NATIONAL LIBRARY
OTTAWA

CANADA

BIBLIOTHÈQUE NATIONALE
OTTAWA

NAME OF AUTHOR...K. A. MURALEEDHARAN........

TITLE OF THESIS...A...TIME...DOMAIN...APPROACH
TO....FREQUENCY..DOMAIN....
..APPROXIMATION..................

UNIVERSITY......UNIVERSITY...OF...ALBERTA

DEGREE FOR WHICH THESIS WAS PRESENTED......Ph. D..........

YEAR THIS DEGREE GRANTED.....1973......................

Permission is hereby granted to THE NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(Signed) K. A. Muraleedharan

PERMANENT ADDRESS:

..ASSISTANT...PROFESSOR
..ENGINEERING.. COLLEGE
...TRIVANDRUM-16

DATED 5th March...1973

NL-91 (10-68)

THE UNIVERSITY OF ALBERTA

A TIME DOMAIN APPROACH TO FREQUENCY DOMAIN APPROXIMATION

by

© K. A. MURALEEDHARAN

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENT
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL ENGINEERING

EDMONTON, ALBERTA

SPRING, 1973

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled A TIME DOMAIN APPROACH TO FREQUENCY DOMAIN APPROXIMATION submitted by K. A. MURALEEDHARAN in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Keith A. Stromsmoe
Supervisor

D. H. Kelly

Z. J. Koles

John F.

A. G. Wacker
External Examiner

Date 5 March 1973

ABSTRACT

The approximation of an ideal frequency response by a realizable filter has wide applications in Engineering. This topic is treated in this study. The general approach used is to convert the problem into the time domain and to find a filter satisfying the desired impulse response such that the error in the frequency domain is minimized. The error criteria used are the Least Integral Squared Error and the Maximum Deviation, in the frequency domain. The Integral Squared Error (ISE) in frequency domain is related to that in time domain by the relation

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}|F(j\omega) - H(j\omega)|^2 d\omega = \int_{-\infty}^{+\infty}|f(t) - h(t)|^2 dt$$

where F(jω) and H(jω) are the desired and approximate frequency responses and f(t) and h(t) are the corresponding impulse responses respectively. The minimization of ISE alone need not limit the maximum deviation in frequency domain. The deviation in frequency domain is related to the average error in time domain by the relation

$$|F(j\omega) - H(j\omega)| \leqslant \int_{-\infty}^{+\infty}|f(t) - h(t)|dt$$

This relation is used to control the frequency domain deviation. It has been shown that the minimization of deviation in the frequency domain can be achieved while keeping the ISE within allowable limits.

The approximation is carried out using a set of orthonormal functions of exponentials. The numerical evaluation of the time domain representations of these orthonormal functions are carried out by a novel method. This method makes the computer evaluation of the transient response of any rational function of complex frequency simpler and more efficient.

The theory developed in this dissertation is applied to the specific example of approximating the ideal low pass filter.

ACKNOWLEDGEMENTS

The author is greatly indebted to his supervisor Dr. K. A. Stromsmoe. His valuable guidance and timely advice during the course of this research are responsible for the successful completion of this dissertation. Special thanks are to Professor Kingma whose lectures inspired the author to work in this field. The author wishes to thank The Canadian Commonwealth Scholarship and Fellowship Administration for their financial assistance and The Government of India and Kerala State Government for deputing the author for higher studies in Canada.

The author thankfully acknowledges the kind cooperation of the Department of Electrical Engineering of The University of Alberta.

TABLE OF CONTENTS

Page

CHAPTER 1 INTRODUCTION 1

CHAPTER 2 LEAST INTEGRAL SQUARED ERROR FILTERS AND EVALUATION OF INTEGRAL SQUARED ERROR . . 4

INTRODUCTION 4

LEAST INTEGRAL SQUARED ERROR FILTERS 5

EXPONENTIAL APPROXIMATION 7

ORTHONORMAL FUNCTIONS OF EXPONENTIALS 10

COMPLIMENTARY FILTER AND ISE 16

CHAPTER 3 A SURVEY OF PREVIOUS RESEARCH 25

PRONY'S METHOD 25

KAUTZ'S METHOD 28

YENGST'S METHOD 30

METHOD OF PERDIKARIS AND LAGO 33

METHOD OF VASILU 37

METHODS OF McDONOUGH AND HUGGINS 38

SEAR'S METHOD 45

MSS METHOD 51

MILLER'S METHOD 56

APPROXIMATION TO IDEAL LOW PASS FILTER USING TIME DOMAIN APPROACH 58

Table of Contents cont'd.

Page

CHAPTER 4 IMPROVING THE FREQUENCY RESPONSE BY MINIMIZING THE AVERAGE ERROR IN TIME DOMAIN 63

PROBLEM STATEMENT 63

AN UPPERBOUND ON DEVIATION 64

MINIMIZATION OF AVERAGE ERROR 68

UNIQUENESS OF THE APPROXIMATION 76

THE GRADIENT OF AVERAGE ERROR 78

CHAPTER 5 EVALUATION OF TRANSIENT RESPONSE 83

INTRODUCTION 83

A BRIEF SURVEY OF PREVIOUS INVESTIGATIONS ON COMPUTATION OF TRANSIENT RESPONSE 83

NEW METHOD 91

CHAPTER 6 COMPUTER ALGORITHMS AND NUMERICAL EXAMPLES 107

INTRODUCTION 107

COMPUTATION OF ISE 107

MINIMIZATION OF ISE 109

FUNCTION MINIMIZATION WITHOUT CALCULATING DERIVATIVES 110

1) Powell's Method 110

Table of Contents cont'd.

Page

2) Zangwill Method 114

MINIMIZATION OF AVERAGE ERROR 122

NUMERICAL EXAMPLES 125

Example 1a 130

Example 1b 134

Example 2a 138

Example 2b 142

SUMMARY AND CONCLUSIONS OF NUMERICAL EXAMPLES . . . 146

CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK 148

RECOMMENDATIONS FOR FUTURE WORK 151

REFERENCES 154

APPENDIX I 161

APPENDIX II 184

LIST OF FIGURES

Figure Page

2.1 Schematic diagram for evaluating ISE 23

3.1 Evaluation of ISE 38

3.2 Evaluation of $\int_0^\infty f(t)\phi_{N+i}(t)dt$ by filtering operation 42

3.3 Simplified diagram to evaluate $\int_0^\infty f(t)\phi_{N+i}(t)dt$ by filtering operation 43

3.4 Schematic diagram to evaluate $\psi_{if}(t)$ and $g_{kf}(t)$ 49

3.5 Evaluation of orthonormal coefficients 50

6.1 Frequency and Impulse responses of ideal low-pass filter. Delay=5 secs. 126

6.2 Truncated Impulse responses of ideal low-pass filter. 128

6.3 Magnitude and Phase responses of $\tilde{H}(j\omega)$, N=5, Delay=π, Minimum ISE=0.00021 132

6.4 Magnitude and Phase responses of $H(j\omega)$, N=5, Delay=π, Minimum γ=0.06163, ISE$\leqslant$0.0005 132

6.5 The least square approximation $\tilde{h}(t)$ and error $\tilde{e}(t)$. N=5, Delay=π, Minimum ISE= 0.00021 133

6.6 The approximation h(t) and error e(t) for minimum average error. N=5, Delay=π,

List of Figures cont'd.

Figure Page

Minimum $\gamma=0.06163$, ISE$\leqslant 0.0005$ 133

6.7 Magnitude and Phase responses of $\tilde{H}(j\omega)$, N=8, Delay=π, Minimum ISE=0.000057 136

6.8 Magnitude and Phase responses of $H(j\omega)$, N=8, Delay=π, Minimum $\gamma=0.02058$, ISE$\leqslant$ 0.0005 136

6.9 The least square approximation $\tilde{h}(t)$ and error $\tilde{e}(t)$. N=8, Delay=π, Minimum ISE= 0.000057 137

6.10 The approximation h(t) and error e(t) for minimum average error. N=8, Delay=π, Minimum $\gamma=0.02058$, ISE$\leqslant 0.0005$ 137

6.11 Magnitude and Phase responses of $\tilde{H}(j\omega)$, N=5, Delay=2π, Minimum ISE=0.00077 140

6.12 Magnitude and Phase responses of $H(j\omega)$, N=5, Delay=2π, Minimum $\gamma=0.25125$, ISE$\leqslant 0.035$ 140

6.13 The least square approximation $\tilde{h}(t)$ and error $\tilde{e}(t)$. N=5, Delay=2π, Minimum ISE= 0.00077 141

6.14 The approximation h(t) and error e(t) for minimum average error. N=5, Delay=2π, Minimum $\gamma=0.25125$, ISE$\leqslant 0.035$ 141

List of Figures cont'd.

Figure Page

6.15 Magnitude and Phase responses of $\tilde{H}(j\omega)$, N=8, Delay=2π, Minimum ISE=0.00052 144

6.16 Magnitude and Phase responses of $H(j\omega)$, N=8, Delay=2π, Minimum γ=0.232, ISE⩽0.03 144

6.17 The least square approximation $\tilde{h}(t)$ and error $\tilde{e}(t)$. N=8, Delay=2π, Minimum ISE=0.00052 145

6.18 The approximation h(t) and error e(t) for minimum average error. N=8, Delay=2π, Minimum γ=0.232, ISE⩽0.03 145

CHAPTER 1

INTRODUCTION

In recent years there has been a renewed interest in the approach of approximating a desired frequency domain characteristic by converting the problem into one of time domain approximation (1-4). This has led to the possibility of realizing filters using least integral squared error criterion in the frequency domain. The family of filters having this frequency domain criterion have not received wide attention. This is partly due to certain problems arising in the solution of the optimum filter parameters. This research is intended to fill this gap. The general approach employed in this work is to convert the problem into a time domain approximation. The least square criterion does not take into account the maximum deviation of the achieved response from the desired frequency response. This work develops a new method of design which minimizes this deviation in the frequency domain while keeping the integral square error within allowable limits. Thus this new method makes use of a combination of least square and Chebyshev criterions in the frequency domain. One problem encountered in this approach is the need to compute on digital computer the transient response of a Laplace transform expressed as rational function. This has been solved by a novel method (5).

Having broadly outlined the scope of this work in this chapter, the basic problem of least square design in the frequency domain is presented in Chapter 2. The conversion from frequency domain to time domain and the equivalent time domain approximation by using a set of orthonormal functions of exponentials are discussed in this chapter. The concept of complimentary filter and the evaluation of the integral squared error using this filter are also reviewed in this chapter. Chapter 3 gives a summary of recent works concerned with this problem. Chapter 4 gives the mathematical basis of the new method of frequency domain design based on minimizing the deviation in the frequency domain while keeping the integral squared error within allowable limits. The basic theory of the new method of numerical evaluation of the transient response is given in Chapter 5. The convergence properties and comparative merits and demerits of this new method are also discussed in this chapter. The implementation of the new method of design on digital computer is done in Chapter 6. The computer algorithm that is used to determine the optimum poles of the filter by minimizing the integral squared error is reviewed in this chapter. The results obtained by applying this new technique to a specific example are also discussed here. Chapter 7 gives a summary of the entire dissertation and discusses the conclusions of this research work. The areas of further work are also mentioned in this chapter.

CHAPTER 2

LEAST INTEGRAL SQUARED ERROR FILTERS AND EVALUATION OF INTEGRAL SQUARED ERROR

INTRODUCTION

In this chapter we define "Least Integral Squared Error Filters" and outline the technique of realizing such filters. The theory of evaluating the "Integral Squared Error" (ISE) by a simple filtering operation is also discussed.

Let us consider the problem of approximating an ideal frequency response $F(j\omega)$ by a physically realizable filter. The impulse response of any physically realizable filter can exist only for positive values of time (6). Let $H(j\omega)$ be the frequency response of such a filter. If $f(t)$ and $h(t)$ are the inverse Fourier Transforms of $F(j\omega)$ and $H(j\omega)$ respectively, we have

$$F(j\omega) = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t}dt$$

and

$$H(j\omega) = \int_{-\infty}^{+\infty} h(t)e^{-j\omega t}dt$$

$$= \int_{0}^{\infty} h(t)e^{-j\omega t}dt \qquad (2.1)$$

Let $E(j\omega)$ be the error in the frequency domain and $e(t)$

be the error in time domain.

Then

$$E(j\omega) = F(j\omega) - H(j\omega)$$

and $$e(t) = f(t) - h(t) \tag{2.2}$$

There exists a vast body of literature on approximating specific examples of $F(j\omega)$ by making use of purely frequency domain techniques (7-10). This research develops a general technique of design of filters that can be adopted to any specific example without difficulty. It may also find application in the design of optimum filters in signal theory (11).

LEAST INTEGRAL SQUARED ERROR FILTERS

All methods of approximating $F(j\omega)$ by $H(j\omega)$ try to minimize the error $E(j\omega)$ in equation (2.2) in some sense. The family of filters having minimum integral squared error in frequency domain has not received attention until recently (1-2). This work develops a method of design based on minimizing the integral squared error $\int_{-\infty}^{+\infty}|E(j\omega)|^2 d\omega$. The filters satisfying this criterion are defined as Least Integral Squared Error Filters.

The integral squared error (ISE) in frequency domain and time domain are related by Parsevel's theorem

(12, 16). It can be easily shown that, if e(t) belongs to $L(-\infty, \infty)$,

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}|E(j\omega)|^2 d\omega = \int_{-\infty}^{+\infty}|e(t)|^2 dt \qquad (2.3)$$

where e(t) and $E(j\omega)$ are defined in equation (2.2). The function h(t) in equation (2.2) exists only for $t \geqslant 0$. Hence it cannot be used to approximate f(t) for $t < 0$. Therefore it is assumed that f(t) vanishes for negative values of t. Equation (2.3) can now be written as

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}|E(j\omega)|^2 d\omega = \int_0^{\infty}|e(t)|^2 dt \qquad (2.4)$$

Equation (2.4) converts the problem of least square design in the frequency domain to a problem of least square design in time domain. This makes it possible to use time domain approximation theory for frequency domain design. If f(t), the inverse of $F(j\omega)$, is known the problem of design in the least ISE sense reduces to finding h(t) which minimizes the expression $\int_0^{\infty}|e(t)|^2 dt$. The choice of h(t) is governed by the factor that the filter is to be realized from passive, lumped linear circuit elements. This requires that H(s), the Laplace transform of h(t), be a rational function of s with all the poles lying in the left half s plane. The impulse response of such circuits can only be a finite sum of damped exponentials. (The case of multiple poles is exceptional and is not considered.) This suggests that the best form of approximation for f(t) is by a sum of damped exponentials.

EXPONENTIAL APPROXIMATION

Let f(t) be approximated by h(t) which is a linear combination of N damped exponentials.

$$h(t) = \sum_{i=1}^{N} \beta_i e^{s_i t} \tag{2.5}$$

s_i is, in general, a complex number with $\mathrm{Re}(s_i)<0$. Let I be the time domain integral squared error.
Then

$$I = \int_0^\infty \{f(t) - \sum_{i=1}^{N} \beta_i e^{s_i t}\}^2 dt \tag{2.6}$$

The parameters $\{\beta_i\}_{i=1}^{N}$ and $\{s_i\}_{i=1}^{N}$ have to be determined such that I is a minimum. This problem was first discussed by Aigrain and Williams and they formulated a set of equations known as Aigrain and Williams equations (13). These equations are of classical interest and are necessary to understand the complexity involved in exponential approximation.

Differentiating the equation (2.6) with respect to β_k and s_k we get,

$$\frac{\partial I}{\partial \beta_k} = \frac{\partial}{\partial \beta_k}\int_0^\infty \{f(t) - \sum_{i=1}^{N} \beta_i e^{s_i t}\}^2 dt$$

$$= 2\int_0^\infty \{f(t) - \sum_{i=1}^{N} \beta_i e^{s_i t}\}(-e^{s_k t})dt \qquad (2.7)$$

$$\frac{\partial I}{\partial s_k} = 2\int_0^\infty \{f(t) - \sum_{i=1}^{N} \beta_i e^{s_i t}\}(-\beta_k t e^{s_k t})dt \qquad (2.8)$$

For I to be minimum,

$$\frac{\partial I}{\partial \beta_k} = 0$$
$$\frac{\partial I}{\partial s_k} = 0, \; k = 1,2,\ldots N \qquad (2.9)$$

It is possible to simplify the equations (2.7-2.9) as follows.

By definition,

$$\int_0^\infty f(t)e^{s_k t}dt = F(-s_k) \qquad (2.10)$$

where F(s) is the Laplace transform of f(t).

