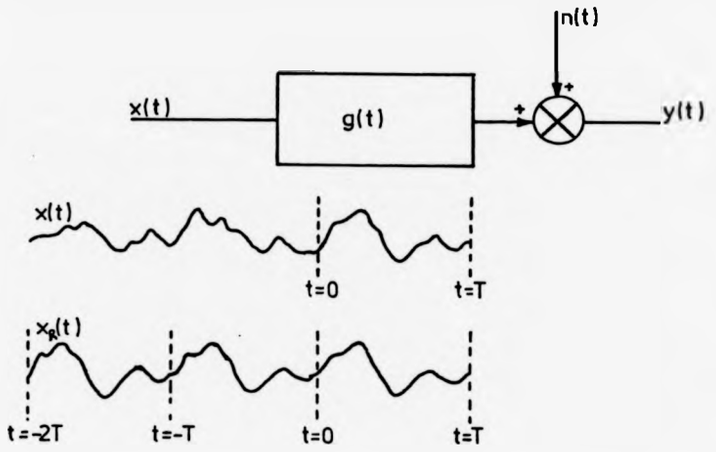


Figure 1. System showing signals x(t), $x_R(t)$.

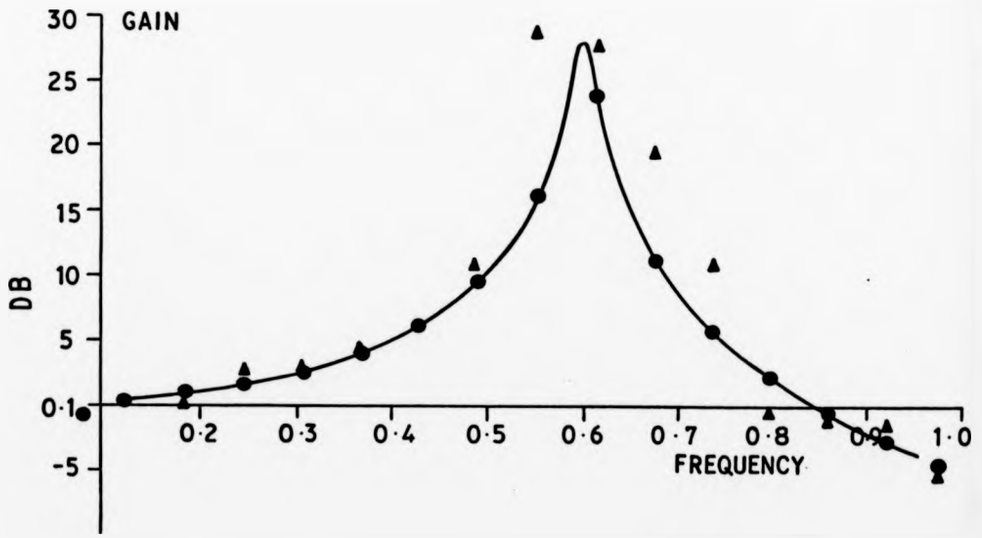


Figure 2. Lightly damped second order system.

———— theoretical response

● ● ● ● ● points obtained by regression method

▲ ▲ ▲ ▲ ▲ points obtained by classical method

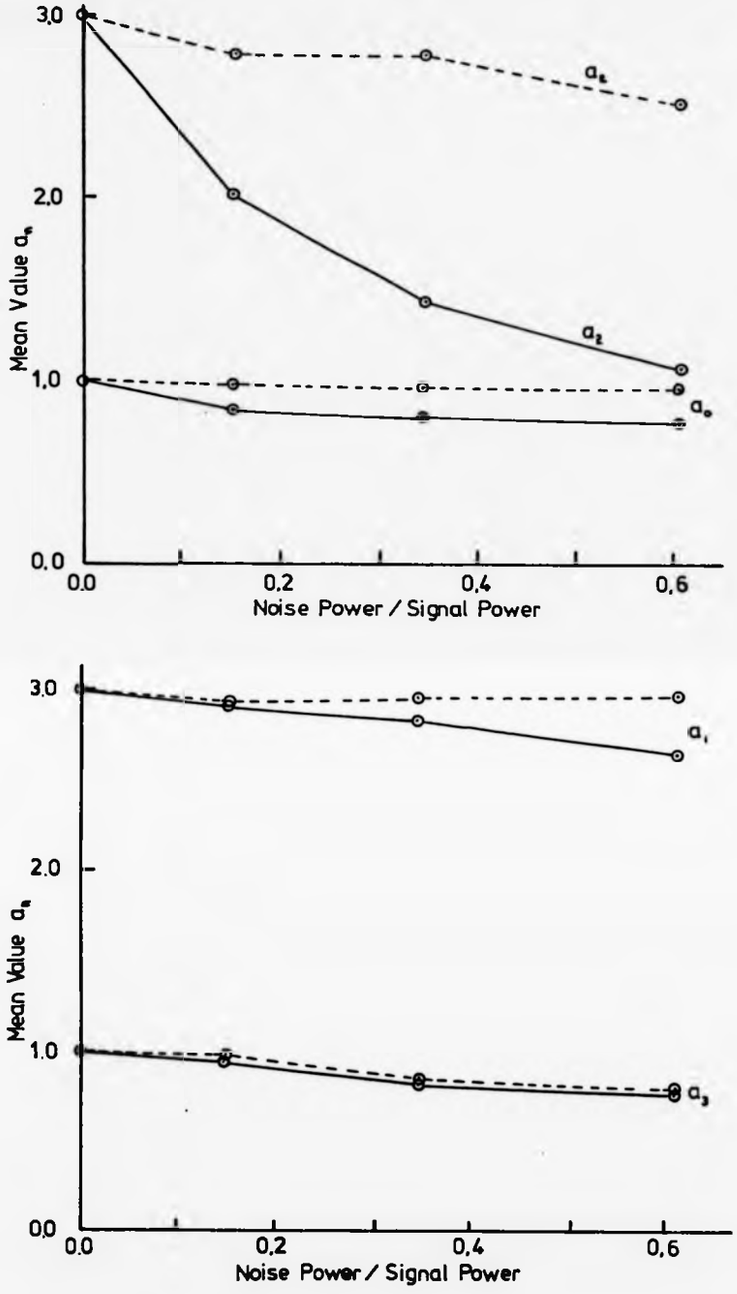


Figure 3. Variation in the mean value of parameter a_n with noise power.

O——O——O linear regression

O----O----O non-linear regression

REFERENCES

[1] Douce, J.L. and Balmer, L. "Transient Effects in Spectrum Estimation". Submitted for publication.

[2] Levy, E.C. "Complex Curve Fitting". I.R.E. Transactions Automatic Control 4 (1959) May, p.37-43.

[3] Marchand, M. "The Identification of Linear Multivariable Systems from Frequency Response Data". Proc. of 3rd I.F.A.C. Symposium, June 1973. Paper TC-1, pp. 591-598.

[4] Draper, N.R. and Smith, H. "Applied Regression Analysis". John Wiley & Son, 1966.

[5] Bard, Y. "Nonlinear Parameter Estimation". Academic Press, 1974.

[6] Lawton, W.H. and Sylvestre, E.A. "Elimination of Linear Parameters in Nonlinear Regression". Technometrics, Vol. 13, No. 3, Aug. 1971.