Similarly,

$$\int_0^\infty e^{s_i t}\, e^{s_k t}\, dt = \frac{-1}{s_k + s_i} \qquad (2.11)$$

$$-\int_0^\infty tf(t)e^{s_k t}dt = \frac{d}{ds}F(s) \text{ at } s = -s_k$$

$$= F'(-s_k) \qquad (2.12)$$

and

$$\int_0^\infty te^{(s_i + s_k)t}dt = \frac{1}{(s_i + s_k)^2} \qquad (2.13)$$

Substituting equations (2.10) and (2.11) in equations (2.7) and (2.9) we get,

$$\sum_{i=1}^{N} \frac{\beta_i}{(s_i + s_k)} = -F(-s_k),\ k = 1,2,\ldots N \qquad (2.14)$$

Similarly substituting equations (2.12) and (2.13) in equations (2.8) and (2.9),

$$\sum_{i=1}^{N} \frac{\beta_i}{(s_i + s_k)^2} = -F'(-s_k),\ k = 1,2,\ldots N \qquad (2.15)$$

Equations (2.14) and (2.15) are known as Aigrain and Williams equations (13). There are 2N unknown complex parameters and 2N equations. The analytical solution of these 2N nonlinear simultaneous equations is extremely difficult; but these equations suggest the probable existence of a set of parameters $\{\beta_i\}_{i=1}^N$ and $\{s_i\}_{i=1}^N$ which will make the ISE, I, a minimum. Moreover any direct solution of these equations requires the availability of F(s) and F'(s) at any point. Because of these factors no attempt is made to solve these equations. Young and Huggins have shown that the complexity involved in tackling these equations could be greatly reduced by making use of orthonormal functions of exponentials (18-21).

ORTHONORMAL FUNCTIONS OF EXPONENTIALS

Let $\{\phi_i(t)\}_{i=1}^{\infty}$ be a set of orthonormal functions defined over $(0, \infty)$. Then

$$\begin{aligned} \int_0^{\infty}\phi_i(t)\phi_j(t)dt &= 1, \text{ if } i = j \\ &= 0, \text{ if } i \neq j \end{aligned} \tag{2.16}$$

Any real function $f(t)$, $t\varepsilon(0,\infty)$, for which $\int_0^{\infty}|f(t)|^2dt < \infty$ can be expanded in terms of these orthonormal functions. If the set of orthonormal functions is complete and if the series $\sum_{i=1}^{\infty} C_i\phi_i(t)$ converges uniformly, we have (12, 14)

$$f(t) = \sum_{i=1}^{\infty} C_i\phi_i(t) \tag{2.17}$$

where

$$C_i = \int_0^{\infty} f(t)\phi_i(t)dt \tag{2.18}$$

If $f(t)$ is approximated by N terms, the approximation $h(t)$ is given by

$$h(t) = \sum_{i=1}^{N} C_i\phi_i(t) \tag{2.19}$$

The error of approximation is

$$e(t) = f(t) - h(t)$$

$$= f(t) - \sum_{i=1}^{N} C_i \phi_i(t) \qquad (2.20)$$

It can be shown that ISE is given by, (14),

$$I = \int_0^\infty |e(t)|^2 dt = \int_0^\infty |f(t)|^2 dt - \sum_{i=1}^{N} C_i^2 \qquad (2.21)$$

It is possible to construct a set of orthonormal functions of exponentials by Gram-Schmidt process (12, 14). Another method of greater practical value was developed by Kautz (17) which was later improved by Young and Huggins (18-20) and Ross (21). This has been widely used in the analysis of speech signals (22-23) but has not been adapted and made use of in filter design. This dissertation makes use of these orthonormal functions.

Let it be required to construct a set of N orthonormal functions from N exponentials of which there are r real exponentials and c pairs of complex conjugate exponentials. It is assumed that all the exponentials have negative real parts.

Consider the frequency domain representations of N functions as given below.

$$\Phi_1(s) = \sqrt{-2s_1}\,\frac{1}{s - s_1}$$

. . . .

$$\Phi_r(s) = \left\{\prod_{i=1}^{r-1} \frac{s + s_i}{s - s_i}\right\} \frac{\sqrt{-2s_r}}{s - s_r}$$

$$\Phi_{r+1}(s) = \left\{\prod_{i=1}^{r} \frac{s + s_i}{s - s_i}\right\} \frac{\sqrt{2p_1}\;s}{s^2 + p_1 s + q_1}$$

$$\Phi_{r+2}(s) = \left\{\prod_{i=1}^{r} \frac{s + s_i}{s - s_i}\right\} \frac{\sqrt{2p_1 q_1}}{s^2 + p_1 s + q_1}$$

where

$$p_1 = -(s_{r+1} + s_{r+2}),$$

$$q_1 = |s_{r+1}|^2 = |s_{r+2}|^2.$$

s_{r+1} and s_{r+2} form a pair of complex conjugate exponentials

. . . .

$$\Phi_{N-1}(s) = \left\{\prod_{i=1}^{N-2} \frac{s + s_i}{s - s_i}\right\} \frac{\sqrt{2p_c}\;s}{s^2 + p_c s + q_c}$$

$$\Phi_N(s) = \left\{\prod_{i=1}^{N-2} \frac{s + s_i}{s - s_i}\right\} \frac{\sqrt{2p_c q_c}}{s^2 + p_c s + q_c} \qquad (2.22)$$

where

$$p_c = -(s_{N-1} + s_N)$$

$$q_c = |s_{N-1}|^2 = |s_N|^2,$$

s_{N-1} and s_N form the last pair of complex conjugate exponentials.

Functions Φ_1 to Φ_r correspond to the set of real exponentials, Φ_{r+1} and Φ_{r+2} to the first pair of complex exponentials and Φ_{N-1} and Φ_N to the last pair of complex exponentials. It will now be proved that the corresponding time domain functions $\{\phi_i(t)\}_{i=1}^N$ form an orthonormal set. This is done by using Parsevel's theorem. By this theorem we have,

$$\int_0^\infty \phi_m(t)\phi_n(t)dt = \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty} \Phi_m(s)\Phi_n(-s)ds \qquad (2.23)$$

where $j = \sqrt{-1}$.

Consider two functions $\Phi_{r+2i-1}(s)$ and $\Phi_{r+2i}(s)$ corresponding to a pair of complex conjugate exponentials s_{r+2i-1} and s_{r+2i}. Using equation (2.22) we get,

$$\Phi_{r+2i-1}(s)\Phi_{r+2i}(-s) = \frac{2p_i\sqrt{q_i}\ s}{(s^2 + p_is + q_i)(s^2 - p_is + q_i)} \qquad (2.24)$$

where $(s^2 + p_is + q_i) = (s - s_{r+2i-1})(s - s_{r+2i}) \qquad (2.25)$

The evaluation of $\int_{-j\infty}^{+j\infty}\Phi_{r+2i-1}(s)\Phi_{r+2i}(-s)ds$ may be done by taking the closed contour consisting of the imaginary axis and left side semicircle at infinity. Thus by Cauchy residue theorem (25) we have

$$\int_{-j\infty}^{+j\infty}\Phi_{r+2i-1}(s)\Phi_{r+2i}(-s)ds = 2\pi j(R_1 + R_2) \tag{2.26}$$

where R_1 and R_2 are the residues of $\{\Phi_{r+2i-1}(s)\Phi_{r+2i}(-s)\}$ at s_{r+2i-1} and s_{r+2i} respectively. Evaluating the residues it can be easily shown that $(R_1 + R_2)$ is zero. This proves that functions $\phi_{r+2i-1}(t)$ and $\phi_{r+2i}(t)$ are orthogonal to each other. Similarly

$$\Phi_{r+2i}(s)\Phi_{r+2i}(-s) = \frac{2p_iq_i}{(s^2 + p_is + q_i)(s^2 - p_is + q_i)} \tag{2.27}$$

and

$$\Phi_{r+2i-1}(s)\Phi_{r+2i-1}(-s) = \frac{-2p_is^2}{(s^2 + p_is + q_i)(s^2 - p_is + q_i)} \tag{2.28}$$

Integrating these functions along the closed contour enclosing the left half s-plane it can be shown that

$$\frac{1}{2\pi j}\int_{-j\infty}^{+j\infty}\Phi_{r+2i}(s)\Phi_{r+2i}(-s)ds = 1$$

$$\text{and} \quad \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty}\Phi_{r+2i-1}(s)\Phi_{r+2i-1}(-s)ds = 1 \tag{2.29}$$

Thus the orthonormality of the functions $\phi_{r+2i-1}(t)$ and $\phi_{r+2i}(t)$ is proved.

If any two functions $\Phi_m(s)$ and $\Phi_n(s)$, which do not belong to the special case discussed above, are considered it can be shown by pole-zero cancellation, that all the poles of $\Phi_m(s)\ \Phi_n(-s)$ lie only on one half (left or right) of s-plane and the function is hence analytic in the opposite half of s-plane. Therefore it can be proved that

$$\int_{-j\infty}^{+j\infty}\Phi_m(s)\Phi_n(-s)ds = 0 \qquad (2.30)$$

Considering any function $\Phi_i(s)$, which is formed from real poles only, we have

$$\Phi_i(s)\Phi_i(-s) = \frac{2s_i}{s^2 - s_i^2} \qquad (2.31)$$

Evaluating the countour integral by Cauchy Residue theorem we can prove

$$\frac{1}{2\pi j}\int_{-j\infty}^{+j\infty}\Phi_i(s)\Phi_i(-s)ds = 1 \qquad (2.32)$$

Thus the orthonormality of functions $\{\phi_i(t)\}_{i=1}^{N}$ is proved.

COMPLIMENTARY FILTER AND ISE

When ISE is minimum with respect to the parameters $\{\beta_i, s_i\}_{i=1}^N$ of equation (2.5)

$$\frac{\partial I}{\partial \beta_i} = 0$$

$$\text{and} \quad \frac{\partial I}{\partial s_i} = 0$$

From equations (2.7) and (2.8) we get

$$\frac{\partial I}{\partial \beta_i} = -2\int_0^\infty e(t)e^{s_i t}dt = 0, \; i = 1,2,\ldots N$$

$$\text{and} \quad \frac{\partial I}{\partial s_i} = -2\beta_i\int_0^\infty e(t)te^{s_i t}dt = 0, \; i = 1,2,\ldots N \qquad (2.33)$$

Equation (2.33) gives the condition for I to be a minimum as

$$\int_0^\infty e(t)e^{s_i t}dt = 0, \; i = 1,2,\ldots N$$

$$\text{and} \quad \int_0^\infty e(t)te^{s_i t}dt = 0, \; i = 1,2,\ldots N \qquad (2.34)$$

In other words, the condition for I to be minimum is that e(t) should be orthogonal to $\{e^{s_i t}\}_{i=1}^N$ and $\{te^{s_i t}\}_{i=1}^N$. These properties of e(t) are made use of by Young and Huggins (18-20) in developing a simple method of evaluating ISE. This is now discussed.

In the formation of the orthonormal functions $\Phi_k(s)$, it is observed that each $\Phi_k(s)$ has an all pass factor of the form $\{\prod_{i=1}^{n} \frac{s + s_i}{s - s_i}\}$. The degree of this factor successively increases. If N exponentials are used for the approximation and it is required to add one real pole or a pair of complex poles, the all pass filter appearing in the corresponding orthonormal function or functions will be $\{\prod_{i=1}^{N} \frac{s + s_i}{s - s_i}\}$. This all pass factor is called the complimentary filter of Nth degree approximation (18-20). This filter has the interesting property that if the function f(t) is time reversed and filtered using this complimentary filter, say G(s), the integrated squared error can be directly evaluated from the output of G(s) (18-20). This property can be proved as follows.

Let v(t) be the time reversed signal of f(t).

$$v(t) = f(-t), \; -\infty < t \leqslant 0 \qquad v(t) = 0 \quad , \; t > 0 \tag{2.35}$$

The approximation h(t) may also be reversed to get the corresponding approximation $v_a(t)$ of v(t). Hence,

$$v_a(t) = h(-t), \; -\infty < t \leqslant 0$$

and
$$v_a(t) = 0, \qquad t > 0 \tag{2.36}$$

Since $h(t) = \sum_{i=1}^{N} C_i \phi_i(t)$

$$h(-t) = \sum_{i=1}^{N} C_i \phi_i(-t) \tag{2.37}$$

Since h(t) can also be expressed as a sum of exponentials as in equation (2.5),

$$h(-t) = \sum_{i=1}^{N} \beta_i e^{-s_i t} \tag{2.38}$$

Comparing equations (2.37) and (2.38) it is seen that the approximation of v(t) can be done solely on negative time using growing exponentials as against the decaying exponentials used for approximation in positive time. In the frequency domain the corresponding functions will be $\{\Phi_i(-s)\}_{i=1}^{N}$ instead of $\{\Phi_i(s)\}_{i=1}^{N}$. Two-sided Laplace transform is used to get the frequency domain functions of negative time functions (24).

Let

$$\tilde{\Phi}_k(s) = \Phi_k(-s) \tag{2.39}$$

The approximation $V_a(s)$ is

$$V_a(s) = \sum_{i=1}^{N} C_i \tilde{\Phi}_i(s) \tag{2.40}$$

Let $\varepsilon(t)$ be the error in approximating v(t). Then,

$$v(t) = v_a(t) + \varepsilon(t), \ -\infty<t\leqslant 0 \qquad (2.41)$$

ISE, I is given by

$$I = \int_0^\infty |e(t)|^2 dt = \int_{-\infty}^{0} |\varepsilon(t)|^2 dt \qquad (2.42)$$

The complimentary filter of Nth degree approximation is

$$G(s) = \prod_{i=1}^{N} \frac{s + s_i}{s - s_i} \qquad (2.43)$$

Let $\bar{a}(t)$ be the output of this filter when v(t), the time reversed f(t), is applied as an input. The output $\bar{a}(t)$ exists for t extending from $-\infty$ to $+\infty$. The Laplace transform $\bar{A}(s)$ of $\bar{a}(t)$ is obtained as

$$\begin{aligned} \bar{A}(s) &= \left\{ \sum_{i=1}^{N} C_i \tilde{\Phi}_i(s) + \bar{\varepsilon}(s) \right\} G(s) \\ &= \left\{ \prod_{i=1}^{N} \frac{s + s_i}{s - s_i} \ \sum_{i=1}^{N} C_i \tilde{\Phi}_i(s) \right\} + \bar{\varepsilon}(s) G(s) \qquad (2.44) \end{aligned}$$

where $\bar{\varepsilon}(s)$ is the Laplace transform of $\varepsilon(t)$.

Any $\tilde{\Phi}_k(s)$, formed from real roots alone is given by

$$\tilde{\Phi}_k(s) = \Phi_k(-s) = \prod_{i=1}^{k-1} \frac{-s + s_i}{-s - s_i} \frac{\sqrt{-2s_k}}{-s - s_k} \qquad (2.45)$$

Any pair of orthonormal functions $\tilde{\Phi}_{r+2k-1}(s)$ and $\tilde{\Phi}_{r+2k}(s)$ corresponding to a pair of complex conjugate exponentials

s_{r+2k-1} and s_{r+2k} are given as

$$\tilde{\Phi}_{r+2k-1}(s) = \Phi_{r+2k-1}(-s)$$

$$= \left\{\prod_{i=1}^{r+2k-2} \frac{-s + s_i}{-s - s_i}\right\} \frac{\sqrt{2p_k}\,(-s)}{s^2 - p_k s + q_K}$$

and $$\tilde{\Phi}_{r+2k}(s) = \Phi_{r+2k}(-s) \qquad (2.46)$$

$$= \left\{\prod_{i=1}^{r+2k-2} \frac{-s + s_i}{-s - s_i}\right\} \frac{\sqrt{2p_k q_k}}{s^2 - p_k s + q_K}$$

where $(s^2 + p_k s + q_K) = (s - s_{r+2k-1})(s - s_{r+2k})$

Examining equations (2.44), (2.45) and (2.46) it is seen that in the expression for $\bar{A}(s)$, any term like

$$\tilde{\Phi}_k(s) \left\{\prod_{i=1}^{N} \frac{s + s_i}{s - s_i}\right\}$$

has all numerator factors of $\tilde{\Phi}_k(s)$ getting cancelled with denominator factors of G_(s) and all the denominator factors of $\tilde{\Phi}_k(s)$ get cancelled with numerators of G(s). Hence the poles of any term like

$$\tilde{\Phi}_k(s) \left\{\prod_{i=1}^{N} \frac{s + s_i}{s - s_i}\right\}$$

are solely determined by the remaining denominator factors of $\prod_{i=1}^{N} \frac{s + s_i}{s - s_i}$. It can thus be concluded that the expression

$$G(s)\left\{\sum_{i=1}^{N} C_i\tilde{\Phi}_i(s)\right\}$$

in equation (2.44) has all its poles lying in the left half s plane. The remaining term in equation (2.44), namely, $\varepsilon(s)G(s)$ may now be considered. Since e(t) has been proved to be orthogonal to $\{e^{s_i t}\}_{i=1}^{N}$, $\varepsilon(t)$ is orthogonal to $\{e^{-s_i t}\}_{i=1}^{N}$.