Transient effects in spectrum estimation

Prof. J.L. Douce, D.Sc., Ph.D., C.Eng., F.I.E.E., Sen.Mem.I.E.E.E., and L. Balmer, M.Sc., C.Eng., M.I.E.E:

Indexing terms: *Errors and error analysis, Spectrum estimation*

Abstract: The paper examines the errors in spectrum and frequency response estimation when short blocks of data are used. General results for the bias and variance of spectral estimates are presented. Analysis is restricted to stable time-invariant linear systems with stationary white excitation.

List of symbols

$x(t)$, $y(t)$ = input and output signals
$g(t)$, $G(s)$ = system impulse response and transfer function, respectively
$X(j\omega)$ = Fourier representation of $x(t)$ over $0 < t < T$

$$= \int_0^T x(t)e^{-j\omega t}\,dt$$

$\hat{\Phi}_{xy}(j\omega)$ = measured cross-spectrum $= (1/T)XY^*$
$R_{xy}(t_1, t_2)$ = crosscorrelation function $= E\{x(t_1) \cdot y(t_2)\}$
$\Phi_{xy}(j\omega)$ = true cross-spectrum
$\bar{\Phi}(j\omega)$ = smoothed measured spectrum

1 Introduction

When a record of finite duration T is available for signal analysis or system identification, errors are introduced into the estimates of the statistical properties of the processes. These lead to systematic errors or bias and to variance in resulting estimates of the frequency response of systems.

Attention is directed in this paper to the errors introduced by end effects or transients, which can be particularly significant when the length of the record or of a single block does not greatly exceed the settling time of the system. The important effects of data windows and smoothing will not be covered in detail [1–3].

The analysis follows the presentation by Papoulis [4]. An alternative derivation of the main results from first principles is given, which leads to some detailed results and aids physical insight into the mechanisms involved.

The statistical properties of auto- and cross-spectral estimates based on a single block of data are examined. From this, the properties of smoothed estimates, smoothed by averaging over several blocks and/or over adjacent frequencies, can be derived.

Given a sample of the process $x(t)$ over the time range $0 < t < T$, we define the measured Fourier transform as

$$X(j\omega) = \int_0^T x(t)e^{-j\omega t}\,dt$$

for $\omega = 2\pi n/T$, with n integer.

The measured autospectrum $\hat{\Phi}_{xx}(j\omega)$ is defined by

$$\hat{\Phi}_{xx}(j\omega) = \frac{1}{T}X(j\omega)X^*(j\omega)$$

For stationary processes, the expected value of this function is related to the autocorrelation function of $x(t)$ by

$$E\{\hat{\Phi}_{xx}(j\omega)\} = \int_{-T}^{T}\left(1 - \frac{|\tau|}{T}\right)R_{xx}(\tau)\,e^{-j\omega\tau}\,d\tau \quad (1)$$

Although the response $y(t)$ of a time-invariant linear system to the stationary time function $x(t)$ is also stationary, the response to $x(t)$ for $t < 0$, uncorrelated with $x(t)$ for $t > 0$, is considered separately. This requires the use of the correlation function $R_{ab}(t_1, t_2) = E\{a(t_1) \cdot b(t_2)\}$.

The expected value of the measured spectrum is now given by

$$E\{\hat{\Phi}_{ab}(j\omega)\} = \frac{1}{T}\int_0^T\int_0^T R_{ab}(t_1, t_2)\,e^{-j\omega(t_1 - t_2)}\,dt_1\,dt_2 \quad (2)$$

2 Measured frequency response function

We consider estimates of the system frequency response function, based on one or more blocks, each of length T, using the estimators given for one block by

$$\hat{G} = \frac{Y(j\omega)}{X(j\omega)} = \frac{X^*Y}{X^*X} = \frac{\hat{\Phi}_{yx}}{\hat{\Phi}_{xx}} \quad (3)$$

and for L blocks by

$$\bar{G} = \frac{\frac{1}{L}\sum X^*Y}{\frac{1}{L}\sum X^*X} = \frac{\bar{\Phi}_{yx}}{\bar{\Phi}_{xx}} \quad (4)$$

We first consider the mean value of the estimate. For the case where the number of blocks is very large, the expected value of the ratio defined in eqn. 4 is equal to the ratio of the expected values. Analysis in a later Section shows that estimates based on a single block have the same expected value. This results since the component of $y(t)$ which is correlated with $x(t)$ in each block is due to the measured excitation $x(t)$ for this block, and is proportional to this excitation.

The expected value of the measured autospectrum of the input signal of constant power spectral density Φ is given, from eqn. 1, by

$$E\{\hat{\Phi}_{xx}\} = \int_{-T}^{T}\left(1 - \frac{|\tau|}{T}\right)\Phi \cdot \delta(\tau) \cdot e^{-j\omega\tau}\,d\tau = \Phi$$

Similarly, from eqn. 2, with

$$R_{xy}(t_1, t_2) = \Phi g(t_1, t_2)$$

then

$$E\{\hat{\Phi}_{yx}\} = \Phi\int_0^T\left(1 - \frac{\tau}{T}\right)g(\tau)\,e^{-j\omega\tau}\,d\tau$$

Paper 3500D (C8), first received 3rd February and in revised form 13th August 1984
Prof. Douce is with the Department of Engineering, University of Warwick, Coventry CV4 7AL, England, and Mr. Balmer is with Coventry (Lanchester) Polytechnic, Priory Street, Coventry CV1 5FB, England

Thus, the expected values of the gain estimate is

$$E\{\bar{G}(j\omega)\} = \int_0^T \left(1 - \frac{\tau}{T}\right) g(\tau)\, e^{-j\omega\tau}\, d\tau \tag{5}$$

Two features in this expression lead to the bias in the estimate: the upper limit of integration is finite, and the second term within the brackets introduces a nonzero contribution.

When the time of observation is sufficiently large, relative to the settling time of the system, the upper limit can be replaced by infinity, and an approximate expression for the error in the gain estimate is

$$-\frac{1}{T}\int_0^\infty \tau g(\tau)\, e^{-j\omega\tau}\, d\tau$$

Using

$$\int_0^\infty t \cdot f(t)\, e^{-st}\, dt = -\frac{d}{ds} F(s)$$

this gives

$$-\frac{j}{T}\frac{dG(j\omega)}{d\omega}$$

This approximate expression has a very simple interpretation. Consider $G(j\omega)$ plotted on the complex plane for all $0 < \omega < \infty$. $dG(j\omega)/d\omega$ is directed along this locus at all frequencies. Hence the bias is perpendicular to the locus, with a clockwise rotation and is of magnitude $(1/T)|dG(j\omega)/d\omega|$.

The coherency function defined by

$$\gamma^2(j\omega) = \frac{E\{\hat{\Phi}_{vx}(j\omega)\} \cdot E\{\hat{\Phi}^*_{vx}(j\omega)\}}{E\{\hat{\Phi}_{xx}(j\omega)\} \cdot E\{\hat{\Phi}_{yy}(j\omega)\}}$$

can readily be evaluated from the above analysis.

The component of $y(t)$ for $t > 0$ which is independent of $x(t)$ for $t > 0$ is due to the system response to nonzero initial conditions at $t = 0$.

Consider the effect of a transient component in $y(t)$ due to the nonzero initial state of the system at $t = 0$. At $t = 0$, the steady state is assumed established, owing to the stationary input $x(t)$, and we set $x(t) = 0$ for $t > 0$.

From Reference 4 (p. 311), the autocorrelation function of the output is related to the crosscorrelation function by

$$R_{yy}(t_1, t_2) = g(t_1) * R_{xy}(t_1, t_2)$$

The crosscorrelation function is given by

$$R_{xy}(t_1, t_2) = \Phi g(t_2 - t_1) \quad \text{for } t_1 < 0$$
$$= 0 \quad \text{for } t_1 > 0$$

The result may be written, for $t_1, t_2 > 0$, as

$$R_{yy}(t_1, t_2) = \Phi \int_0^\infty g(t_2 + u) g(t_1 + u)\, du \tag{6}$$

Hence the expected value of the measured spectrum of this component can be found, using eqn. 2.

In a similar manner, the build-up with time of the component of $y(t)$ due to $x(t)$ for $t > 0$ can be determined. However, as $R_{yy}(t_1, t_2)$ for the total signal $y(t)$ is known to be stationary, it can be deduced immediately that the transient component is the negative of that given by eqn. 6.

The unmeasured portion of $y(t)$ due to $x(t)$ (for $0 < t < T$) is that which occurs for $t > T$. The autocorrelation function of this is obtained from eqn. 6, with the times t_1 and t_2 on the left-hand side replaced by $(t_1 - T)$ and $(t_2 + T)$, respectively.

It is important to note that, although expected values of the two transient components of $R_{yy}(t_1, t_2)$ have identical expressions, the two components are independent for sufficiently large T.

3 Alternative derivation

Using convolution, the above results can be derived from first principles. This permits some more detailed appreciation of the results of one individual experiment, as well as presenting a physical appreciation of the underlying mechanisms involved.

(*a*) *Bias:* Consider the input signal component

$$x(t) = A \sin(\omega t + \theta) \quad 0 < t < T$$

where, as before, ω is a harmonic of the fundamental frequency $2\pi/T$. If the measurement time is sufficiently long, then this component is uncorrelated with any other frequency component of the white input signal over the given time interval. It is also uncorrelated with $x(t)$ for $t < 0$. It follows that the component is correlated only with that component of the output signal $y(t)$ which is due to this input. It will be shown how bias in frequency response estimates arises owing to the transient component of the response due to this applied signal, and how the response depends, in an individual experiment, on the phase angle θ.

In terms of the system impulse response $g(t)$, the output is given by

$$y(t) = A \int_0^t g(\tau) \sin[\omega(t - \tau) + \theta]\, d\tau$$

The in-phase and quadrature components of the response, over the time interval $0 < t < T$, using $\sin(\omega t + \theta)$ and $\cos(\omega t + \theta)$ as references, are

$$\frac{2}{T}\int_0^T y(t) \begin{Bmatrix} \sin \\ \cos \end{Bmatrix} (\omega t + \theta)\, dt$$

Substituting for $y(t)$ and interchanging the order of integration, the in-phase and quadrature components of the output signal are given by

$$R = \frac{A}{T}\int_0^T g(\tau) \left\{ \int_\tau^T \cos \omega\tau - \cos[2\omega t - \omega\tau + 2\theta]\, dt \right\} d\tau$$

and

$$I = \frac{A}{T}\int_0^T g(\tau) \left\{ \int_\tau^T - \sin \omega\tau + \sin[2\omega t - \omega\tau + 2\theta]\, dt \right\} d\tau$$

respectively, giving

$$R = \frac{A}{T}\int_0^T (T - \tau) g(\tau) \cos \omega\tau\, d\tau + \frac{A}{\omega T} \cos 2\theta \times \int_0^T g(\tau) \sin \omega\tau\, d\tau \tag{7a}$$

and

$$I = \frac{A}{T}\int_0^T (T - \tau) g(\tau) \sin \omega\tau\, d\tau - \frac{A}{\omega T} \sin 2\theta \times \int_0^T g(\tau) \sin \omega\tau\, d\tau \tag{7b}$$

respectively.

Proceeding as before, we now assume that:

(1) The period of observation is significantly longer than the setting time of the system, so that the upper limits of integration can be taken as infinite

(2) The integral $\int_0^\infty \tau g(\tau)\, e^{-j\omega\tau}\, d\tau$ converges: hence

$$\int_0^\infty \tau g(\tau) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} (\omega t)\, d\tau = \begin{matrix} -\mathrm{Re} \\ +\mathrm{Im} \end{matrix} \left[\frac{dG(s)}{ds}\right]_{s=j\omega}$$

Dividing eqns. 7*a* and *b* by *A*, the real and imaginary parts of the estimated gain become

$$R[\hat{G}(j\omega)] = R[G(j\omega)] + \frac{1}{T} R[G'(j\omega)] - \frac{1}{\omega T} \times \cos 2\theta\, I[G(j\omega)]$$

and

$$I[\hat{G}(j\omega)] = I[G(j\omega)] + \frac{1}{T} I[G'(j\omega)] + \frac{1}{\omega T} \times \sin 2\theta\, I[G(j\omega)]$$

respectively, where $G'(j\omega)$ is written for $[dG(s)/ds]_{s=j\omega}$. Combining these expressions results in

$$\hat{G}(j\omega) = G(j\omega) + \frac{1}{T} G'(j\omega) - \frac{1}{\omega T} e^{-2j\theta} I[G(j\omega)] \qquad (8)$$

The three terms on the right-hand side correspond to the true value, the bias and a term of constant magnitude and phase angle depending on the phase of the input signal. If θ is uniformly distributed, then the final term has zero mean.

(*b*) *Complete response:* the input is modelled by a series of independent impulses of strength h. h is zero mean and of mean-square value related to the power spectral density of the input by $E\{h^2\} = \Phi$.

The input may, as before, be considered as two components, the first existing over the time interval $-\infty < t < 0$ and the second over the time of observation $0 < t < T$.

Consider an impulse applied at time τ, with $0 < \tau < T$. This gives a contribution to the measured amplitude spectrum of $y(t)$ equal to

$$y(\tau, j\omega) = h \int_\tau^T g(t - \tau)\, e^{-j\omega t}\, dt$$

and writing $u = t - \tau$, we obtain

$$y(\tau, j\omega) = h\, e^{-j\omega\tau} \int_0^{T-\tau} g(u)\, e^{-j\omega u}\, du \qquad (9)$$

This contributes to the measured cross-spectrum between $x(\tau) = h\delta(t - \tau)a$ term

$$\hat{\Phi}_{yx}(\tau, j\omega) = \frac{h^2}{T} \int_0^{T-\tau} g(u)\, e^{-j\omega u}\, du$$

This is now averaged with respect to τ, and taking expectation with respect to h^2, we find that

$$E\{\hat{\Phi}_{yx}(j\omega)\} = \Phi \cdot \frac{1}{T} \int_0^T \int_0^{T-\tau} g(u)\, e^{-j\omega u}\, du\, d\tau$$

$$= \Phi \frac{1}{T} \int_0^T \int_0^{T-u} g(u)\, e^{-j\omega u}\, d\tau\, du$$

$$= \Phi \int_0^T \left(1 - \frac{u}{T}\right) g(u)\, e^{-j\omega u}\, du$$

Similarly, the measured autospectrum of $y(t)$ due to $x(t)$ for $0 < t < T$ is, for the impulse at $t = \tau$ from eqn. 9, given by

$$\frac{1}{T} Y(\tau, j\omega) \cdot Y^*(\tau, j\omega) = \frac{h^2}{T} \int_0^{T-\tau} g(u)\, e^{-j\omega u}\, du \cdot \int_0^{T-\tau} g(v)\, e^{j\omega v}\, dv$$

Again, this is averaged with respect to τ over the range $0 < \tau < T$, and h^2 is replaced by its expected value.