Or

$$\int_{-\infty}^{0}\varepsilon(t)e^{-s_i t}dt = 0, \; i = 1,2,\ldots N \qquad (2.47)$$

This orthogonality expressed in the frequency domain becomes

$$\int_{-j\infty}^{+j\infty} \bar{\varepsilon}(s)\frac{1}{s - s_i}ds = 0, \; i = 1,2,\ldots N \qquad (2.48)$$

Evaluating this integral we get

$$\bar{\varepsilon}(s_i) = 0, \; i = 1,2,\ldots N \qquad (2.49)$$

This proves that $\bar{\varepsilon}(s)$ has zeros at $\{s_i\}_{i=1}^{N}$.

Considering the factor $G(s)\bar{\varepsilon}(s)$, it is found that all the denominator factors of G(s) get cancelled with the zeros of $\bar{\varepsilon}(s)$. Hence all the poles of $G(s)\bar{\varepsilon}(s)$ are determined by the poles of $\bar{\varepsilon}(s)$ and so $G(s)\bar{\varepsilon}(s)$ has poles only in the right half s-plane. Thus $\bar{A}(s)$ can be split into two terms A'(s) and A''(s) such that

$$\bar{A}(s) = A'(s) + A''(s) \tag{2.50}$$

where $A'(s) = G(s)\bar{\varepsilon}(s)$ has all poles in the right half s-plane and $A''(s) = G(s)\sum_{i=1}^{N} C_i\tilde{\Phi}_i(s)$ has all poles in the left half s-plane. The output $\bar{a}(t)$, of the complimentary filter, with reversed f(t) as input, extends from $-\infty$ to $+\infty$. This is also split into two parts such that

$$\bar{a}(t) = a'(t) + a''(t)$$

$$a'(t) = 0, \; t>0$$

and

$$a''(t) = 0, \; t<0, \tag{2.51}$$

a'(t) and a''(t) correspond to A'(s) and A''(s) respectively. It is easily seen that a'(t) is the output of the complimentary filter during negative time and a''(t) during positive time.

By Parsevel's theorem

$$\int_{-\infty}^{0} |a'(t)|^2 dt = \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty} A'(s)A'(-s)ds$$

$$= \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty} G(s)\bar{\varepsilon}(s)G(-s)\bar{\varepsilon}(s)ds$$

Since $G(s)G(-s) = 1$, we obtain

$$\int_{-\infty}^{0}|a'(t)|^2dt = \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty} \bar{\varepsilon}(s)\bar{\varepsilon}(-s)ds$$

$$= \int_{-\infty}^{0}|\varepsilon(t)|^2dt$$

$$= \int_{0}^{\infty}|e(t)|^2dt.$$

Hence ISE is given by

$$I = \int_{0}^{\infty}|e(t)|^2dt = \int_{-\infty}^{0}|\bar{a}(t)|^2dt \qquad (2.52)$$

where $\bar{a}(t)$ is the output of the complimentary filter with reversed f(t) as input.

This is schematically represented in Fig. (2.1). Methods of implementing this on digital computer are discussed in Chapter 6.

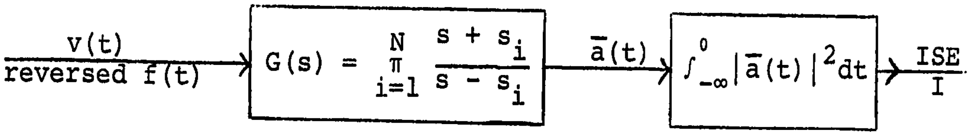

$v(t)$
reversed $f(t)$
$G(s) = \prod_{i=1}^{N} \frac{s + s_i}{s - s_i}$
$\bar{a}(t)$
$\int_{-\infty}^{0}|\bar{a}(t)|^2dt$
$\frac{ISE}{I}$

Figure 2.1. Schematic diagram for evaluating ISE

The advantage of employing this technique of computing I is that it can be directly evaluated from the pole positions even without forming the functions $\{\phi_i(t)\}_{i=1}^{N}$ and evaluating the coefficients $\{C_i\}_{i=1}^{N}$. Moreover, in minimizing I, as given by Fig. 2.1, we need only consider

N parameters, instead of the original 2N parameters of equations (2.14) and (2.15).

CHAPTER 3

A SURVEY OF PREVIOUS RESEARCH

In this chapter we discuss some of the important methods that have been used to determine the parameters of H(s) in the least square sense. Some of these methods tackle the problem purely as a time domain approximation, while others take the frequency domain behaviour into consideration.

PRONY'S METHOD

The first attempt to approximate a time function by the sum of weighted exponentials was due to Prony (26). Prony chose the function

$$h(t) = \sum_{i=1}^{N} \beta_i e^{s_i t} \qquad (3.1)$$

such that h(t) passes through 2N equally spaced points of the function f(t) where f(t) is the function to be approximated by h(t). Let f(t) be specified at points $\{t_i\}_{i=0}^{2N-1}$ where $t_i = iT$, T being the interval between two successive points. We then have,

$$h(t_k) = \sum_{i=1}^{N} \beta_i e^{s_i t_k}$$

$$= \sum_{i=1}^{N} \beta_i (e^{s_i T})^k$$

$$= f(t_k),\ k = 0,1,\ldots,2N-1 \quad (3.2)$$

Let $e^{s_i T} = x_i$. Equation (3.2) can now be written as

$$\sum_{i=1}^{N} \beta_i x_i^k = f(t_k),\ k = 0,1,\ldots,2N-1 \quad (3.3)$$

Equation (3.3) is nonlinear in x_i. This nonlinearity can be removed by the following procedure.

Let $\{\alpha_i\}_{i=0}^{N}$ be a new set of variables defined by

$$\sum_{i=0}^{N} \alpha_i x^i = \prod_{i=1}^{N} (x - x_i),\ \alpha_N = 1 \quad (3.4)$$

where $\{x_i\}_{i=1}^{N}$ are the roots of the equation

$$\sum_{i=0}^{N} \alpha_i x^i = 0 \quad (3.5)$$

Hence we have

$$\sum_{i=0}^{N} \alpha_i x_k^i = 0,\ k = 1,2,\ldots,N \quad (3.6)$$

Multiplying both sides of equation (3.3) by α_i and changing the index k to k+i, equation (3.3) can be expressed as

$$\alpha_i f(t_{k+i}) = \alpha_i \sum_{j=1}^{N} \beta_j x_j^{k+i} \qquad (3.7)$$

From equation (3.7) we get

$$\sum_{i=0}^{N} \alpha_i f(t_{k+i}) = \sum_{j=1}^{N} \beta_j x_j^k \sum_{i=0}^{N} \alpha_i x_j^i$$

Since, by equation (3.6), $\sum_{i=0}^{N} \alpha_i x_j^i = 0$ we have

$$\sum_{i=0}^{N} \alpha_i f(t_{k+i}) = 0, \; k = 0,1,\ldots,N-1 \qquad (3.8)$$

Since α_N is assumed to be unity, equation (3.8) may be written as

$$\sum_{i=0}^{N-1} \alpha_i f(t_{k+i}) = -f(t_{k+N}) \quad k = 0,1,\ldots,N-1 \qquad (3.9)$$

Equation (3.9) gives N simultaneous linear equations in N unknowns $(\alpha_i)_{i=0}^{N-1}$. Once α_i's are known, the solution of the polynomial equation (3.5) gives $(e^{s_i T})_{i=1}^{N}$ and the β_i values are obtained as the solution of (3.3).

The basic philosophy of Prony's approach is made clear by equations (3.9) and (3.5). Equation (3.9) defines a difference equation and (3.5) corresponds to the characteristic equation of this difference equation. This idea of determining the exponentials of the approximation from the coefficients of an equivalent difference or differential equation is common to most of the methods reviewed in this chapter. Prony's method has also been extended to the case where the available number of points of f(t) is more than 2N, the number of unknown parameters.

KAUTZ'S METHOD (17)

If f(t), the function to be approximated is the sum of N weighted exponentials, then f(t) is the solution of an Nth order, constant coefficient linear differential equation. This means

$$\sum_{i=0}^{N} \alpha_i D^i f(t) = 0, \ \alpha_N = 1 \qquad (3.10)$$

where $D \equiv \frac{d}{dt}$.

The characteristic equation of equation (3.10) is

$$\sum_{i=0}^{N} \alpha_i s^i = 0 \qquad (3.11)$$

and the solution of equation (3.11) gives $(s_i)_{i=1}^{N}$.

But, in general f(t) is not an exact sum of exponentials and equation (3.10) is not satisfied. Equation (3.10) is now modified as

$$\sum_{i=0}^{N} \alpha_i D^i f(t) = e(t), \; \alpha_N = 1 \qquad (3.12)$$

Here e(t) is not the exact error of approximation but is a measure of how closely f(t) can be expressed as a sum of N exponentials. The $(\alpha_i)_{i=0}^{N-1}$ are chosen so that $\int_0^\infty |e(t)|^2 dt$ is minimum. Differentiating the expression for $\int_0^\infty |e(t)|^2 dt$ with respect to α_k and equating it to zero we get,

$$\frac{\partial}{\partial \alpha_k} \int_0^\infty |e(t)|^2 dt = 2\int_0^\infty e(t) \frac{\partial}{\partial \alpha_k} e(t) dt = 0 \qquad (3.13)$$

$$k = 0,1,\ldots,N-1$$

Substituting equation (3.12) into (3.13) we obtain

$$\sum_{i=0}^{N-1} \left\{\int_0^\infty D^i f(t) D^k f(t) dt\right\} \alpha_i = -\int_0^\infty D^N f(t) D^k f(t) dt \qquad (3.14)$$

$$k = 0,1,\ldots,N-1$$

Equation (3.14) gives N equations in N unknowns. If N derivatives of f(t) are known, equation (3.14) can be solved for $\{\alpha_i\}_{i=0}^{N-1}$. Hence the method is useful when f(t) is analytically known.

This method obviously cannot claim to have minimized the ISE between f(t) and the approximation h(t). If f(t) has any discontinuity, the derivatives of f(t) at these points of discontinuity are undefined and hence the method has to be modified. Another point worth noting is that, instead of $\alpha_N = 1$ in equation (3.12) any of the α_i can be taken as unity and this will lead to a different set of solutions for the exponentials (26). An interesting review of this and similar methods, which basically make use of the approach due to Prony, can be found in McDonough (26).

YENGST'S METHOD (27)

Yengst's method resembles the approach of Prony in that he assumes a difference equation, the solution of which gives h(t), which approximates the desired function f(t). The Nth order difference equation used by Yengst is

$$a_1 f_{i-1} + a_2 f_{i-2} + \dots + a_N f_{i-N} = h_i, \; i \geqslant N \qquad (3.15)$$

where $f_{i-1}, f_{i-2}, \dots, f_{i-N}$ are equally spaced values of f(t) at times (i-1)T, (i-2)T, ..., (i-N)T and h_i is the value of the approximation at time $t = iT$. (T is the interval between two consecutive points.)

The error at the i'th instant is

$$e_i = f_i - h_i \qquad (3.16)$$

Let q be the total number of samples of f(t) available. The coefficients $(a_i)_{i=1}^{N}$ are so chosen that $\sum_{i=N}^{q} e_i^2$ is minimized. Substituting equation (3.15) into (3.16) we have,

$$\sum_{i=N}^{q} e_i^2 = \sum_{i=N}^{q} \left(f_i - \sum_{k=1}^{N} a_k f_{i-k}\right)^2 \qquad (3.17)$$

Equating the derivatives with respect to the coefficients $(a_j)_{j=1}^{N}$ to zero,

$$\frac{\partial}{\partial a_j} \sum_{i=N}^{q} \left(f_i - \sum_{k=1}^{N} a_k f_{i-k}\right)^2 = 0$$

$$j = 1,2,\ldots N$$

Or

$$\sum_{i=N}^{q} \left(f_i - \sum_{k=1}^{N} a_k f_{i-k}\right) f_{i-j} = 0$$

This gives

$$\sum_{i=N}^{q} \sum_{k=1}^{N} a_k f_{i-k} f_{i-j} = \sum_{i=N}^{q} f_i f_{i-j} \qquad (3.18)$$

$$j = 1,2,\ldots, N$$

Equation (3.18) can be expressed as a matrix equation as follows

$$\begin{bmatrix} \sum_{i=N}^{q} f_{i-1}^2 & \sum_{i=N}^{q} f_{i-1}f_{i-2} & \cdots\cdots & \sum_{i=N}^{q} f_{i-1}f_{i-N} \\ \sum_{i=N}^{q} f_{i-1}f_{i-2} & \sum_{i=N}^{q} f_{i-2}^2 & \cdots\cdots\cdots & \sum_{i=N}^{q} f_{i-2}f_{i-N} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \sum_{i=N}^{q} f_{i-1}f_{i-N} & \sum_{i=N}^{q} f_{i-1}f_{i-N} & \cdots\cdots & \sum_{i=N}^{q} f_{i-N}^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_N \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=N}^{q} f_i f_{i-1} \\ \sum_{i=N}^{q} f_i f_{i-2} \\ \cdots\cdots \\ \cdots\cdots \\ \cdots\cdots \\ \sum_{i=N}^{q} f_i f_{i-N} \end{bmatrix} \qquad (3.19)$$

If the N x N matrix of equation (3.19) becomes singular no solution for $(a_i)_{i=1}^{N}$ exists. This means that the given function f(t) can be approximated by a smaller number of exponentials. If the matrix is non singular, equation (3.19) can be solved for $(a_i)_{i=1}^{N}$. The exponentials are then obtained from the solution of the polynomial equation

$$Z^n + a_1 Z^{n-1} + \dots. + a_{N-1} Z + a_N = 0 \qquad (3.20)$$

Yengst (27) has shown that $(e^{s_i T})_{i=1}^{N}$ are the solutions of the equation (3.20). It may be noted that equation (3.20) corresponds to the characteristic equation of the difference equation (3.15).

In Yengst's method, the zeros of H(s) are determined by the successive application of the initial value theorem to obtain expressions of initial values of h(t) and its (N-1) derivatives at origin. These are then equated to the corresponding values of f(t). The solution of the N simultaneous linear equations, thus formed, gives the zeros of H(s). This method is easy to apply if f(t) is known analytically. If f(t) is specified as samples, Yengst suggests a polynomial interpolation of f(t) at t = 0. This method of determination of zeros is not very convenient when f(t) is not known analytically.

METHOD OF PERDIKARIS AND LAGO

A recent paper by Perdikaris and Lago (28), on time domain synthesis, is very closely related to the method of Yengst. This paper makes use of Z-domain approach instead of the s-domain approach of Yengst. An Nth order difference equation is first assumed.

$$y_i = \sum_{k=1}^{N} a_k y_{i-k} \quad N \leqslant i \leqslant q \tag{3.21}$$

$y_{i-k} = f_{i-k}$ is the $(i-k)^{th}$ sample of $f(t)$.

Equation (3.21) can be expressed as

$$y_{N+n} = \sum_{k=1}^{N} a_k y_{N+n-k} \quad 0 \leqslant n \leqslant q-N \tag{3.22}$$

By making use of the properties of Z-transforms (29) we get

$$z(y_{n+N}) = Z^N \left(Y(Z) - \sum_{k=0}^{N-1} y_k Z^{-k} \right) \tag{3.23}$$

where $z(y_n) = Y(Z)$.

Equations (3.22) and (3.23) give

$$\left(Z^N - \sum_{k=1}^{N} a_k Z^{N-k} \right) Y(Z) = \sum_{i=0}^{N-1} y_i Z^{N-i} - \sum_{k=1}^{N} a_k \sum_{i=0}^{N-k-1} y_i Z^{N-k-i} \tag{3.24}$$

If we consider H(Z) as the equivalent Z-transform of H(s) we have

$$H(Z) = Y(Z) = \frac{\sum_{i=0}^{N-1} y_i Z^{N-i} - \sum_{k=1}^{N} a_k \sum_{i=0}^{N-k-1} y_i Z^{N-k-i}}{Z^N - \sum_{k=1}^{N} a_k Z^{N-k}} \qquad (3.25)$$

The error criterion used by Perdikaris and Lago for determining $(a_i)_{i=1}^{N}$ is exactly the same as used by Yengst. Hence the error defined as $\sum_{i=N}^{q}(y_i - \sum_{k=1}^{N} a_k y_{i-k})^2$ leads to the same set of equations as those arrived at by Yengst in equation (3.19). Once $(a_i)_{i=1}^{N}$ are known the poles in the Z-domain are determined and the corresponding poles in s-domain are found from the relation

$$s_i = \frac{1}{T}\log Z_i \qquad (3.26)$$

The Z-transform expression of equation (3.25) is made use of to find the zeros of H(s). Since we have the relation

$$H(Z) \to H(s) \qquad (3.27)$$

and H(Z) is completely known from equation (3.25), the zeros of H(s) can be found by splitting H(Z) into partial fractions. The typical factors appearing in the partial fraction expansion of H(Z) and their s-domain equivalents are as below.

$$\frac{Z}{Z - e^{-aT}} \rightarrow \frac{1}{s + a}$$

$$\frac{Z^2 - Ze^{-aT}\cos bT}{Z^2 - 2Ze^{-aT}\cos bT + e^{-2aT}} \rightarrow \frac{s + a}{(s + a)^2 + b^2} \qquad (3.28)$$

and

$$\frac{Ze^{-aT}\sin bT}{Z^2 - 2Ze^{-aT}\cos bT + e^{-2aT}} \rightarrow \frac{b}{(s + a)^2 + b^2}$$

By making use of the partial fraction expansion of H(Z) and equation (3.28), H(s) can be found.