4 Spectral windows

So far, the raw Fourier representation of $x(t)$ $(0 < t < T)$ has been considered. To improve the form of the equivalent filter in the frequency domain, particularly to reduce sidelobes, the time functions are multiplied by a non-rectangular window $w(t)$ $(0 < t < T)$, zero for t outside this range.

The effect of windowing the time functions on spectral estimates is well documented [2, 6, 7].

The conventional approach assumes a long record, so that transient effects may be ignored.

The analysis of Sections 2 and 3 can readily be extended to handle an arbitrary time window, as follows.

With a window $w(t)$, equal to zero except over the range $0 < t < T$, the expected value of the correlation function is modified to

$$R_w(t_1, t_2) = R(t_1, t_2) \cdot w(t_1) \cdot w(t_2)$$

From eqn. 2, the expected value of the modified measured spectral density becomes

$$E\{\hat{\phi}_{ab}(j\omega)\} = \frac{1}{T} \int_0^T \int R_{ab}(t_1, t_2) w(t_1) w(t_2) \times e^{-j\omega(t_1 - t_2)}\, dt_1\, dt_2$$

If the process is stationary, with $\tau = t_1 - t_2$, this gives

$$\frac{1}{T} \int_{-T}^{T} R_{ab}(\tau)\, e^{-j\omega\tau} \left\{ \int_{-\infty}^{\infty} w(t_2) \cdot w(t_2 - \tau)\, dt_2 \right\} d\tau \qquad (10a)$$

For the rectangular or 'do-nothing' window, eqn. 1 follows by setting $w(t) = 1$ for $0 < t < T$.

For the Hanning window

$$w(t) = (1 - \cos 2\pi t/T)$$

the inner integral in eqn. 10*a* becomes

$$T\left\{\left(1 - \frac{\tau}{T}\right)\left(1 + \frac{1}{2}\cos\omega_0\tau\right) + \frac{3}{4\pi}\sin\omega_0\tau\right\} \qquad (10b)$$

where $\omega_0 = 2\pi/T$. This is used later to evaluate the coherency with this window in an example.

Introduction of this window considerably reduces bias in many cases. This is due to the constant value of expr. 10*b* for small values of τ, which reduces the error in the expected value of the measured correlation functions. The variance of spectral estimates is also reduced, as the end portions of the records, which contain the transient terms, are given a small weight.

5 Example

Consider the first-order system

$$G(s) = \frac{Y(s)}{X(s)} = \frac{1}{1 + sT_1}$$

With white noise of power per unit bandwidth $\Phi_{xx}(j\omega) = \Phi$ applied to the system, then

$$R_{yx}(\tau) = \frac{\Phi}{T_s} e^{-\tau/T_s} \quad \text{for } \tau > 0$$

$$R_{yy}(\tau) = \frac{\Phi}{2T_s} e^{|\tau|/T_s}$$

Thus, with measurements based on a block of length T, from eqn. 1, we find

$$E[\hat{\Phi}_{yx}(j\omega)] = T^2 \int_0^T \left(1 - \frac{\tau}{T}\right) R_{yx}(\tau) e^{-j\omega\tau} d\tau$$

$$= \Phi_{xx} \left\{ \frac{1}{(1 + j\omega T_s)} - \frac{T_s}{T} \frac{(1 - e^{-T/T_s})}{(1 + j\omega T_s)^2} \right\}$$

Dividing by Φ_{xx} gives the expected value of the estimate of the gain at the harmonic frequency ω.

The second term in the resulting expression is the bias. The approximate expression for the bias is

$$\frac{-j}{T} \frac{dG(j\omega)}{d\omega} = -\frac{T_s}{T} \frac{1}{(1 + j\omega T_s)^2}$$

Thus the approximate expression will be valid under normal experimental conditions, as the difference is a factor of $(1 - e^{-T/T_s})$, and it is normal for $T \gg T_s$.

In Fig. 1 is shown the frequency-response locus and the

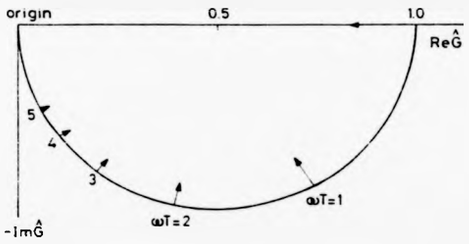


Fig. 1 *Frequency response and bias for $T = 10\ T_s$*

bias for $T = 10T_s$. In this particular case, the frequency-response locus is a semicircle, and the bias is always directed towards the centre.

In Section 3, the response to the sinusoid $x(t) = A \sin(\omega t + \theta)$ applied at $t = 0$ was considered, and eqn. 8 gives the approximate expression for the estimated gain. In addition to the bias, there is a further error, equal to

$$-\frac{1}{\omega T} I[G(j\omega)] \cdot e^{-2j\theta}$$

This term describes a circle on the complex plane of radius $(T_s/T)(1/(1 + \omega^2 T_s^2))$ equal to the (approximate) magnitude of the bias. Thus, for particular values of θ, the estimate is exact. These values correspond to the continuously applied signal, producing zero output at time $t = 0$, so there is no transient term in the observed response.

The expected value of the measured autospectrum of the output is readily shown to be

$$E[\hat{\Phi}_{yy}(j\omega)] = \Phi_{xx} \left\{ \frac{1}{1 + \omega^2 T_s^2} - \frac{T_s}{T} \frac{(1 - \omega^2 T_s^2)(1 - e^{-T/T_s})}{(1 + \omega^2 T_s^2)^2} \right\} \quad (11)$$

An approximate expression for the bias in this quantity, assuming $e^{-T/T_s} \simeq 0$, can be derived as for the cross-spectrum, with note of the real nature of the autospectrum.

This leads to the (exact) expression for the coherency function

$$\gamma^2(j\omega) = \frac{(1 + \omega^2 T_s^2) + T_s/T(1 - e^{-T/T_s})\left[\frac{T_s}{T}(1 - e^{-T/T_s}) - 2\right]}{(1 + \omega^2 T_s^2) - \frac{T_s}{T}(1 - e^{-T/T_s})(1 - \omega^2 T_s^2)}$$

When $e^{-T/T_s} \simeq 0$, setting $\omega = 2\pi n/T$, this becomes

$$\gamma^2(j\omega) = \frac{1 - 2T_s/T + [(2\pi n)^2 + 1](T_s/T)^2}{1 - T_s/T + (2\pi n T_s/T)^2(1 + T_s/T)}$$

For $T = 10T_s$, this is approximately equal to 0.9 for all integer n.

Returning to the exact expression, it is notable that, for $T \ll T_s$, we have

$$\gamma^2 \to T/2T_s \quad \text{for } \omega = 0$$

$$\gamma^2 \to 0.5 \quad \text{for } \omega = 2\pi n/T,\ n \neq 0$$

A computer simulation has verified this result.

We may examine the autospectrum of y, given by eqn. 11, in more detail. This expression is the result of transforming the steady-state autocorrelation function over the finite time of observation. As previously discussed, the autocorrelation function of $y(t)$, $R_{yy}(t_1, t_2)$, for $t_1, t_2 > 0$, can be considered as the sum of two time-varying functions, one representing the decaying response to initial conditions established by $x(t)$, $t < 0$, which is uncorrelated with $x(t)$ for $t > 0$, and the second term representing the build-up of the response to $x(t)$ for $t > 0$ towards the steady-state conditions.

Owing to initial conditions at $t = 0$, the autocorrelation function is given by eqn. 6 as

$$R_{yy}(t_1, t_2) = \Phi \int_0^\infty g(t_1 + u) g(t_2 + u)\, du$$

with

$$g(t) = \frac{1}{T_s} e^{-t/T_s} \quad \text{for } t > 0$$

and

$$R_{yy}(t_1, t_2) = \frac{\Phi}{2T_s} e^{-(t_1 + t_2)/T_s}$$

From eqn. 1

$$E\{\hat{\Phi}_{yy}(j\omega)\} = \frac{\Phi_{xx}(T_s/T)}{2(1 + \omega^2 T^2)} [1 - e^{-T/T_s}]^2$$

Owing to $x(t)$ for $t > 0$, $R_{yy}(t_1, t_2)$ obeys

$$R_{yy}(t_1, t_2) = g(t_1) * R_{xy}(t_1, t_2)$$

In this situation, $R_{xy}(t_1, t_2) = 0$ for t_1 or $t_2 < 0$, we find

$$R_{xy}(t_1, t_2) = \Phi g(t_2 - t_1) \quad t_1 \text{ and } t_2 > 0$$

giving

$$R_{yy}(t_1, t_2) = \frac{\Phi}{2T_s} \{e^{-|t_2 - t_1|/T_s} - e^{-(t_1 + t_2)/T_s}\}$$

The two terms on the right-hand side can be associated with the steady-state and transient effects, respectively.

For this system, the transients have a particularly simple effect on the gain estimates, and possess a structure not revealed by conventional analysis. It is necessary to use the principles considered in Section 3, letting the component of interest in $x(t)$, $0 < t < T$, be written as

$$x(n, t) = A_n \sin(n\omega_0 t + \theta_n)$$

For simplicity, consider T/T_s large, so $e^{-T/T_s} \simeq 0$.

The total transient in this system is of the form $(h/T_s)\, e^{-t/T_s}$, for $t > 0$, where h is a zero-mean random variable. Measuring the Fourier component of this transient at frequency $n\omega_0$ gives a resultant $h \cdot G(jn\omega_0)$.

With reference to the input component differing from $\cos n\omega_0$ by an angle $(-\pi/2) + \theta_n$, the resultant is $h \cdot G(jn\omega_0) \cdot e^{j((\pi/2) - \theta_n)}$.

Letting $G(jn\omega_0) = |G| e^{-j\phi_n}$, it follows that the transients add to the estimate a term which always lies along the line at an angle $((\pi/2) - \theta_n - \phi_n)$.

An ensemble of experiments can be considered in which the phase angle is held constant. This is readily simulated digitally. In this situation, the total contribution to the measured frequency response at one frequency is shown in Fig. 2, where subscripts are omitted.

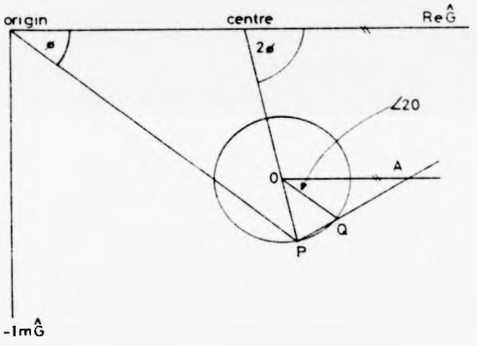


Fig. 2 *Resultant measured response at one frequency*

ϕ = phase lag of system
P = true frequency response
PO = bias
OQ = component which is function of angle 2θ
QA = effect of uncorrelated transients
$\angle OAQ = (\pi/2 - \theta - \phi)$

The construction of the measured response proceeds as follows: from the true value P, a line PO, representing the bias, is drawn towards the centre of the semicircle representing $G(j\omega)$. From O, a line OQ is drawn to represent the final term in eqn. 8. It is shown above that the length OQ = length OP. From Q, a line QA at an angle $((\pi/2) - \theta - \phi)$ to the horizontal is drawn to represent the contribution of the transient. Elementary trigonometry shows that this line passes through the true value P.

The effect of the Hanning window on the coherency function has been evaluated for this example using eqn. 10. Fig. 3 shows the results for a measurement time $T = 10\ T_s$, at the harmonic frequencies $f = 0, 1/T$ etc. to $20/T$.

6 Conclusions

The statistics of spectral estimates based on short time samples of input-output records can be obtained using correlation techniques or equivalently from the impulse response of the system. The second method gives some additional insight into the detailed nature of the errors.

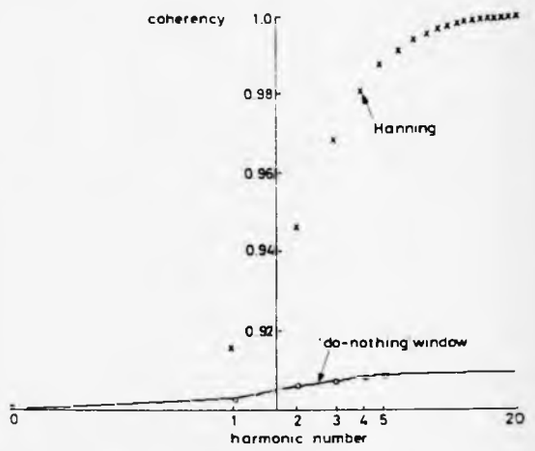


Fig. 3 *Coherency as a function of frequency*

The coherency between input and output can be evaluated as a function of system response and period of observation.

No mention has been made in this paper of methods that can be used to reduce the errors in the estimates. Specifically, many techniques have been developed and successfully applied invoking parameteric models of spectra and systems. The model can take the form of an autoregressive, moving-average or autoregressive moving-average model [1].

The authors are proceeding with further work on applying parameterisation methods to frequency-response estimates. The novelty of the approach is the inclusion of extra parameters to model the components of the measured spectra associated with the significant transient effects.

Although attention has been confined throughout this paper to white excitation, much of the analysis, including the derivation of the coherency function, is applicable for a coloured input signal. A major difference is noted when the transient is considered in detail, as the response to the unmeasured input present before $t = 0$ is correlated with the measured input, and hence with the response of the system to this input.

7 References

1 Special issue: 'Spectral estimation', *Proc. IEEE*, September 1982, **70**
2 GECKINLI, N.C., and YAVUZ, D.: 'Discrete Fourier transformation and its applications to power spectra estimation' (Elsevier, 1983)
3 CHILDERS, D.G. (Ed.): 'Modern spectral analysis' (IEEE Press, New York, 1978)
4 PAPOULIS, A.: 'Probability, random variables and stochastic processes' (McGraw Hill, 1965)
5 GOODMAN, N.R.: 'On the joint estimation of the spectra, cospectrum and quadrature spectrum of a two-dimensional stationary Gaussian process', Ph.D. thesis, Princeton University, 1957
6 NUTTALL, A.H., and CARTER, G.C.: 'Spectral estimation using combined time and lag weighting', *Proc. IEEE*, 1982, **70**, pp. 1115–1125
7 BLACKMAN, R.B., and TUCKEY, J.W.: 'The measurement of power spectra' (Dover Publications Inc., New York, 1958)
8 AKAIKE, H.: 'On the statistical estimation of the frequency response function of a system having multiple input', *Ann. Inst. Stat. Maths.*, 1965, **17**, pp. 155–210

ADAPTIVE CONTROL IN THE FREQUENCY DOMAIN

*Balmer, Leslie
**Douce, John L.