An alternate method of determining zeros of H(s), suggested by Perdikaris and Lago, is to force h(t) to pass through N points as specified by the desired function f(t). Since the poles of H(s) are already known, h(t) can be expressed as a sum of damped sine, cosine and exponential terms. The coefficients of these terms constitute the N unknowns which determine the zeros of H(s). Hence N equations are formed by choosing h(t) such that

$$h(t_i) = f(t_i),\ i = 1,2,\ldots N \qquad (3.29)$$

The solution of equation (3.29) gives these coefficients and the zeros of H(s) are found from them. This method has the advantage that h(t) can be made to approximate f(t) very closely in a short range but h(t) may deviate from f(t) considerably in other regions.

METHOD OF VASILU

A different approach of time domain synthesis is found in a paper by Vasilu (30). In this method the impulse response of a pulse transfer function H(Z) approximates the desired function f(t). The poles of the function H(Z) are assumed to be uniformly distributed on a circle of radius smaller than unity. Thus H(Z) has the form

$$H(Z) = \frac{b_N Z^N + b_{N-1} Z^{N-1} + \ldots + b_1 Z}{Z^N - a_0} \tag{3.30}$$

where $|a_0| < 1$.

The choice of a_0 is made arbitrarily. It is proved in (30) that if f(t) is zero for $t > T_0$, the error in the range $t > T_0$ can be reduced by choosing a_0 very small. But too small a value for a_0 will result in larger errors for $t < T_0$. The parameters $(b_i)_{i=1}^{N}$ are chosen so that the mean square error between the impulse response of H(Z) and the desired function f(t) at sampling instants is minimized.

METHODS OF McDONOUGH AND HUGGINS (26-31)

All the methods discussed so far do not minimize the actual ISE as defined by equation (2.6). We will now review some of the recent methods which minimize the ISE.

McDonough and Huggins make use of the complimentary filter to find a set of optimum poles. From equation (2.52) we have

$$I = \int_0^\infty |f(t) - h(t)|^2 dt = \int_{-\infty}^0 |\bar{a}(t)|^2 dt \qquad (3.31)$$

where $\bar{a}(t)$ is the output of the complimentary filter G(s) when the reversed signal f(-t) is applied as an input.

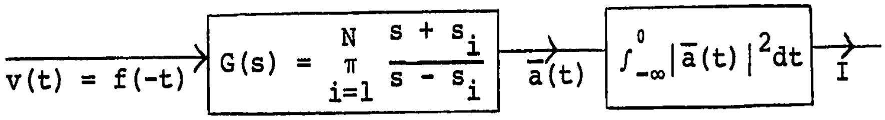

$v(t) = f(-t)$
$G(s) = \prod_{i=1}^{N} \frac{s + s_i}{s - s_i}$
$\bar{a}(t)$
$\int_{-\infty}^{0} |\bar{a}(t)|^2 dt$
I

Figure 3.1 Evaluation of ISE

Figure 3.1 shows how to evaluate I for a set of N exponentials. One method of computing the exponentials by minimizing I is to use gradient techniques. But, it is found that error I is very insensitive to pole positions over a wide range around the optimum poles (26). The values of the gradients at these points are nearly zeros. Hence gradient techniques are not suited to minimize I. This has led to the development of various iteration schemes of computing the best pole positions. These are now discussed.

The condition for I to be minimum with respect to $(\beta_i)_{i=1}^{N}$ and $(s_i)_{i=1}^{N}$ are expressed as (equation (2.34))

$$\int_0^\infty e(t)e^{s_i t}dt = 0 \qquad (3.32)$$

$$\int_0^\infty e(t)te^{s_i t}dt = 0 \qquad (3.33)$$

$$i = 1,2,\ldots N$$

McDonough and Huggins (26, 31) make use of equations (3.32) and (3.33) to develop an iterative scheme as follows

Let f(t), defined over $(0, \infty)$ be a signal vector in a Hilbert space S. The exponentials $(e^{s_i t})_{i=1}^{N}$ span a subspace S_N of S. Equation (3.32) implies that for I to be minimum the projection of e(t) on S_N should be the null vector. Let us now consider a subspace S_{2N} of S such that S_{2N} is spanned by $(e^{s_i t})_{i=1}^{N}$ and $(te^{s_i t})_{i=1}^{N}$. Equations (3.32) and (3.33) imply that, for I to be minimum, e(t) has to be orthogonal to the subspace S_{2N}. Let us assume that e(t) is always chosen such that equation (3.32) is satisfied.

Since e(t) = f(t) - h(t), h(t) is the projection of f(t) on S_N. If e(t) has a component on S_{2N}, it has to be the same as the projection of e(t) on $S_{2N} - S_N$. This is because e(t) is always chosen as orthogonal to S_N.

Hence we have,

Projection of e(t) on $S_{2N} - S_N$

= (Projection of f(t) on $S_{2N} - S_N$) - (Projection of h(t) on $S_{2N} - S_N$)

Since h(t) lies entirely on S_N, h(t) has no component on $S_{2N} - S_N$. Thus we get a modified condition for optimality of the poles. This condition is that, for ISE to be minimum, f(t) has to be orthogonal to the subspace $S_{2N} - S_N$.

McDonough and Huggins chose a set of orthonormal functions as bases for space $S_{2N} - S_N$ as follows (31). The functions $\{\phi_i(t)\}_{i=1}^{N}$ are already chosen as bases for S_N. S_{2N} is the space spanned by $(e^{s_i t})_{i=1}^{N}$ and $(te^{s_i t})_{i=1}^{N}$. If a set of orthonormal functions $\{\phi_i(t)\}_{i=N+1}^{2N}$ are defined such that they are linear combinations of $(e^{s_i t})_{i=1}^{N}$ and $(te^{s_i t})_{i=1}^{N}$, then these functions are bases for $S_{2N} - S_N$. Let us consider the following functions

$$\Phi_{N+i}(s) = G(s)\Phi_i(s), \ i = 1,2,\ldots N \qquad (3.34)$$

where

$$G(s) = \prod_{i=1}^{N} \frac{s + s_i}{s - s_i}$$

and $\Phi_i(s)$, $i = 1,2,\ldots N$ are as defined by equations 2.22. Hence the denominator of $\Phi_{N+i}(s)$ has terms like $(s - s_i)^2$. Therefore $\phi_{N+i}(t)$ is a linear combination of $(e^{s_i t})_{i=1}^{N}$ and $(te^{s_i t})_{i=1}^{N}$. Moreover

$$\int_{-j\infty}^{+j\infty}\Phi_{N+i}(s)\Phi_{N+k}(-s)ds = \int_{-j\infty}^{+j\infty}G(s)\Phi_i(s)G(-s)\Phi_k(-s)ds$$

$$= \int_{-j\infty}^{+j\infty}\Phi_i(s)\Phi_k(-s)ds$$

$$= 2\pi j\delta_{ik} \qquad (3.35)$$

where $j = \sqrt{-1}$

Hence $\phi_k(t)$, $k = 1,2,\ldots 2N$ form an orthonormal set and $\phi_k(t)$, $k = N+1, \ldots, 2N$ are bases for the subspace $S_{2N} - S_N$.

The condition for $f(t)$ to be orthogonal to the subspace $S_{2N} - S_N$ can now be stated as

$$\int_0^\infty f(t)\phi_{N+i}(t)dt = 0,\ i = 1,2,\ldots N \qquad (3.36)$$

where $\Phi_{N+i}(s) = G(s)\Phi_i(s)$

Equation (3.36) can be written as follows

$$\int_0^\infty f(t)\phi_{N+i}(t)dt = \int_{-\infty}^{0} f(-\tau)\phi_{N+i}(-\tau)d\tau$$

$$= \int_{-\infty}^{t} f(-\tau)\phi_{N+i}(t-\tau)d\tau \text{ at } t = 0$$

$$= 0,\ i = 1,2,\ldots,\ N \qquad (3.37)$$

If we consider the reversed signal of f(t) as

$$v(\tau) = f(-\tau),\ -\infty < \tau < 0 \qquad (3.38)$$

Equation (3.37) can now be written as

$$\int_0^\infty f(t)\phi_{N+i}(t)dt = \int_{-\infty}^{t} v(\tau)\phi_{N+i}(t-\tau)d\tau \Big|_{t=0}$$

$$\triangleq d_i \qquad (3.39)$$

The right hand side of equation (3.39) can be described as a filtering operation as shown in Figure 3.2

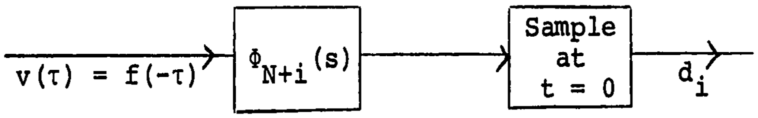


Figure 3.2 Evaluation of $\int_0^\infty f(t)\phi_{N+i}(t)dt$ by filtering operation

Since $\Phi_{N+i}(s) = G(s)\Phi_i(s)$, the filtering operation of Fig. 3.2 can be modified as shown in Fig. (3.3).

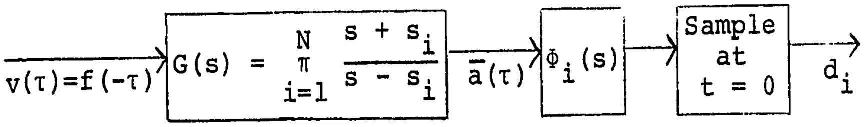


Figure 3.3 Simplified diagram to evaluate $\int_0^\infty f(t)\phi_{N+i}(t)dt$ by filtering operation

The optimality condition is stated as

$$d_i = 0,\ i = 1,2,\ldots N \tag{3.40}$$

The evaluation of the output of the filter with reversed f(t) as input is not difficult. In all practical cases f(t) can be assumed to be zero for $t > T_0$ where T_0 is the duration of the signal f(t).

McDonough defines a function $\sum_{i=1}^{N} d_i^2$ and computes the zeros of this function (26). It is found that the analytical expression for $\sum_{i=1}^{N} d_i^2$ is too difficult to work with even for N = 3. The method is found to lead to computational difficulties for larger N (26).

McDonough and Huggins later modified their method into an iterative scheme (31).

Let

$$\bar{G}(s) = \sum_{i=0}^{N} g_i s^i / D(s),\ g_N = 1 \tag{3.41}$$

where

$$D(s) = \prod_{i=1}^{N} (s - s_i) \tag{3.42}$$

If $\bar{G}(s)$ is used for the filtering operation of Fig. 3.3, instead of G(s), the sample at t = 0 is given by

$$d_k = \int_{-j\infty}^{+j\infty} F(-s) \left\{ \sum_{i=0}^{N} g_i s^i / D(s) \right\} \Phi_k(s) \frac{ds}{2\pi j} \qquad (3.43)$$

$$k = 1,2,\ldots N$$

where F(s) is the Laplace transform of f(t).

If d_k satisfies the optimality condition, equation (3.43) can be written as

$$\sum_{i=0}^{N-1} \left[\int_{-j\infty}^{+j\infty} F(-s) \left\{ \frac{s^i}{D(s)} \Phi_k(s) \right\} \frac{ds}{2\pi j} \right] g_i$$

$$= -\int_{-j\infty}^{+j\infty} F(-s) \left\{ \frac{s^N}{D(s)} \Phi_k(s) \right\} \frac{ds}{2\pi j} \qquad k = 1,2,\ldots N \qquad (3.44)$$

Equation (3.44) gives N linear equations in $(g_i)_{i=0}^{N-1}$. If $(g_i)_{i=0}^{N-1}$, as obtained by solving equation (3.44), satisfy the condition,

$$\sum_{i=0}^{N} g_i s^i = (-1)^N D(-s) \qquad (3.45)$$

we get

$$\bar{G}(s) = G(s) \qquad (3.46)$$

and the corresponding values of $(s_i)_{i=1}^{N}$ give the optimum point.

An initial estimate of $(s_i)_{i=1}^{N}$ is substituted into D(s) and $\Phi_k(s)$, $k = 1,2,\ldots,N$. Equation (3.44) can now be solved for g_i, $i = 0,1,\ldots(N-1)$. When equation (3.45) is satisfied the iteration is stopped. Otherwise these g_i values are used to make a new estimate of $(s_i)_{i=1}^{N}$ and a new set of values of g_i are found and the iteration is continued.

The evaluation of expressions like

$$\int_{-j\infty}^{+j\infty} F(-s)\left\{\frac{s^i\Phi_K(s)}{D(s)}\right\}\frac{ds}{2\pi j}$$

can be done by the filtering scheme similar to that in Fig. 3.2. Some numerical difficulties are reported in the realization of filters $\frac{s^i}{D(s)}$ and $\Phi_k(s)$ in cascade (31). The initial estimate of pole positions has to be near the optimum to achieve convergence. When f(t) is nearly exponential the convergence is fast.

SEAR'S METHOD

An iterative scheme, similar to the one discussed above, was first suggested by Sears (32).

From equation (3.31)

$$I = \int_{-\infty}^{0}|\bar{a}(t)|^2 dt$$

where $\qquad \bar{A}(s) = F(-s)G(s)$

Sears assumes a set of even number of exponentials.

Therefore

$$G(s) = \frac{s^N + \sum_{i=0}^{N-1} (-1)^i a_i s^i}{s^N + \sum_{i=0}^{N-1} a_i s^i} \tag{3.47}$$

The output of the complimentary filter can be expressed as

$$\bar{a}(t) = L^{-1}\left\{F(-s)G(s)\right\} = L^{-1} F(-s)\left\{\frac{s^N + \sum_{i=0}^{N-1} (-1)^i a_i s^i}{s^N + \sum_{i=0}^{N-1} a_i s^i}\right\} \tag{3.48}$$

where L^{-1} stands for the inverse Laplace transform.

For I to be minimum

$$\frac{\partial I}{\partial a_k} = 0,\ k = 0,1,\ldots N-1 \tag{3.49}$$

$$\frac{\partial I}{\partial a_k} = 2\int_{-\infty}^{0} \bar{a}(t)\frac{\partial}{\partial a_k} \bar{a}(t)dt$$

$$= 2\int_{-\infty}^{0} \bar{a}(t)L^{-1}\left\{\frac{\partial}{\partial a_k}\left(F(-s)\frac{s^N + \sum_{i=0}^{N-1} (-1)^i a_i s^i}{s^N + \sum_{i=0}^{N-1} a_i s^i}\right)\right\}dt$$

$$= 0 \tag{3.50}$$

Equation (3.50) may be expressed as

$$\int_{-\infty}^{0} L^{-1}\{\bar{A}(s)\}L^{-1}\left\{F(-s)\frac{s^{N+k}\left[(-1)^{k}-1\right]+\sum\limits_{i=0}^{N-1}\left[(-1)^{k}-(-1)^{i}\right]a_{i}s^{i+k}}{(s^{N}+\sum\limits_{i=0}^{N-1}a_{i}s^{i})^{2}}\right\}dt$$

$$= 0, \quad k = 0,1,\ldots,N-1 \qquad (3.51)$$

Let

$$\frac{s^{k}}{s^{N}+\sum\limits_{i=0}^{N-1}a_{i}s^{i}} = \Psi_{k}(s) \qquad (3.52)$$

and

$$\frac{s^{N+k}\left[(-1)^{k}-1\right]+\sum\limits_{i=0}^{N-1}\left[(-1)^{k}-(-1)^{i}\right]a_{i}s^{i+k}}{(s^{N}+\sum\limits_{i=0}^{N-1}a_{i}s^{i})^{2}} = G_{K}(s) \qquad (3.53)$$

Using equation (3.52) $\bar{A}(s)$ can be expressed as

$$\bar{A}(s) = F(-s)\Psi_{N}(s) + \sum_{i=0}^{N-1}(-1)^{i}F(-s)\Psi_{i}(s)a_{i}$$

Hence (3.51) can now be written as

$$\int_{-\infty}^{0} L^{-1}\left\{F(-s)\Psi_{N}(s) + \sum_{i=0}^{N-1}(-1)^{i}\Psi_{i}(s)F(-s)a_{i}\right\}L^{-1}\left\{G_{k}(s)F(-s)\right\}dt = 0$$

$$k = 0,1,2,\ldots N-1$$

Or

$$\sum_{i=0}^{N-1} (-1)^i a_i \int_{-\infty}^{0} L^{-1}\left\{\Psi_i(s)F(-s)\right\} L^{-1}\left\{G_K(s)F(-s)\right\} dt$$

$$= \int_{-\infty}^{0} L^{-1}\left\{F(-s)\Psi_N(s)\right\} L^{-1}\left\{G_K(s)F(-s)\right\} dt \qquad (3.54)$$

$$k = 0,1,\ldots N-1$$

The equation (3.54) can be expressed in a matrix form as

$$\begin{bmatrix} \\ \\ P \\ \\ \\ \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ .. \\ .. \\ .. \\ a_{N-1} \end{bmatrix} = \begin{bmatrix} \int_{-\infty}^{0} L^{-1}\{F(-s)\Psi_N(s)\}L^{-1}\{G_0(s)F(-s)\}dt \\ \int_{-\infty}^{0} L^{-1}\{F(-s)\Psi_N(s)\}L^{-1}\{G_1(s)F(-s)\}dt \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \int_{-\infty}^{0} L^{-1}\{F(-s)\Psi_N(s)\}L^{-1}\{G_{N-1}(s)F(-s)\}dt \end{bmatrix} \qquad (3.55)$$

where P is a N x N matrix.

The $(\ell, j)^{th}$ element of P matrix is given by

$$p_{\ell,j} = (-1)^{\ell-1} \int_{-\infty}^{0} L^{-1}\{\Psi_\ell(s)F(-s)\}L^{-1}\{G_j(s)F(-s)\}dt \quad (3.56)$$

If $(a_i)_{i=0}^{N-1}$ are known, the P matrix and the vector on right hand side of equation (3.55) can be calculated. If these values of a_i, $i = 0,1,\ldots,N-1$ are at the optimum point, equation (3.55) is satisfied. If not, the solution of this equation gives a new estimate of $(a_i)_{i=0}^{N-1}$. The evaluation of terms like $\int_{-\infty}^{0} L^{-1}\{F(-s)\Psi_i(s)\}L^{-1}\{F(-s)G_K(s)\}dt$ is carried out as follows.

Let

$$L^{-1}\{\Psi_i(s)F(-s)\} = \psi_{if}(t)$$

and

$$L^{-1}\{G_k(s)F(-s)\} = g_{kf}(t) \qquad (3.57)$$

$\psi_{if}(t)$ and $g_{kf}(t)$ can be generated by the filtering operations shown in Fig. 3.4. The evaluation of the expression is done by integrating the product of $\psi_{if}(t)$ and $g_{kf}(t)$.

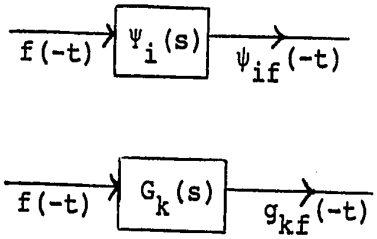


Figure 3.4. Schematic diagram to evaluate $\psi_{if}(t)$ and $g_{kf}(t)$

Convergence properties of this method are discussed in (33). It has been proved that if the first step of interation gives a reduced error the method will converge. If not, the initial estimate is changed and the iteration is repeated. Hence the success of the method depends on making a 'good' initial estimate of the pole positions.

The methods of McDonough, Huggins and Sears are very important. None of the previous techniques reported in the literature sought to find an exact solution of the exponential representation of signals by minimizing the actual ISE. The new methods are computationally more tedious than the previous methods, but this is inherent in the very nature of the problem as indicated by Aigrain and Williams equations.

In all of these new methods the zeros of H(s) are found by first determining the coefficients of the orthonormal expansion, as given by equation (2.18), that is

$$C_i = \int_0^\infty f(t)\phi_i(t)dt$$

The evaluation of C_i is done by a filtering operation similar to the one in Fig. 3.2 (33). The actual scheme for evaluating C_i, i = 1,2,...,N is shown in Fig. 3.5.

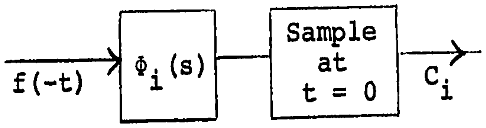


Figure 3.5. Evaluation of orthonormal coefficients

Once C_i values are found,

$$H(s) = \sum_{i=1}^{N} C_i \Phi_i(s) \tag{3.58}$$

Equation (3.58) gives the zeros of H(s).

MSS METHOD

An alternate approach to arrive at an exact solution was suggested by McBridge, Schaefgen and Steiglitz in (34). (For convenience this will be called MSS method). In this method, a rational Laplace transform H(s) is first assumed

$$H(s) = \frac{b_1 s^{N-1} + b_2 s^{N-2} + \ldots + b_N}{s^N + a_1 s^{N-1} + \ldots + a_N}$$

$$= \frac{b_1 s^{-1} + b_2 s^{-2} + \ldots + b_N s^{-N}}{1 + a_1 s^{-1} + \ldots + a_N s^{-N}}$$

$$= \frac{N(s)}{D(s)} \tag{3.59}$$

The ISE is given by

$$I = \int_0^\infty |f(t) - h(t)|^2 dt \tag{3.60}$$

where h(t) is the impulse response of H(s).

Considering the error e(t)

$$e(t) = f(t) - h(t)$$

$$E(s) = F(s) - H(s)$$

$$= F(s) - \frac{b_1 s^{-1} + b_2 s^{-2} + \ldots + b_N s^{-N}}{1 + a_1 s^{-1} + a_2 s^{-2} + \ldots + a_N s^{-N}} \quad (3.61)$$

$$\frac{\partial E(s)}{\partial b_i} = - \frac{s^{-i}}{D(s)} = -P_{bi}(s)$$

$$(3.62)$$

$$\frac{\partial E(s)}{\partial a_i} = \frac{s^{-i}N(s)}{D(s)} = P_{ai}(s)$$

From equation (3.62) we get,

$$p_{bi}(t) = - \frac{\partial e(t)}{\partial b_i}$$

$$(3.63)$$

and

$$p_{ai}(t) = \frac{\partial e(t)}{\partial a_i}$$

Since $I = \int_0^\infty e^2(t)dt$, I is minimum when,

$$\frac{\partial I}{\partial b_i} = 2\int_0^\infty e(t)\frac{\partial e(t)}{\partial b_i}dt$$

$$= -2\int_0^\infty e(t)p_{bi}(t)dt$$

$$= 0$$

and
$$\frac{\partial I}{\partial a_i} = 2\int_0^\infty e(t)\frac{\partial e(t)}{\partial a_i}dt$$

$$= 2\int_0^\infty e(t)p_{ai}(t)dt$$

$$= 0 \qquad (3.64)$$

$$i = 1,2,\ldots N$$

An analytical solution of equation (3.64) is difficult because of the nonlinearity. An iterative solution was suggested in MSS method by defining a new error $e_1(t)$ such that $e_1(t)$ tends to $e(t)$ as the final solution is approached. The new error $e_1(t)$ is defined as

$$E_1(s) = \frac{D_j(s)}{D_{j-1}(s)}F(s) - \frac{N_j(s)}{D_{j-1}(s)} \qquad (3.65)$$

where j is the number of iteration. When final solution is reached $D_j(s) = D_{j-1}(s)$ and $E_1(s) = E(s)$. A new ISE is defined as

$$I_1 = \int_0^\infty e_1^2(t)dt \qquad (3.66)$$

$$\frac{\partial E_1(s)}{\partial b_i} = -\frac{s^{-i}}{D_{j-1}(s)} = -P_{bi}(s)$$

$$(3.67)$$

$$\frac{\partial E_1(s)}{\partial a_i} = \frac{s^{-i}}{D_{j-1}(s)}F(s) = P_{ai}(s)$$

In deriving equation (3.67) it is assumed that the parameters of $D_{j-1}(s)$ are not varied in the j^{th} iteration. $E_1(s)$ can now be written as,

$$E_1(s) = \frac{1 + a_1 s^{-1} + \ldots + a_N s^{-N}}{D_{j-1}(s)} F(s) - \frac{b_1 s^{-1} + b_2 s^{-2} + \ldots + b_N s^{-N}}{D_{j-1}(s)}$$

$$= P_{a0}(s) + a_1 P_{a1}(s) + \ldots + a_N P_{aN}(s)$$

$$- b_1 P_{b1}(s) - b_2 P_{b2}(s) \ldots - b_N P_{bN}(s) \qquad (3.68)$$

The following 2N dimensional vectors λ and $\bar{p}(t)$ are defined

$$\lambda = [a_1 \;\; a_2 \ldots a_N \;\; -b_1 \;\; -b_2 \ldots -b_N]^t$$

and $\bar{p}(t) = [p_{a1}(t) \;\; p_{a2}(t) \ldots p_{aN}(t) \;\; p_{b1}(t) \;\; p_{b2}(t)$

$$\ldots . p_{bN}(t)]^t \qquad (3.69)$$

where λ^t is the transpose of λ.

$e_1(t)$ can now be expressed as a matrix equation

$$e_1(t) = \lambda^t \, \bar{p}(t) + p_{a0}(t) \qquad (3.70)$$

$$I_1 = \int_0^\infty |\lambda^t \, \bar{p}(t) + p_{a0}(t)|^2 dt \qquad (3.71)$$

From equations (3.70) and (3.71) we obtain

$$\text{Grad } I_1 = 2[\int_0^\infty \bar{p}(t)\bar{p}^t(t)dt]\lambda + 2\int_0^\infty p_{a0}(t)\bar{p}(t)dt$$

$$= 2P \;\; + 2C \qquad (3.72)$$

where

$$P = \int_0^\infty \bar{p}(t)\bar{p}^t(t)dt \text{ is a } 2N \times 2N \text{ matrix}$$

and $$C = \int_0^\infty p_{a0}(t)\bar{p}(t)dt \text{ is a } 2N \text{ vector.} \qquad (3.73)$$

Equating Grad $I_1 = 0$ in equation (3.72) an iteration equation is obtained

$$\lambda = P^{-1}C \qquad (3.74)$$

Examining equations (3.62), (3.64), (3.66) and (3.67) it is found that Grad $I_1 \neq$ Grad I. Hence this method cannot converge to the true minimum of I. This difficulty is overcome by making the following changes.

Let $$\tilde{p}_{ai}(s) = \frac{s^{-i}N_{j-1}(s)}{D_{j-1}^2(s)} \qquad (3.75a)$$

Then $$\frac{\partial I}{\partial a_i} = 2\int_0^\infty e_1(t)\tilde{p}_{ai}(t)dt$$

$$\frac{\partial I}{\partial b_i} = -2\int_0^\infty e_1(t)p_{bi}(t)dt \qquad (3.75b)$$

The vector $\bar{p}(t)$ is changed by replacing $p_{ai}(t)$ by $\tilde{p}_{ai}(t)$.

$p_{a0}(t)$ is also replaced by $\tilde{p}_{a0}(t)$. Corresponding changes are made in the matrix P and vector C.

The second method of MSS cannot be started with an initial estimate of zero or unity for H(s). In general it is found that method 1 is well suited for initial points far from optimum and method 2 converges faster near the optimum. It should be pointed out that in MSS method the iteration is to be carried out for 2N variables while in previous methods only N variables needed to be considered.

MILLER'S METHOD (35-36)

Miller modifies the MSS method by introducing Aigrain and Williams equations to the error $E_1(s)$. Using equations (2.7), (2.8) and (2.9) we obtain a different form for Aigrain and Williams equations. These are

$$E(-s_i) = F(-s_i) - H(-s_i) = 0$$
$$E'(-s_i) = F'(-s_i)-H'(-s_i) = 0 \qquad (3.76)$$

where $E'(s) = \frac{d}{ds}E(s)$, $i = 1,2,\ldots N$.

Assuming that $E_1(s)$, as defined by equation (3.65) should satisfy equation (3.76) at the optimum point, we get

$$E_1(-s_i) = 0 \qquad (3.77a)$$

$$E_1'(-s_i) = 0 \qquad (3.77b)$$

where $(s_i)_{i=1}^{N}$ give the exponentials for minimum ISE. Substituting $E_1(s)$ and $E_1{}'(s)$ from equation (3.65) into equation (3.77)

$$D_j(-s_i)F(-s_i) - N_j(-s_i) = 0$$

$$F'(-s_i)D_j(-s_i) + F(-s_i)D_j'(-s_i) = N_j'(-s_i) \qquad (3.78)$$

$$i = 1,2,\ldots N$$

$D_{j-1}(s)$ used in equation (3.65) does not appear in equation (3.78). Miller has shown (35) that equation (3.78) can be directly derived from Aigrain and Williams equations as given by equation (3.76). If $(s_i)_{i=1}^{N}$ are considered as an estimate of the pole positions, equations (3.78) gives 2N simultaneous linear equations in 2N unknowns $(a_i, b_i)_{i=1}^{N}$. The solution of equation (3.78) thus gives a new estimate of the pole positions. Iteration is continued until there is no further change in pole positions.

Miller's approach could be used only when F(s) and F'(s) are known. When f(t) is analytically known this method is useful. Miller establishes the link between MSS method and Aigrain and Williams equations. Equations (3.78) gives an iterative method of solving the Aigrain and Williams equations. Both MSS and Miller methods converge slowly (36).

APPROXIMATION TO IDEAL LOW PASS FILTER USING TIME DOMAIN APPROACH

Two different methods of approximating the ideal low-pass filter, by time domain techniques have been reported in the literature (1-3). In (3) Ulstad approximates the delayed and truncated $\frac{\sin t}{t}$ function by the sum of damped sine and cosine functions. The techniques of nonlinear programming are used to find the sine and cosine functions and their weights such that ISE is minimized. These values are then used as the starting point for minimizing the Chebyshev error in time domain.

In a recent paper, Pottle and Wong make use of time domain techniques to achieve optimum least-squares approximations to ideal low-pass filter (2). The approximation is by means of orthonormal functions $\phi_k(t)$, $k = 1, 2, \ldots N$ discussed in Chapter 2. The ISE in the frequency domain and time domain are given as follows

$$I = \frac{1}{2\pi}\int_{-\infty}^{+\infty}|F(j\omega) - H(j\omega)|^2 d\omega$$

$$= \int_{-\infty}^{+\infty}|f(t) - h(t)|^2 dt \qquad (3.79)$$

$F(j\omega)$ is the desired frequency response and $H(j\omega)$ is the approximation. The functions f(t) and h(t) are the inverse of F(jω) and H(jω) respectively.

$$h(t) = \sum_{i=1}^{N} C_i \phi_i(t)$$

and

$$H(j\omega) = \sum_{i=1}^{N} C_i \Phi_i(j\omega) \qquad (3.80)$$

$$C_i = \int_0^\infty f(t)\phi_i(t)dt$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(-j\omega)\Phi_i(j\omega)d\omega \qquad (3.81)$$

The expression for I can be written as

$$I = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |F(j\omega)|^2 d\omega - \sum_{i=1}^{N} C_i^2 \qquad (3.82)$$

The frequency response considered by Pottle and Wong in (2) is

$$F(j\omega) = e^{-j\omega t_0}, \quad |\omega| \leqslant 1$$

$$= 0, \quad |\omega| > 1 \qquad (3.83)$$

The inverse f(t) of F(jω) is

$$f(t) = \frac{1}{\pi}\frac{\sin(t - t_0)}{(t - t_0)} \qquad (3.84)$$

Equations (3.81) and (3.82) can be written as

$$C_i = \frac{1}{2\pi}\int_{-1}^{+1} e^{j\omega t_0}\Phi_i(j\omega)d\omega \qquad (3.85)$$

and

$$I = \frac{1}{\pi} - \sum_{i=1}^{N} C_i^2 \qquad (3.86)$$

The problem of choosing the parameters of H(s) by minimizing I in equation (3.86) was first discussed by Pottle and Thorp in (1) and recently by Pottle and Wong in (2). Pottle and Thorp make use of the steepest descent technique and Newton's method of minimization. The optimization is done with respect to the parameters p_i and q_i of functions $\Phi_i(j\omega)$, i = 1,2,...N. The delay t_0 in equation (3.83) may be kept constant or may be considered as another parameter.

The results obtained by Pottle and Thorp by applying this minimization technique are not satisfactory. It is found that in most cases reported in (1) the optimum point has multiple poles. This, obviously, is a contradiction because the orthonormal functions $\phi_i(t)$, i = 1,2,...N are only linear combinations of $(e^{s_i t})_{i=1}^{N}$ and hence $\Phi_i(s)$ i = 1,2,...,N cannot have multiple poles.

An improved version of this method appears in the recent paper by Pottle and Wong (2). The general approach in (2) is the same as in (1). The minimization is carried out by using Fletcher-Powell algorithm. The optimum points, as reported in (2), do not have multiple poles.

The following points may be noted in connection with the methods in (1) and (2).

1. The function f(t) in equation (3.79) exists for $-\infty < t < \infty$. The orthonormal functions $\phi_i(t)$, $i = 1,2,\ldots,N$ and h(t) exist only for $t \geqslant 0$. Hence the condition of completeness, required for any orthonormal expansion, is not satisfied (14).