* Coventry (Lanchester) Polytechnic, Coventry, England.
** University of Warwick, Coventry, England.

A method of adaptive control is described where the error power is minimised, this being obtained using frequency domain techniques. Results are presented for a simulated system and performance is compared when parametric and non-parametric models are used for the system identification.

LIST OF SYMBOLS

$G(j\omega)$	System frequency response function
$D(j\omega)$	Controller frequency response function
K	Controller gain
K_I	Integral gain
T_D	Derivative time constant
Φ_{rr}	Input spectral density
Φ_{ru}, Φ_{rc}	Cross spectral densities, input to control signal and input to output
b_o, a_1, a_2, a_3	Parameters in frequency response function

$$G(j\omega) = \frac{b_o}{1 + a_1 j\omega + a_2 (j\omega)^2 + a_3 (j\omega)^3}$$

INTRODUCTION

Most of the recent work relating to adaptive control has concentrated on time domain methods. However, frequency response models are still extensively used to describe systems and many practical performance criteria are based on these models. This paper examines the use of such a model in an adaptive control scheme. Possible advantages of such a method are that if a non-parametric model is used very few prior assumptions need be made with respect to the system and performance criteria related to frequency domain models can easily be incorporated into the performance criterion.

THE GENERAL SYSTEM

The method of adaption can be understood by reference to Fig. 1. The system has an open-loop frequency response function $G(j\omega)$ and the parameters are to be optimised in the controller $D(j\omega)$, here shown as a three term controller. An estimate of the frequency response function $\hat{G}(j\omega)$ is obtained by taking data blocks $r(t)$, $u(t)$, $c(t)$. (Details of the method of estimation will be considered in the next section). From $G(j\omega)$ the frequency response function $E(j\omega)/R(j\omega)$ can be obtained for a given controller $D(j\omega)$ as

$$\frac{E}{R}(j\omega) = \frac{1}{1 + D(j\omega)\, G(j\omega)} \qquad (1)$$

Using an estimate of input power spectral density Φ_{rr} the total error power, V, can be expressed as

$$V = \int_{-\infty}^{+\infty} \left| \frac{E}{R}(j\omega) \right|^2 \Phi_{rr}\, d\omega \qquad (2)$$

This error power is a function of controller parameters and an "off-line" minimisation routine can be used to obtain the parameter values that minimise V. The values are then implemented in the controller in the system. As more data becomes available this is combined with previous data blocks in the estimation routine to give a more refined estimate $\hat{G}(j\omega)$ and hence improved controller settings.

THE ESTIMATION ROUTINE

Various methods have been investigated for obtaining the frequency response estimate $\hat{G}(j\omega)$, these can be broadly divided into non-parametric and parametric methods.

Non-parametric methods have the advantage that no specific model need be presupposed in formulating the estimation problem. The estimate used for $\hat{G}(j\omega)$ is given as:

$$\hat{G}(j\omega) = \frac{\Sigma\Phi_{rc}}{\Sigma\Phi_{ru}} \qquad (3)$$

where the summation is performed over successive data blocks. The cross-spectra are obtained by using the Fast Fourier Transform (FFT) on the data blocks r(t), u(t), c(t) after modifying these by a suitable window. Recent papers by the authors [1], [2] have investigated the statistics of the estimate given by (3).

It has been shown that for short data blocks it is subject to bias and variance even in the noise free case. The error transfer function (eqn. 1) is very sensitive to variations in $\hat{G}(j\omega)$ at frequencies where $\hat{G}(j\omega)$ is close to -1 + jo. Hence the variance in $\hat{G}(j\omega)$ can cause large errors in the parameter settings. These errors will be reduced as more blocks of data are taken into account as the bias in $\hat{G}(j\omega)$ is not usually a problem.

In parametric methods a model is assumed in the formulation of the problem and the requirement is to estimate the parameters of the model from the short term data blocks. This method has been investigated extensively for the case of auto and cross-spectral estimates [3]. The authors have investigated a method [4] by which a model is assumed for the frequency response function. However the novel feature of this approach is that parameters are included to account for the effect of the transients due to the finite data block length. With such an approach it is possible to obtain exact estimates in the noise free case. When noise is present in the system the estimates become subject to variance. This can be reduced by using data in successive blocks but care must be taken in utilising this data as the parameters due to the transients change from block to block.

MINIMISATION AND STABILITY

It is required to derive the values of the controller parameters that minimise the error power as obtained from equations (1) and (2). The problems associated with this minimisation can be best understood by reference to a specific system (this system is considered in more detail in the following section). Consider $G(j\omega) = 1/(1 + j\omega)^3$, the input band limited noise and the controller a simple gain controller gain K. The variation of normalised error power against K is shown in Fig. 2. The required minimum occurs at K = 2.7 and instability occurs for values of K greater than 8. However the error power as calculated from equations (1) and (2) falls for K greater than 8.

The minimum was obtained by use of the "Hooke and Jeeves" direct search method. With this method there is a danger that should the search point move to a value of K greater than 8 then a false minimum may be obtained and if the corresponding K value is implemented into the controller instability of the system will result. As this controller setting is held throughout a complete data block this would prove disastrous in a practical situation. A test must be performed at each search point to check for potential instability and the result used as a constraint to restrict the search region. Constraints must also be applied to prevent negative values of the controller parameters being implemented.

The method of calculating the error power and checking for stability depend on whether $\hat{G}(j\omega)$ is available in parametric or non-parametric form. In parametric form an Algorithm due to Astrom [5] can be used to obtain error power and to check for stability. For the non-parametric form the error power is approximated by replacing the integral of equation 2 by a summation over the frequency points. In order to check for stability a numerical form of the Nyquist Criterion was developed.

RESULTS FOR A SPECIFIC SYSTEM

The system shown in Fig. 3 was imulated using a digital simulation with a sampling time of 0.1 second. For such a system the optimum parameter settings are $K = 3.4$, $K_I = 0.4$, $T_D = 1.0$. The minimum is relatively insensitive to small variations of K_I. T_D does not produce a minimum but gives very small decrease in error power as it is increased to infinity. A constraint was imposed to limit T_D to unity. At optimum setting the normalised error power is approximately 0.1. Initially the controller parameters were set to $K = 0.2$, $K_I = 0.0$, $T_D = 0.0$ (normalised error power approximately 0.8) and the system allowed to adapt from these settings.

Fig. 4 shows results using a non-parametric model with data blocks of length 102.4 seconds (1024 data points) and no noise present. As can be seen convergence to the optimum settings is very slow. If the block length is decreased, or if noise is added then the parameters will not converge to optimum.

If a parametric model is used performance, for the noise free system, is considerably improved. The parameters reach their optimum value in one data block of length 12.8 seconds (128 points). This is the shortest block that could be used because of the resolution of the frequency points. With 7 parameters to identify (4 for the system and 3 for the transients) 4 frequency points are required (each point other than zero frequency contributes a real and imaginary part). With 64 points the frequency spacing is 0.156 Hz and the points cause ill-conditioned equations in the regression analysis.

If noise is added the optimum settings cannot be achieved in one data block. The time required depends upon the noise level and the length of the block. Fig. 5 shows results when noise is added. The noise represented a noise/signal power at the output of 3:1 at the initial setting and 0.3:1 at the optimum setting. As can be seen, convergence time is reduced compared to the non-parametric case even though no noise was present in that case. Fig. 6 shows the parameter estimates on successive data blocks under these conditions.

CONCLUSIONS

A method of using frequency domain estimates for adaptive control has been investigated. Results have been presented for a specific system

using parametric and non-parametric modelling. Although the non-parametric model has the advantage that few prior assumptions need be made regarding the system the results obtained are inferior to those obtained using a parametric model.

REFERENCES

1. Douce, J.L. and Balmer, L. "Transient Effects in Spectrum Estimation". I.E.E. Proceedings, Vol. 132, Pt. D, No. 1, Jan 1985, p25-27.

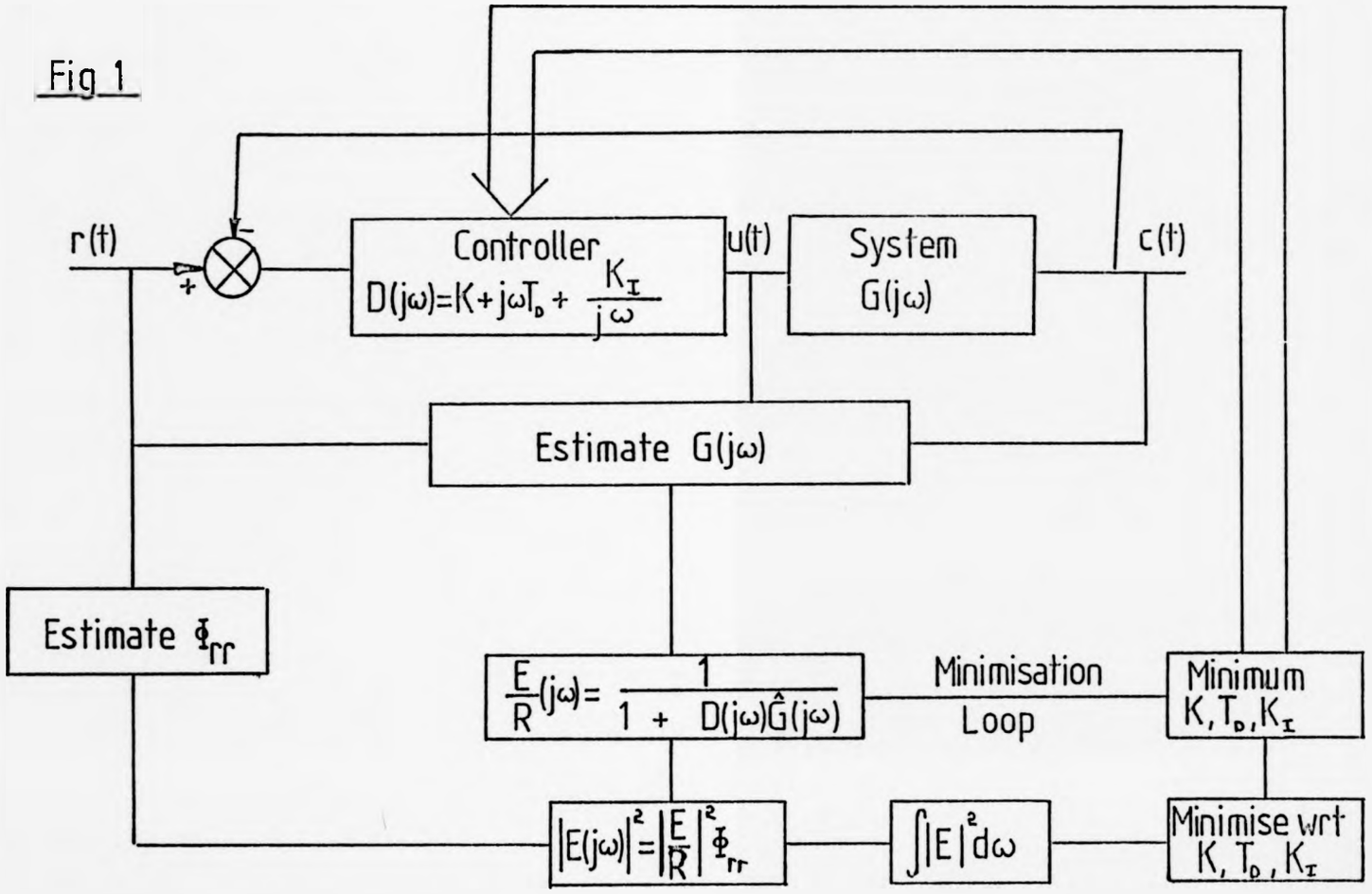
2. Douce, J.L. and Balmer, L. "Bias of Frequency Response Estimates". Systems Science, Vol. 8, No. 2-3, p95-102.

3. Special Issue on Spectral Estimation. Proc. IEEE, Sept. 1982, Vol. 70, No. 9.

4. Douce, J.L. and Balmer, L. "Frequency Response Estimation Using Short Data Blocks". 3rd International Conference on Systems Engineering. Wright State University. Sept. 1984.

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

Fig 1


r(t)
+
−
Controller
$D(j\omega)=K+j\omega T_D+\frac{K_I}{j\omega}$
u(t)
System
$G(j\omega)$
c(t)
Estimate $G(j\omega)$
Estimate Φ_{rr}
$\frac{E}{R}(j\omega)=\frac{1}{1+D(j\omega)\hat{G}(j\omega)}$
Minimisation
Loop
Minimum
K, T_D, K_I
$|E(j\omega)|^2=\left|\frac{E}{R}\right|^2\Phi_{rr}$
$\int|E|^2d\omega$
Minimise wrt
K, T_D, K_I