2. The ISE evaluated by equation (3.82) corresponds to the total error for $-\infty < t < \infty$.

3. The coefficients $(C_i)_{i=1}^{N}$ correspond only to that part of f(t) for which $t \geqslant 0$.

4. Whenever a pair of poles occur near the imaginary axis, $\Phi_i(j\omega)$ will have peaks at the corresponding values of ω. This leads to serious difficulties in the numerical evaluation of $(C_i)_{i=1}^{N}$ and the gradients (1,2).

The paper by Pottle and Thorp (1) is the first attempt, reported in the literature, to approximate a desired frequency response in the least integral squared error sense. This method and its later modification by Pottle and Wong (2) have the advantage that when the signal is band limited, all integrations over the semi infinite time axis could be converted to finite integrations in the frequency domain. But the numerical difficulties reported in both (1) and (2) limit the applications of the method.

All the methods reviewed in this chapter approximate a given impulse response by the inverse of a rational Laplace transform H(s). In most cases the poles and zeros of H(s) are determined in the least square sense. In some cases, the poles are located first and the zeros are then chosen to satisfy certain conditions in the time domain. The determination of optimum poles always involve tedious computations which sometimes lead to numerical difficulties. The problem of finding H(s) so as to satisfy specific conditions in the frequency domain, in addition to the conditions on ISE, has not yet been reported in the literature.

CHAPTER 4

IMPROVING THE FREQUENCY RESPONSE BY MINIMIZING AVERAGE ERROR IN TIME DOMAIN

It has already been shown that minimizing the ISE in the time domain is equivalent to minimizing the ISE in the frequency domain. From the point of view of the frequency domain performance the minimization of ISE alone need not produce the best result. For example, at some value ω, the magnitude response of $H(j\omega)$ may have a sharp deviation from the desired response even if the ISE is a minimum (2). This chapter develops a new method of minimizing this deviation, keeping the ISE within allowable limits.

PROBLEM STATEMENT

Let $F(j\omega)$ be the ideal frequency response to be approximated and $H(j\omega)$ be the approximation. The deviation $E(j\omega)$ for $-\infty < \omega < \infty$ is defined as

$$E(j\omega) = F(j\omega) - H(j\omega) \tag{4.1}$$

The problem is to find the parameters of $H(j\omega)$ such that the upper bound of $|E(j\omega)|$, $\omega \varepsilon (-\infty, \infty)$ is minimized subject to the condition that the ISE is less than a preassigned value. The problem as stated above considers both the ISE and the deviations in the frequency response. The solution of this problem, thus offers a compromise between the least square and Chebyshev criteria in the frequency domain.

AN UPPER BOUND ON DEVIATION

The error $E(j\omega)$ is defined as

$$E(j\omega) = F(j\omega) - H(j\omega)$$

where $$e(t) = f(t) - h(t) \qquad (4.2)$$

$E(j\omega)$ can be expressed as

$$E(j\omega) = \int_{-\infty}^{+\infty} e(t)e^{-j\omega t}dt$$

$$= \int_{-\infty}^{+\infty} \left[e(t)e^{-j\omega t}\right]^{\frac{1}{2}} \left[e(t)e^{-j\omega t}\right]^{\frac{1}{2}} dt$$

Since $e(t)\varepsilon L^2$, we can apply Schwarz's inequality giving

$$|E(j\omega)| \leqslant \left[\int_{-\infty}^{+\infty} |\{e(t)e^{-j\omega t}\}^{\frac{1}{2}}|^2 dt\right]^{\frac{1}{2}} \left[\int_{-\infty}^{+\infty} |\{e(t)e^{-j\omega t}\}^{\frac{1}{2}}|^2 dt\right]^{\frac{1}{2}}$$

$$= \int_{-\infty}^{+\infty} |\{e(t)e^{-j\omega t}\}^{\frac{1}{2}}|^2 dt$$

$$= \int_{-\infty}^{+\infty} |e(t)| dt$$

Thus we have

$$|E(j\omega)| \leqslant \int_{-\infty}^{+\infty}|e(t)|dt \tag{4.3}$$

The function h(t) of equation (4.2) is formed from a set of N exponentials as discussed in Chapter 2. Hence

$$h(t) = \sum_{i=1}^{N} C_i\phi_i(t) \tag{4.4}$$

where $\phi_i(t)$, i = 1,2,...,N are the N orthonormal functions. Using equation (4.4) equation (4.2) can be expressed as

$$e(t) = f(t) - \sum_{i=1}^{N} C_i\phi_i(t) \tag{4.5}$$

Therefore

$$\begin{aligned}\int_{-\infty}^{+\infty}|e(t)|dt &= \int_{-\infty}^{+\infty}|f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|dt \\ &\leqslant \int_{-\infty}^{+\infty}|f(t)|dt + \int_{-\infty}^{+\infty}|\sum_{i=1}^{N} C_i\phi_i(t)|dt \\ &\leqslant \int_{-\infty}^{+\infty}|f(t)|dt + \sum_{i=1}^{N}|C_i|\int_{-\infty}^{+\infty}|\phi_i(t)|dt \\ &= \int_{-\infty}^{+\infty}|f(t)|dt + \sum_{i=1}^{N}|C_i|\int_{0}^{\infty}|\phi_i(t)|dt \qquad (4.6)\end{aligned}$$

Each of the functions $\phi_i(t)$, i = 1,2,...N is a sum of damped exponentials and exists only for $t \geqslant 0$. From Chapter 2, we have

$$\phi_k(t) = \sum_{i=1}^{n} R_i e^{s_i t}$$

$$= \sum_{i=1}^{n} R_i e^{\{Re(s_i) + jIm(s_i)\}t} \qquad (4.7)$$

where R_i is a complex constant and $1 \leqslant k \leqslant n \leqslant N$.

$$\int_0^\infty |\phi_k(t)|dt = \int_0^\infty |\sum_{i=1}^{n} R_i e^{\{Re(s_i) + jIm(s_i)\}t}|dt$$

$$\leqslant \sum_{i=1}^{n} |R_i| \int_0^\infty e^{Re(s_i)t} dt$$

$$= \sum_{i=1}^{n} |R_i| \frac{1}{-Re(s_i)} \qquad (4.8)$$

By using equation (4.8) we get

$$\sum_{i=1}^{N} |C_i| \int_0^\infty |\phi_i(t)|dt < \infty \qquad (4.9)$$

(assuming each C_i to be finite)

Using equations (4.6) and (4.9) we have the result that $\int_{-\infty}^{+\infty}|e(t)|dt$ is finite, when $\int_{-\infty}^{+\infty}|f(t)|dt$ is finite. This leads to the conclusion that the deviation $E(j\omega)$ has a finite upper bound when $\int_{-\infty}^{+\infty}|f(t)|dt$ exists. Moreover if e(t) is unidirectional $\int_{-\infty}^{+\infty}|e(t)|dt$ is the same as the magnitude of the deviation at $\omega = 0$. Hence, in the most general case $\int_{-\infty}^{+\infty}|e(t)|dt$ is the least upper bound of $|E(j\omega)|$. This expression, $\int_{-\infty}^{+\infty}|e(t)|dt$, is defined as the average error of approximation. Thus the minimization of average error offers a method of minimizing the deviation in the frequency domain. This is now considered in greater detail.

Since $\phi_i(t) = 0$ for $t < 0$, equation (4.6) can be modified as follows.

$$\int_{-\infty}^{+\infty} |e(t)|dt = \int_{-\infty}^{+\infty} |f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|dt$$

$$= \int_{-\infty}^{0} |f(t)|dt + \int_0^{\infty} |f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|dt \quad (4.10)$$

Examining equation (4.10) we see that the choice of $C_i\phi_i(t)$, $i = 1,2,...N$ has no influence over the term $\int_{-\infty}^{0} |f(t)|dt$. For a causal system $f(t) = 0$ for $t < 0$ and this term vanishes. Hence in order to minimize the average error we need only consider the expression $\int_0^{\infty} |f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|dt$. Let γ represent this average error

$$\gamma = \int_0^{\infty} |f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|dt \quad (4.11)$$

The problem of minimizing the upper-bound of deviation in frequency domain subject to a constraint on ISE can now be restated as follows.

Choose the parameters of $h(t) = \sum_{i=1}^{N} C_i\phi_i(t)$ such that the average error $\gamma = \int_0^{\infty} |f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|dt$ is minimized subject to the condition that the integral squared error $I = \int_0^\infty |f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|^2 dt \leqslant \mu$, where μ is the allowable ISE.

MINIMIZATION OF AVERAGE ERROR

The determination of the parameters of $H(s) = \sum_{i=1}^{N} C_i\Phi_i(s)$ is carried out in two steps. The first step is to find a set of poles of H(s) which will minimize the ISE. These poles are the solution of the Aigrain and Williams equations (2.14-2.15). Even though the uniqueness of the solutions of equations (2.14-2.15) has not yet been proved mathematically, all the researches reviewed in Chapter 3 have reported that in all of the cases considered the various iterative schemes have converged to the same point. Hence it may be assumed that the solution of the Aigrain and Williams equations or equivalent equations gives a set of poles yielding the least ISE. Once the poles $(s_i)_{i=1}^{N}$ are known a set of orthonormal functions $\phi_i(t)$, i = 1,2,...N can be formed. The second step is the determination of the coefficients $(C_i)_{i=1}^{N}$ such that γ is minimized, subject to the constraint on ISE.

Let $(s_i)_{i=1}^{N}$ be the set of poles giving the least ISE and $\phi_i(t)$, i = 1,2,...N be the corresponding orthonormal functions. Let $\tilde{I}$ be the least ISE and $(\tilde{C}_i)_{i=1}^{N}$ be the coefficients of orthonormal expansion at the least ISE.

$$\tilde{C}_i = \int_0^\infty f(t)\phi_i(t)dt \qquad (4.12)$$

The expression for the least ISE is

$$\tilde{I} = \int_0^\infty |f(t)|^2 dt - \sum_{i=1}^{N} \tilde{C}_i^2 \qquad (4.13)$$

Let I be the ISE at some other set of N coefficients $(C_i)_{i=1}^{N}$. Then,

$$\begin{aligned} I &= \int_0^\infty |f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|^2 dt \\ &= \int_0^\infty |f(t)|^2 dt - 2\int_0^\infty f(t)\left\{\sum_{i=1}^{N} C_i\phi_i(t)\right\}dt + \int_0^\infty \left\{\sum_{i=1}^{N} C_i\phi_i(t)\right\}^2 dt \end{aligned} \qquad (4.14)$$

Using equation (4.12)

$$\begin{aligned} \int_0^\infty f(t)\left\{\sum_{i=1}^{N} C_i\phi_i(t)\right\}dt &= \sum_{i=1}^{N} C_i\int_0^\infty f(t)\phi_i(t)dt \\ &= \sum_{i=1}^{N} C_i\tilde{C}_i \end{aligned} \qquad (4.15)$$

$$\begin{aligned} \int_0^\infty \left\{\sum_{i=1}^{N} C_i\phi_i(t)\right\}^2 dt &= \int_0^\infty \left\{\sum_{i=1}^{N} C_i^2\phi_i^2(t)\right\} dt \\ &+ \int_0^\infty \left\{\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j\neq 1}}^{N} C_iC_j\phi_i(t)\phi_j(t)\right\}dt \end{aligned} \qquad (4.16)$$

Since the functions $\phi_i(t)$, i = 1,2,...N are orthonormal

$$\int_0^\infty \phi_i^2(t)dt = 1,$$

and $$\int_0^\infty \phi_i(t)\phi_j(t) = 0, \; i \neq j \qquad (4.17)$$

$$i = 1,2,...N$$

$$j = 1,2,...N$$

Substituting equation (4.17) into (4.16)

$$\int_0^\infty \left\{ \sum_{i=1}^{N} C_i\phi_i(t) \right\}^2 dt = \sum_{i=1}^{N} C_i^2 \qquad (4.18)$$

Equation (4.14) can now be written as

$$I = \int_0^\infty |f(t)|^2 dt - 2\sum_{i=1}^{N} C_i\tilde{C}_i + \sum_{i=1}^{N} C_i^2 \qquad (4.19)$$

From equation (4.13)

$$\int_0^\infty |f(t)|^2 dt = \tilde{I} + \sum_{i=1}^{N} \tilde{C}_i^2$$

Thus the expression for I is obtained as

$$I = \tilde{I} + \sum_{i=1}^{N} \tilde{C}_i^2 - 2\sum_{i=1}^{N} C_i\tilde{C}_i + \sum_{i=1}^{N} C_i^2 \qquad (4.20)$$

If μ is the allowable ISE, the expression for the coefficients $(C_i)_{i=1}^{N}$ which will yield an ISE $\leqslant \mu$ can be obtained from equation (4.20). That is,

$$\tilde{I} + \sum_{i=1}^{N} (\tilde{C}_i^2 - 2C_i\tilde{C}_i + C_i^2) = I \leqslant \mu \qquad (4.21)$$

Equation (4.21) can be written as

$$\sum_{i=1}^{N} (C_i - \tilde{C}_i)^2 \leqslant (\mu - \tilde{I}) = \delta^2 \qquad (4.22)$$

Obviously, μ can only be chosen such that $\mu \geqslant \tilde{I}$.

We now prove the following theorem on the existence of a minimum of the average error with a constraint on ISE.

THEOREM

If a function f(t), $f(t) \varepsilon L$, is approximated by a function $h(t) = \sum_{i=1}^{N} C_i\phi_i(t)$, where $\phi_i(t)$, $i = 1,2,\ldots,N$ are N orthonormal functions constructed from a set of N distinct exponentials giving the least ISE $\tilde{I}$, then for every ISE $\mu \geqslant \tilde{I}$, there exists at least one set of coefficients $(C_i)_{i=1}^{N}$ such that the average error γ at this set of coefficients is a minimum.

PROOF

For every ISE $\mu \geqslant \tilde{I}$, the coefficients $(C_i)_{i=1}^{N}$ are given by equation (4.22). According to this equation the point $(C_i)_{i=1}^{N}$ lies inside or on the sphere of radius δ, in the N dimensional Euclidean space. This sphere, defined by equation (4.22), forms a compact set in a metric space. We will prove that the average error γ, considered as a function of $(C_i)_{i=1}^{N}$, is continuous on this compact set. Since any real valued continuous function defined on a compact set in a metric space has its infimum and supremum on that set (37, 38) γ has at least one minimum inside or on the sphere defined by equation (4.22).

By definition,

$$\gamma = \int_0^\infty |e(t)|dt$$

$$= \int_0^\infty |f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|dt$$

Here γ may be considered as the L_1 norm of e(t). Let us define the norm of e(t) as

$$||e(t)|| = \int_0^\infty |e(t)|dt \qquad (4.23)$$

γ can be considered as a function of a vector $\overline{C}$ where

$$\overline{C} = [C_1 \quad C_2 \quad C_3 \quad \ldots\ldots \quad C_N]^t \qquad (4.24)$$

Hence

$$\gamma(\overline{C}) = ||e(t)||$$

$$= ||f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|| \qquad (4.25)$$

Let $\overline{C}_1$ and $\overline{C}_2$ be two points satisfying equations (4.22). Hence ISE at $\overline{C}_1$ or $\overline{C}_2$ is less than or equal to μ.

Let $$\overline{C}_1 = [C_{11} \;\; C_{21} \;\; C_{31} \;\; \ldots . \;\; C_{N1}]^t$$

and $$\overline{C}_2 = [C_{12} \;\; C_{22} \;\; C_{32} \;\; \ldots . \;\; C_{N2}]^t \tag{4.26}$$

Then

$$\gamma(\overline{C}_1) = ||f(t) - \sum_{i=1}^{N} C_{i1}\phi_i(t)||$$

and $$\gamma(\overline{C}_2) = ||f(t) - \sum_{i=1}^{N} C_{i2}\phi_i(t)|| \tag{4.27}$$

$$\left|\gamma(\overline{C}_1) - \gamma(\overline{C}_2)\right| = \left| \, ||f(t) - \sum_{i=1}^{N} C_{i1}\phi_i(t)|| - ||f(t) - \sum_{i=1}^{N} C_{i2}\phi_i(t)|| \, \right| \tag{4.28}$$

Let $$e_1(t) = f(t) - \sum_{i=1}^{N} C_{i1}\phi_i(t)$$

and $$e_2(t) = f(t) - \sum_{i=1}^{N} C_{i2}\phi_i(t) \tag{4.29}$$

From the properties of norms,

$$||e_1(t)|| = ||e_1(t) - e_2(t) + e_2(t)||$$

$$\leqslant ||e_1(t) - e_2(t)|| + ||e_2(t)|| \tag{4.30}$$

Similarly

$$||e_2(t)|| \leqslant ||e_2(t) - e_1(t)|| + ||e_1(t)|| \tag{4.31}$$

From equations (4.30) and (4.31) we get

$$||e_1(t)|| - ||e_2(t)|| \leqslant ||e_1(t) - e_2(t)||$$

and $$||e_2(t)|| - ||e_1(t)|| \leqslant ||e_2(t) - e_1(t)|| \qquad (4.32)$$

Since $$||e(t)|| = ||-e(t)||$$

$$\Big|\,||e_1(t)|| - ||e_2(t)||\,\Big| \leqslant ||e_1(t) - e_2(t)|| \qquad (4.33)$$

Using equation (4.33), equation (4.28) can be written as

$$\begin{aligned}\Big|\gamma(\overline{C}_1) - \gamma(\overline{C}_2)\Big| &= \Big|\,||e_1(t)|| - ||e_2(t)||\,\Big| \\ &\leqslant ||e_1(t) - e_2(t)|| \\ &= ||\sum_{i=1}^{N}(C_{i1}-C_{i2})\phi_i(t)|| \\ &\leqslant \sum_{i=1}^{N}||(C_{i1}-C_{i2})\phi_i(t)|| \\ &= \sum_{i=1}^{N}\Big|C_{i1}-C_{i2}\Big|\,||\phi_i(t)|| \\ &\leqslant \underset{i=1,N}{\text{Max}}|C_{i1}-C_{i2}|\sum_{i=1}^{N}||\phi_i(t)||\end{aligned} \qquad (4.34)$$

By definition

$$||\phi_i(t)|| = \int_0^{\infty}|\phi_i(t)|dt \qquad (4.35)$$

By equation (4.8)

$$\int_0^\infty |\phi_i(t)|dt \leqslant \sum_{k=1}^{n} |R_k| \frac{1}{-\mathrm{Re}(s_k)} \qquad (4.36)$$

$$1 \leqslant i \leqslant n \leqslant N$$

where R_k is a complex constant and s_k is the k^{th} exponential. Hence $||\phi_i(t)||$ is finite for $i = 1,2,\ldots N$.

$$\text{Let} \qquad \sum_{i=1}^{N} ||\phi_i(t)|| < K \qquad (4.37)$$

where K is a positive constant.