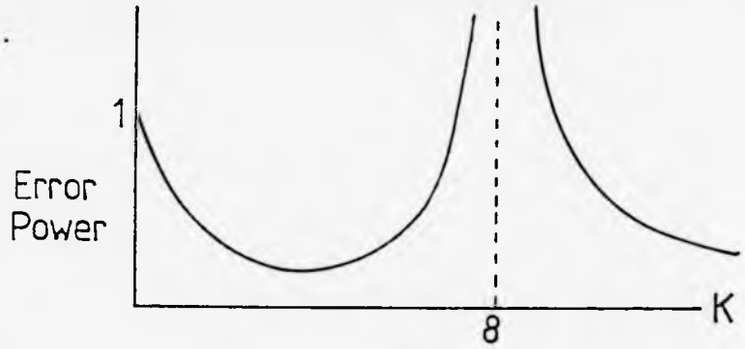


Fig 2

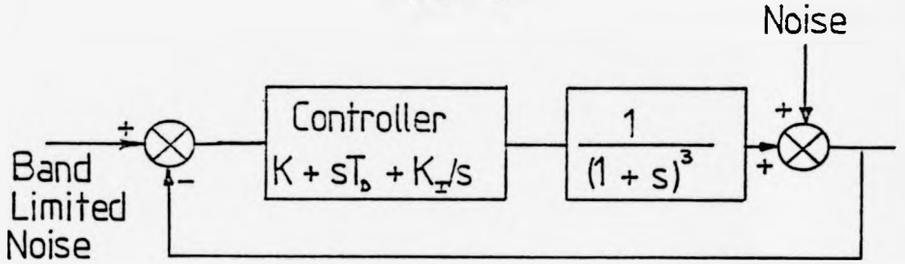


Fig 3

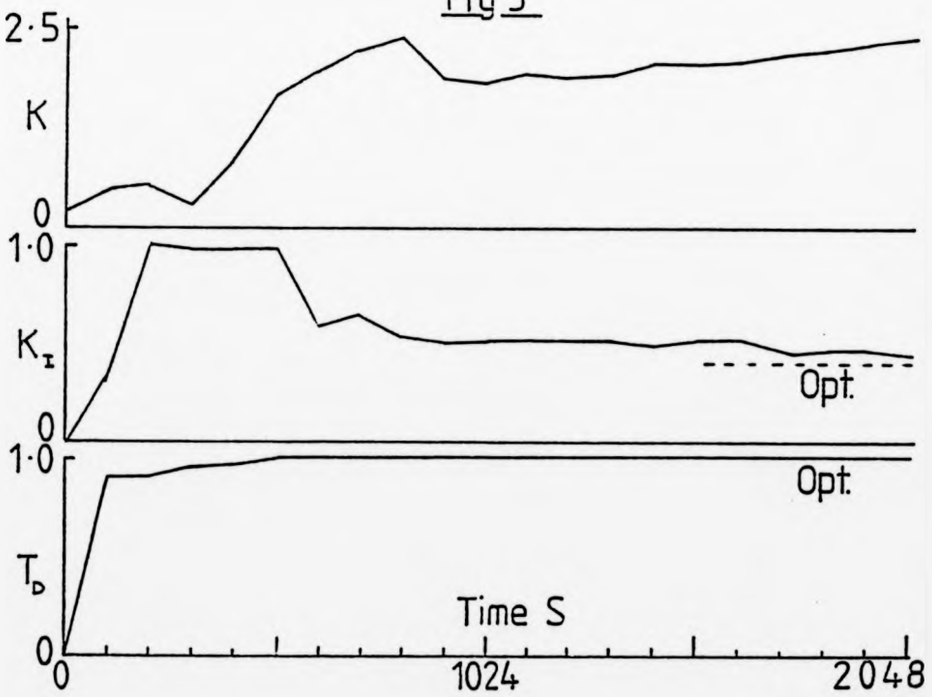


Fig 4

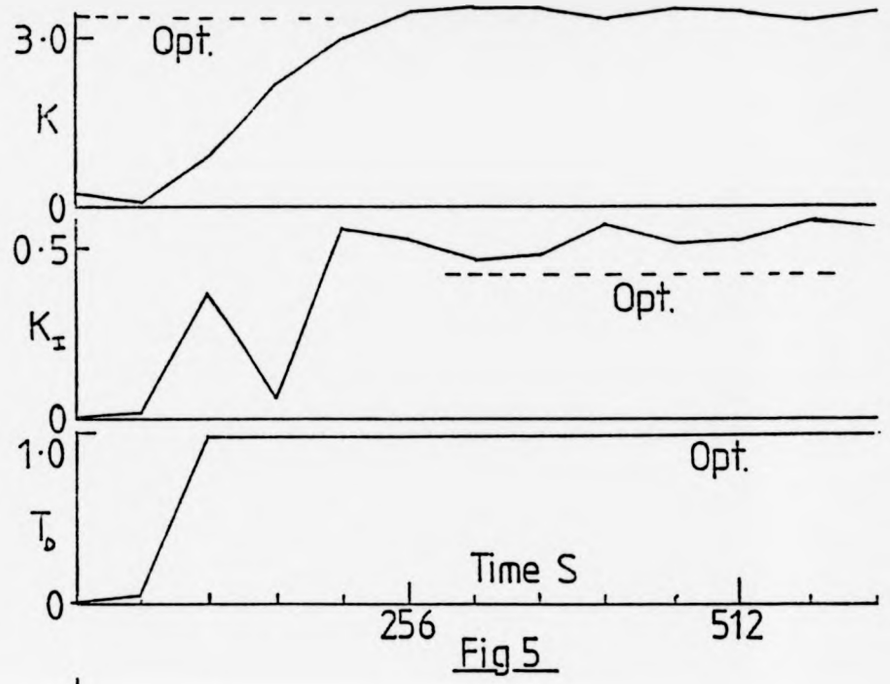

3·0
Opt.
K
0
0·5
K_I
Opt.
0
1·0
Opt.
T_D
Time S
0
256
512

Fig 5

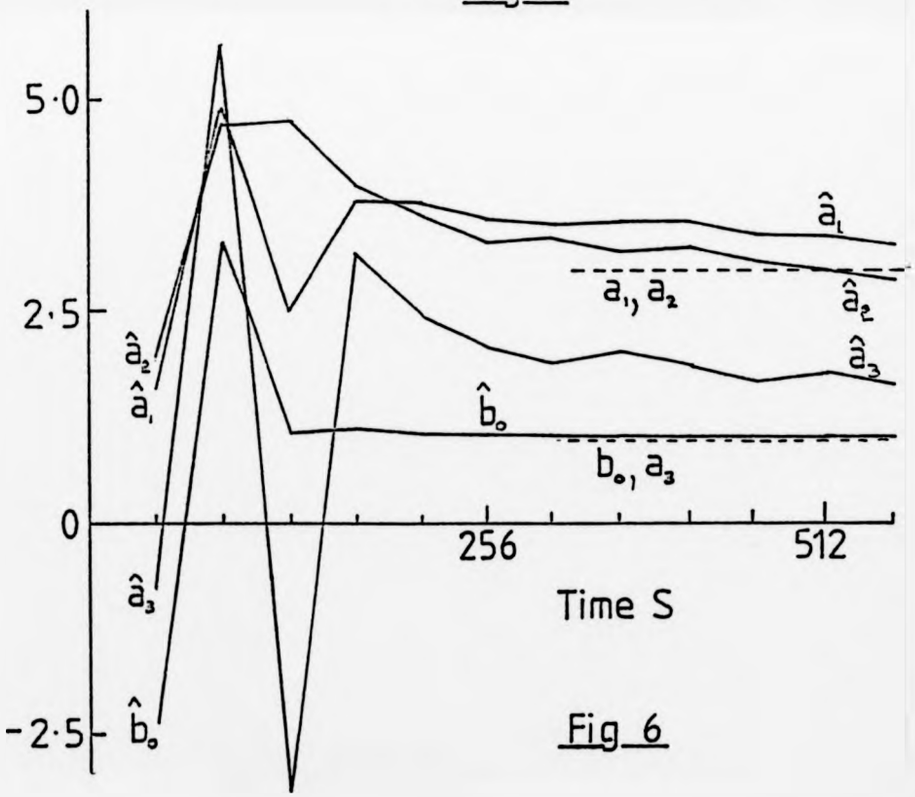

5·0
2·5
0
-2·5
$\hat{a}_1$
a_1, a_2
$\hat{a}_2$
$\hat{a}_3$
$\hat{b}_0$
b_0, a_3
$\hat{a}_2$
$\hat{a}_1$
$\hat{a}_3$
$\hat{b}_0$
256
512
Time S

Fig 6