Equation (4.34) can now be written as

$$|\gamma(\overline{C}_1) - \gamma(\overline{C}_2)| < K \max_{i=1,N} |c_{i1} - c_{i2}| \qquad (4.38)$$

Equation (4.38) may be interpreted as follows.

Given a $\xi > 0$, there exists $\Delta > 0$ such that

$$|\gamma(\overline{C}_1) - \gamma(\overline{C}_2)| < \xi$$

and

$$\max_{i=1,N} |c_{i1} - c_{i2}| < \Delta \qquad (4.39)$$

where

$$\frac{\xi}{K} < \Delta$$

Hence $\gamma(\overline{C})$ is a continuous function of $\overline{C}$ (39). By assumption the points $\overline{C}_1$ and $\overline{C}_2$ are in the compact set defined by the sphere of equation (4.22). By equation (4.39), $\gamma(\bar{C})$ is a continuous function on the compact subset defined by the above sphere. Hence γ has an infimum value on this compact subset. This proves that in every sphere defined by equation (4.22) γ has at least one minimum value.

UNIQUENESS OF THE APPROXIMATION

We can make use of the general theory of approximation in the mean (37, 38) to discuss the quality of the above approximation. By Jackson's theorem (38) the expression

$$\gamma = \int_0^\infty \left| f(t) - \sum_{i=1}^{N} C_i\phi_i(t) \right| dt$$

possesses a unique best approximation, if $\phi_i(t)$, $i = 1,2, \ldots N$ satisfy the Harr Condition. The Harr condition is stated below.

Let $\phi_i(t)$, $i = 1,2,\ldots,N$ be a set of functions in $C[a, b]$. These functions satisfy Harr condition if

$$D = \begin{vmatrix} \phi_1(t_1) & \phi_2(t_1) & \cdots & \phi_N(t_1) \\ \phi_1(t_2) & \phi_2(t_2) & \cdots & \phi_N(t_2) \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \phi_1(t_N) & \phi_2(t_N) & & \phi_N(t_N) \end{vmatrix} \neq 0 \qquad (4.40)$$

for all t_i, $i = 1,2,\ldots N$ such that

$$a \leqslant t_1 \leqslant t_2 \leqslant \ldots \leqslant t_N \leqslant b.$$

In our case $a = 0$ and b tends to ∞. The functions $\phi_i(t)$, $i = 1,2,\ldots N$ are, in general, oscillatory and cross the t-axis many times. The determinant D is zero if any column of the determinant becomes zero. This happens when any function $\phi_i(t)$ has N zero crossings. Hence, in general, the Harr condition is not satisfied.

However, we always approximate the semi infinite interval $[0, \infty]$ by a finite interval $[0, T_0]$. Hence it is sufficient if the Harr condition is satisfied in this interval. The functions $\phi_i(t)$, $i = 1,2,\ldots,N$ are linearly independent by definition. Hence if none of the functions $\phi_i(t)$, $i = 1,2,\ldots,N$ has more than $(N-1)$ zero crossings in the interval $[0, T_0]$ there exists a unique best approximation. Since γ is the L_1 norm, and L_1 norm forms a convex set, this unique best approximation is the same as a local minimum of γ (37 - 38). But this best solution of γ with respect to the coefficients $(C_i)_{i=1}^{N}$ need not lie inside or on the sphere of equation (4.22). Since γ, the L_1 norm of $e(t)$, considered as a function of $(C_i)_{i=1}^{N}$ is a continuous convex function, γ should have a best approximation in every sphere of equation (4.22) even if the unconstrained unique minimum lies outside this sphere.

THE GRADIENT OF AVERAGE ERROR

In the previous sections the existence and uniqueness properties of the minimum of γ with a constraint on ISE are discussed. In order to compute the minimum point we can make use of the fact that at a local minimum

$$\text{Grad } \gamma = [0] \tag{4.41}$$

where

$$\text{Grad } \gamma = \left[\frac{\partial\gamma}{\partial C_1} \quad \frac{\partial\gamma}{\partial C_2} \quad \cdots \quad \frac{\partial\gamma}{\partial C_N}\right]^t \tag{4.42}$$

Since $\frac{\partial}{\partial C_k}|e(t)|$ does not exist when $e(t) = 0$ it is necessary to consider the computation of Grad γ in some detail.

$$\gamma = \int_0^\infty |f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|dt$$

$$\frac{\partial\gamma}{\partial C_k} = \frac{\partial}{\partial C_k}\int_0^\infty |f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|dt \tag{4.43}$$

If at any point $(C_k)_{k=1}^N$, $\frac{\partial}{\partial C_k}|f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|$ is continuous with respect to C_k and t, $t\varepsilon[0, \infty]$, $\frac{\partial\gamma}{\partial C_k}$ exists at this point and hence the differentiation and integration in equation (4.43) can be interchanged (39).

Assuming $|f(t) - \sum_{i=1}^{N} C_i\phi_i(t)| \neq 0$

$$\frac{\partial}{\partial C_k}|f(t) - \sum_{i=1}^{N} C_i\phi_i(t)| = \frac{\partial}{\partial C_k}\{\text{Sgn } e(t)\}\left\{f(t) - \sum_{i=1}^{N} C_i\phi_i(t)\right\} \tag{4.44}$$

where $\text{Sgn } e(t) = \text{Sgn}\left\{f(t) - \sum_{i=1}^{N} C_i\phi_i(t)\right\} = \pm 1,\ e(t) \gtrless 0$

If $e(t) > 0$

$$\frac{\partial}{\partial C_k}|f(t) - \sum_{i=1}^{N} C_i\phi_i(t)| = -\phi_k(t) \tag{4.45}$$

If $e(t) < 0$

$$\frac{\partial}{\partial C_k}|f(t) - \sum_{i=1}^{N} C_i\phi_i(t)| = \phi_k(t) \tag{4.46}$$

From equations (4.45) and (4.46)

$$\frac{\partial}{\partial C_k}|f(t) - \sum_{i=1}^{N} C_i\phi_i(t)| = -\{\text{Sgn } e(t)\}\phi_k(t) \tag{4.47}$$

We are approximating the semi infinite interval $[0, \infty]$ by a finite interval $[0, T_0]$. For some $(C_i)_{i=1}^{N}$, let $e(t) = 0$, at $t_1, t_2, \ldots t_x$ such that

$$0 \leqslant t_1 < t_2 < t_3 < \ldots < t_x \leqslant T_0$$

γ can be written as

$$\gamma = \int_0^{t_1 - \xi} |e(t)|dt + \int_{t_1+\xi}^{t_2-\xi} |e(t)|dt + \ldots + \int_{t_x+\xi}^{T_0} |e(t)|dt \tag{4.48}$$

where $\xi \simeq 0$.

Therefore

$$\frac{\partial\gamma}{\partial C_k} = \frac{\partial}{\partial C_k}\int_0^{t_1 - \xi} |e(t)|dt + \frac{\partial}{\partial C_k}\int_{t_1+\xi}^{t_2-\xi} |e(t)|dt + \ldots$$

$$\ldots + \frac{\partial}{\partial C_k}\int_{t_x+\xi}^{T_0} |e(t)|dt \tag{4.49}$$

Let us consider any one of the above intervals, say $t_y < t < t_z$. Since in this interval the error $|e(t_, C_K)|$ is a continuous function of t and C_k we can always find a neighbourhood of C_k such that for any $t\varepsilon(t_y, t_z)$, $\{\text{Sgn } e(t)\}\phi_k(t)$ is a continuous function of t and C_k in this region. Hence in this neighbourhood

$$\frac{\partial}{\partial C_k}\int_{t_y+\xi}^{t_z-\xi} |e(t)|dt$$

exists and is given by

$$\frac{\partial}{\partial C_k}\int_{t_y+\xi}^{t_z-\xi} |e(t)|dt = \int_{t_y+\xi}^{t_y-\xi} \frac{\partial}{\partial C_k}|e(t)|dt$$

$$= \int_{t_y+\xi}^{t_z-\xi} -\{\text{Sgn } e(t)\}\phi_K(t)dt \tag{4.49}$$

Hence $\frac{\partial\gamma}{\partial C_k}$ can be written as

$$\frac{\partial\gamma}{\partial C_k} = \int_0^{t_1-\xi} -\{\mathrm{Sgn}\ e(t)\}\phi_k(t)dt + \int_{t_1+\xi}^{t_2-\xi} -\{\mathrm{Sgn}\ e(t)\}\phi_k(t)dt$$

$$+ \ldots + \int_{t_x+\xi}^{T_0} -\{\mathrm{Sgn}\ e(t)\}\phi_K(t)dt \qquad (4.50)$$

where $\xi \simeq 0$.

Assuming that the integral in the following equation is always evaluated as given in equation (4.50) we can write

$$\frac{\partial\gamma}{\partial C_k} = \int_0^{T_0} -\{\mathrm{Sgn}\ e(t)\}\phi_k(t)dt \qquad (4.51)$$

Hence

$$\mathrm{Grad}\ \gamma = \left[\frac{\partial\gamma}{\partial C_1}\ \ \frac{\partial\gamma}{\partial C_2}\ \ \cdots\ \ \frac{\partial\gamma}{\partial C_N}\right]^t$$

$$= -\int_0^{T_0} \mathrm{Sgn}\ e(t)\Big[\phi_1(t)\ \ \phi_2(t)\ \ \ldots.\ \ \phi_N(t)\Big]^t \quad (4.52)$$

In this chapter we have developed a method of minimizing the upper bound of the deviation in the frequency domain, keeping ISE within allowable limits. The implementation of this method requires the computation of the best pole positions of least ISE and the availability of the orthonormal functions $\phi_k(t)$, $k = 1,2,\ldots N$ in the time domain. The next chapter develops a new method of obtaining the time domain representations of functions like $\Phi_k(s)$. The minimization of average error with constraints on ISE is discussed with examples in Chapter 6.

CHAPTER 5

EVALUATION OF TRANSIENT RESPONSE

INTRODUCTION

This chapter develops a new numerical method of computing the transient response of a Laplace transform, expressed as a rational function of complex frequency. This method was primarily developed to compute the time domain representations of functions $\Phi_k(s)$, $k = 1,2,\ldots N$ discussed in previous chapters. However this new technique of computing the transient response is very general and computationally more efficient than some of the previous methods (5).

A BRIEF SURVEY OF PREVIOUS INVESTIGATIONS ON COMPUTATION OF TRANSIENT RESPONSE

The evaluation of the inverse of a Laplace transform expressed as

$$Y(s) = \frac{\sum\limits_{i=1}^{n} b_i s^{n-i}}{s^n + \sum\limits_{i=1}^{n} a_i s^{n-i}}, \quad a_i \neq 0 \qquad (5.1)$$

is commonly encountered in many branches of Engineering and Science. The classical approach of evaluating the transient response is to express the rational function Y(s) as partial fractions and find the inverse of each factor. Generally this method is very complicated. Moreover when two of the roots of the denominator are very close to each other, the evaluation of the corresponding residues of the partial fraction expansion leads to numerical difficulties. Because of these problems, other numerical methods of evaluating the transient response have been developed.

Corrington proves (40) that a rational operational form of equation (5.1) leads to a linear n-term difference equation of the form given below

$$y(t) = \sum_{i=1}^{n} (-1)^{i+1} F_{n,i}\, y(t - i\Delta t), \; t > n\Delta t \qquad (5.2)$$

The coefficients $F_{n,i}$ are real constants, independent of t but dependent on Δt. These coefficients may be found from the following relation,

$$\prod_{k=1}^{n} (s + e^{s_k \Delta t}) = \sum_{k=0}^{n} F_{n,k} s^{n-k} \qquad (5.3)$$

where $(s_k)_{k=1}^{n}$ are the n poles of Y(s). Corrington suggests an easier way of determining the coefficients $F_{n,i}$, i = 1,2,...n. In this method Y(s) is expressed as

$$Y(s) = \sum_{i=0}^{\infty} \frac{c_i}{s^{i+1}} \qquad (5.4)$$

The coefficients c_i, $i = 1,2,\ldots n$ are found by dividing the denominator into the numerator. If the c_i coefficients are known y(t) can be written as

$$y(t) = \sum_{i=0}^{\infty} \frac{c_i t^i}{i!} \qquad (5.5)$$

Equations (5.2) and (5.5) are used to form the (n-1) simultaneous equations in $(F_{n,i})_{i=1}^{n-1}$. Since $F_{n,n}$ is readily known from equation (5.3), these (n-1) equations can be solved to find $F_{n,i}$, $i = 1,2,\ldots(n-1)$. Further computations of y(t) are carried out by using equation (5.2). Aaron and Kaiser have pointed out some of the numerical difficulties one encounters while using this method (41). The division of two polynomials to find the c_i coefficients is difficult and leads to large errors (41).

Another method of evaluating the transient response is reported by Liou (42). This method makes use of the state space approach. Liou considers the differential equation which gives rise to the rational function Y(s) of equation (5.1). This differential equation is

$$D^n y(t) + \sum_{i=1}^{n} a_i D^{n-i} y(t) = 0 \qquad (5.6)$$

where $D^i y(t) = \frac{d^i}{dt^i} y(t)$. The numerator coefficients of equation (5.1), $(b_i)_{i=1}^{n}$, may be expressed in terms of the initial values of y(t) and its first (n-1) derivatives. Let $\overline{Y}(t)$ be a n-dimensional vector defined as

$$\overline{Y}(t) = [y(t) \quad Dy(t) \; \ldots . \; D^{n-1}y(t)]^t \qquad (5.7)$$

$\overline{Y}(t)$ can be expressed as a state space equation

$$\dot{\overline{Y}}(t) = A\overline{Y}(t) \qquad (5.8)$$

where $\dot{\overline{Y}}(t) = \frac{d}{dt}\overline{Y}(t)$ and A is a nxn matrix. A is given as,

$$A = \begin{bmatrix} 0 & 1 & 0 \ldots\ldots\ldots\ldots 0 & 0 \\ 0 & 0 & 1 \ldots\ldots\ldots\ldots 0 & 0 \\ \vdots & \vdots & \vdots \qquad\qquad\quad \vdots & \vdots \\ 0 & 0 & 0 \ldots\ldots\ldots\ldots 0 & 1 \\ -a_n & -a_{n-1} & -a_{n-2} \qquad\quad -a_2 & -a_1 \end{bmatrix} \qquad (5.9)$$

The solution of the state space equation (5.8) is

$$\overline{Y}(t) = e^{At}\overline{Y}(0^+) \qquad (5.10)$$

where $\overline{Y}(0^+)$ is the initial condition vector. If we consider two points t = iT and t = (i + 1)T, we have

$$\begin{aligned} \overline{Y}(iT) &= e^{A(iT)}\overline{Y}(0^+) \\ \overline{Y}\{(i+1)T\} &= e^{A\{(i+1)T\}}\overline{Y}(0^+) \end{aligned} \tag{5.11}$$

Equations (5.11) give,

$$\overline{Y}\{(i+1)T\} = e^{AT}\overline{Y}(iT) \tag{5.12}$$

Once the initial vector $\overline{Y}(0^+)$ is known, successive values of $\overline{Y}(iT)$ can be computed using equation (5.12). The initial vector $\overline{Y}(0^+)$ is given by

$$\overline{Y}(0^+) = \begin{bmatrix} y(0^+) \\ y'(0^+) \\ \cdot \\ \cdot \\ \cdot \\ y^{n-1}(0^+) \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 - a_1 y(0^+) \\ \cdot \\ \cdot \\ \cdot \\ b_n - a_1 y^{n-2}(0^+) \ldots a_{n-1} y(0^+) \end{bmatrix} \tag{5.13}$$

where $y^i(0^+) = \left.\frac{d^i}{dt^i}y(t)\right|_{t = 0^+}$

The transient response $y(iT)$ is given by the first element of the vector $\overline{Y}(iT)$.

One difficulty of the method is the necessity of evaluating the state transition matrix e^{AT}. Liou suggests the power series evaluation of this matrix. e^{AT} is expanded as follows.

$$e^{AT} = I + \frac{AT}{1!} + \frac{A^2T^2}{2!} + \ldots$$

$$= \sum_{i=0}^{\infty} \frac{A^iT^i}{i!} \qquad (5.14)$$

where $A^\circ = I$, Identity matrix.

The matrix e^{AT} is approximated by a matrix M, where

$$e^{AT} \simeq M = \sum_{i=0}^{K} \frac{A^iT^i}{i!} \qquad (5.15)$$

Liou has developed a technique of choosing K such that the error involved in the approximation e^{AT} as in equation (5.15) is within a specified limit.

The computation of e^{AT} can be simplified by using the relations between the last column and other columns of e^{AT} matrix. These relations are proved by Thomson (43).

Let $m_{in}(t)$, $i = 1,2,\ldots,n$ be the last column of e^{AT} matrix. Thomson proves (43)

$$m_{in}(t) = \sum_{k=0}^{\infty} c_{k+i-1} \frac{t^k}{k!} \qquad (5.16)$$

where $c_0 = c_1 = \ldots = c_{n-2} = 0$, $c_{n-1} = 1$.

and

$$c_k = -\sum_{j=1}^{n} a_j c_{k-j}, \; k \geqslant n \qquad (5.17)$$

The elements of the other columns of e^{AT} matrix can be expressed in a recursive form as follows

$$m_{i,j} = m_{i+1,j+1} + a_{n-j}m_{in}, \quad 1 \leqslant i \leqslant j < n$$

$$m_{i,1} = -a_n m_{i-1,n} \qquad 1 < i \leqslant n \tag{5.18}$$

$$m_{i,j} = m_{i-1,j-1} - a_{n-j+1}m_{i-1,n}, \quad 1 < j < i \leqslant n$$

Valand makes use of the results of Thomsón to modify the approach of Liou in order to avoid the computation of e^{AT} matrix (44).

Equation (5.1) can be expressed as

$$Y(s) = \sum_{i=1}^{n} b_i s^{n-i} Y_L(s) \tag{5.19}$$

where $Y_L(s) = \dfrac{1}{s^n + \sum\limits_{i=1}^{n} a_i s^{n-i}}$ (5.20)

The initial vector of $Y_L(s)$, as defined by Liou is

$$\bar{Y}_L(0^+) = [0 \quad 0 \; \ldots \; 1]^t \tag{5.21}$$

Making use of equation (5.10) we obtain

$$\bar{Y}_L(t) = e^{At}\bar{Y}_L(0^+)$$

Since $\bar{Y}_L(0^+)$ has unity for the last element and all other elements are zeros

$$\bar{Y}_L(t) = [m_{1n}(t) \quad m_{2n}(t) \; \ldots \; m_{nn}(t)]^t \tag{5.23}$$

where $m_{in}(t)$ is the element of i^{th} row and n^{th} (last) column of e^{At}. From the definition of $\overline{Y}_L(t)$ in equation (5.7), we have

$$\frac{d^i}{dt^i} y_L(t) = m_{i+1,n}(t) \tag{5.24}$$

From (5.19),

$$y(t) = \sum_{i=1}^{n} b_i L^{-1}\ s^{n-i} Y_L(s)$$

$$= \sum_{i=1}^{n} b_i D^{n-i} y_L(t) \tag{5.25}$$

Substituting (5.24) into (5.25)

$$y(t) = \sum_{i=1}^{n} b_i m_{n-i+1,n}(t) \tag{5.26}$$

In this method the c_k values are computed first, using equation (5.17). Factors like $(c_{k+i-1}/i!)$ are then computed for i = 0, 1, ... (n-1) and stored. The values of $m_{in}(t)$, i = 1,2,...,n can now be computed using equations (5.16) and y(t) is obtained from equation (5.26). This method does not involve the computation of e^{AT} matrix. However the terms $m_{in}(t)$, i = 1,2,...,n must be found for each value of t. This involves the summing of n series at each value of t.

NEW METHOD

In the new method, y(t) is regarded as the output response of the system

$$\frac{\sum_{i=1}^{n} b_i s^{n-i}}{s^n + \sum_{i=1}^{n} a_i s^{n-i}},$$

when an impulse input is applied. Let us consider a general system

$$\frac{Y(s)}{U(s)} = \frac{\sum_{i=0}^{n} b_i s^{n-i}}{s^n + \sum_{i=1}^{n} a_i s^{n-i}} \qquad (5.27)$$

where Y(s) and U(s) are the Laplace transforms of the system output and input respectively. The system output may be evaluated by using the state space approach (45-46). The general differential equation of the system represented by the equation (5.27) is

$$D^n y(t) + \sum_{i=1}^{n} a_i D^{n-i} y(t) = \sum_{i=0}^{n} b_i D^{n-i} u(t) \qquad (5.28)$$

Let a new variable z(t) be defined such that

$$D^n z(t) + \sum_{i=1}^{n} a_i D^{n-i} z(t) = u(t) \qquad (5.29a)$$

and

$$\sum_{i=0}^{n} b_i D^{n-i} z(t) = y(t) \qquad (5.29b)$$

Substituting y(t), as given by equation (5.29) into (5.28)

$$
\begin{aligned}
D^n y(t) &+ \sum_{i=1}^{n} a_i D^{n-i} y(t) \\
&= D^n \left\{ \sum_{k=0}^{n} b_k D^{n-k} z(t) \right\} + \sum_{i=1}^{n} a_i D^{n-i} \left\{ \sum_{k=0}^{n} b_k D^{n-k} z(t) \right\} \\
&= \sum_{k=0}^{n} b_k D^{n-k} \left\{ D^n z(t) \right\} + \sum_{k=0}^{n} b_k D^{n-k} \left\{ \sum_{i=1}^{n} a_i D^{n-i} z(t) \right\} \\
&= \sum_{k=0}^{n} b_k D^{n-k} \left\{ D^n z(t) + \sum_{i=1}^{n} a_i D^{n-i} z(t) \right\} \\
&= \sum_{k=0}^{n} b_k D^{n-k} u(t) \qquad (5.30)
\end{aligned}
$$

Equation (5.30) shows that equations (5.28) and (5.29) are equivalent. Hence if we solve z(t) from equation (5.29a) and substitute into (5.29b), y(t) can be obtained. Consider the equation

$$D^n z(t) + \sum_{i=1}^{n} a_i D^{n-i} z(t) = u(t) \qquad (5.31)$$

Let us define a set of state variables $y_i(t)$, i = 1,2,..., n such that

$$y_i(t) = D^{i-1} z(t) \qquad (5.32)$$

We then have

$$
\begin{aligned}
\dot{y}_1(t) &= Dz(t) = y_2(t) \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
\dot{y}_i(t) &= D^i z(t) = y_{i+1}(t) \\
&\cdots\cdots\cdots\cdots\cdots\cdots \qquad\qquad\qquad (5.33) \\
\dot{y}_n(t) &= D^n z(t) = -\sum_{i=1}^{n} a_i D^{n-i} z(t) \;+ u(t) \\
&= \left\{-\sum_{i=1}^{n} a_i y_{n-i+1}(t)\right\} + u(t)
\end{aligned}
$$

Equation (5.33) may be expressed as a state space equation.

$$
\begin{bmatrix} \dot{y}_1(t) \\ \dot{y}_2(t) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \dot{y}_{n-1}(t) \\ \dot{y}_n(t) \end{bmatrix}
=
\begin{bmatrix}
0 & 1 & 0\ldots\ldots\ldots\ldots0 & 0 \\
0 & 0 & 1\ldots\ldots\ldots\ldots0 & 0 \\
\cdot & \cdot & \cdot \qquad\qquad\quad \cdot & \cdot \\
\cdot & \cdot & \cdot \qquad\qquad\quad \cdot & \cdot \\
\cdot & \cdot & \cdot \qquad\qquad\quad \cdot & \cdot \\
\cdot & \cdot & \cdot \qquad\qquad\quad \cdot & \cdot \\
\cdot & \cdot & \cdot \qquad\qquad\quad \cdot & \cdot \\
0 & 0 & 0\ldots\ldots\ldots\ldots0 & 1 \\
-a_n & -a_{n-1} & -a_{n-2}\ldots\ldots-a_2 & -a_1
\end{bmatrix}
\begin{bmatrix} y_1(t) \\ y_2(t) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_{n-1}(t) \\ y_n(t) \end{bmatrix}
$$

$$
+ \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \end{bmatrix} u(t) \qquad\qquad (5.34)
$$

or

$$\dot{\overline{Y}} = A\overline{Y} + Bu(t) \tag{5.35}$$

where

$$\overline{Y} = [y_1 \quad y_2 \cdots y_n]^t$$

and $$B = [0 \quad 0 \ldots . \; 1]^t$$

The A matrix is the same as the one defined by equation (5.9). The output y(t) is obtained from (5.29) as

$$\begin{aligned} y(t) &= \sum_{i=0}^{n} b_i D^{n-i} z(t) \\ &= b_0 D^n z(t) + \sum_{i=1}^{n} b_i y_{n-i+1}(t) \\ &= b_0 \left\{ -\sum_{i=1}^{n} a_i y_{n-i+1}(t) + u(t) \right\} + \sum_{i=1}^{n} b_i y_{n-i+1}(t) \\ &= \left\{ \sum_{i=1}^{n} (b_i - a_i b_0) y_{n-i+1}(t) \right\} + b_0 u(t) \end{aligned} \tag{5.36}$$

Equation (5.36) may be written as

$$y(t) = C\overline{Y} + b_0 u(t) \tag{5.37}$$

where

$$C^t = \begin{bmatrix} b_n - a_n b_0 \\ \cdots\cdots\cdots \\ b_i - a_i b_0 \\ \cdots\cdots\cdots \\ b_1 - a_1 b_0 \end{bmatrix} \tag{5.38}$$

The equations (5.35) and (5.37) can be made use of to compute y(t) for any input u(t). The solution of the state space equation (5.35) is given as (45-46)

$$\bar{Y}(t) = e^{At}\bar{Y}(o) + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau \qquad (5.39)$$

The output y(t) is then given by

$$y(t) = \{Ce^{At}\bar{Y}(o) + C\int_0^t e^{A(t-\tau)}Bu(\tau)d\tau\} + b_o u(t) \quad (5.40)$$

For the special case of the transient response of the system defined by equation (5.1) $b_o = 0$ and $u(t) = \delta(t)$, the impulse at t = 0. The initial vector $\bar{Y}(o)$ can be considered as that at $t = 0^-$. Hence

$$\bar{Y}(o) = \bar{Y}(o^-) = [o] \qquad (5.41)$$

This gives

$$\begin{aligned} y(t) &= C\int_{0-}^{t} e^{A(t-\tau)}B\delta(\tau)d\tau \\ &= Ce^{At}B \qquad (5.42) \end{aligned}$$

where

$$C = [b_n \ \ b_{n-1} \ldots b_1] \qquad (5.43)$$

The matrix A and vector B are given by the equations (5.9) and (5.35) repsectively.

Let e^{At} in equation (5.42) be approximated by

K terms. The approximate y(t) is then given by

$$y(t) = C\left\{\sum_{i=0}^{K-1} \frac{A^i t^i}{i!}\right\}B$$

$$= C\left[I + \frac{At}{1!} + \frac{A^2t^2}{2!} + \ldots + \frac{A^{K-1}t^{K-1}}{(K-1)!}\right]B \qquad (5.44)$$

Since factors like $\frac{t^i}{i!}$ are scalars y(t) can be written as

$$y(t) = C\left[IB + \frac{AB}{1!}t + \frac{A^2B}{2!}t^2 + \ldots + \frac{A^{K-1}B}{(K-1)!}t^{K-1}\right] \qquad (5.45)$$

The matrix inside the bracket in equation (5.45) can be expressed as the product of a partitioned matrix and another vector. Thus we get

$$y(t) = C\left[IB \,\vdots\, AB \,\vdots\, A^2B \,\vdots \cdots\cdots \vdots\, A^{K-1}B\right]\left[1 \;\; \frac{t}{1!} \;\; \frac{t2}{2!} \;\; \cdot\cdot \frac{t^{K-1}}{(K-1)!}\right]^t \qquad (5.46)$$

Since the vector B has zeros for the first (n-1) elements and unity for the last element we have

$$\left[IB \,\vdots\, AB \,\vdots\, A^2B \,\vdots \cdots\cdots \vdots\, A^{K-1}B\right] = \left[A_0 \,\vdots\, A_1 \,\vdots\, A_2 \,\vdots \cdots\cdots \vdots\, A_{K-1}\right] = \Phi$$

$$(5.47)$$

where A_i is the last column of A^i matrix, $A^0 = I$, the Identity matrix and Φ is a nxK matrix. The i^{th} column of the Φ matrix is the last column of the matrix A^{i-1}. Hence there exists a recursive relation between any two successive columns of the Φ matrix. This relation is proved as follows.

$$A^{i} = AA^{i-1} \tag{5.48}$$

Let the last column of A^{i-1} be given in terms of the elements of the Φ matrix. That is

$$A_{i-1} = [\phi_{1i} \quad \phi_{2i} \cdots \phi_{ni}]^{t} \tag{5.49}$$

where ϕ_{jk} is the element of j^{th} row and k^{th} column of Φ matrix.

From equation (5.48) we get

$$A_i = A\,A_{i-1}$$

That is

$$A_i = \begin{bmatrix} \phi_{1,i+1} \\ \phi_{2,i+1} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \phi_{n,i+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \ldots\ldots\ldots 0 & 0 \\ 0 & 0 & 1 \ldots\ldots\ldots 0 & 0 \\ \cdot & \cdot & \cdot \qquad\quad \cdot & \cdot \\ \cdot & \cdot & \cdot \qquad\quad \cdot & \cdot \\ \cdot & \cdot & \cdot \qquad\quad \cdot & \cdot \\ & & \ldots\ldots\ldots 0 & 1 \\ -a_n & -a_{n-1} & -a_{n-2} \ldots\ldots -a_2 & -a_1 \end{bmatrix} \begin{bmatrix} \phi_{1i} \\ \phi_{2i} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \phi_{ni} \end{bmatrix} \tag{5.50}$$

The evaluation of equation (5.50) gives the recursive relation as

$$\begin{bmatrix} \phi_{1,i+1} \\ \phi_{2,i+1} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \phi_{n,i+1} \end{bmatrix} = \begin{bmatrix} \phi_{2,i} \\ \phi_{3,i} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ -\sum\limits_{i=1}^{n} a_i \phi_{n-i+1,i} \end{bmatrix} \tag{5.51}$$

This recursive relation can also be expressed as

$$\phi_{j,i+1} = \phi_{j+1,i}, \quad j = 1,2,\ldots,(n-1) \tag{5.52}$$

and $\phi_{n,i+1} = -\sum\limits_{i=1}^{n} a_i \phi_{n-i+1,i}$

The relation (5.52) is valid for all values of i. Hence knowing the first column of Φ, which is the last column of A^0, that is, the last column of the Identity matrix, the transient response can be computed from

$$y(t) = C\Phi\psi \tag{5.53}$$

where $\psi = \left[1 \ \frac{t}{1!} \ \frac{t^2}{2!} \ \cdots \ \frac{t^{K-1}}{(K-1)!}\right]^t$

Expanding $C\Phi$, we have

$$C\Phi = [b_n \quad b_{n-1}\cdots b_1] \begin{bmatrix} \phi_{11}\cdots\phi_{1i}\cdots\phi_{1K} \\ \phi_{21}\cdots\phi_{2i}\cdots\phi_{2K} \\ \cdot \quad\quad \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \quad\quad \cdot \\ \phi_{n1}\cdots\phi_{ni}\cdots\phi_{nK} \end{bmatrix} \tag{5.54}$$

Using equations (5.53) and (5.54) y(t) can be expressed as

$$y(t) = \sum_{i=1}^{K} h_i t^{i-1} \qquad (5.55)$$

where $h_i = \sum_{j=1}^{n} (b_j \phi_{n-j+1,i})/(i-1)!$ (5.56)

The evaluation of each column of Φ matrix and the corresponding h_i coefficient may be done simultaneously. Once the h_i coefficients are computed and stored the transient response y(t) can be evaluated for any t.

The error involved in approximating y(t) in the above manner is now studied. We make use of the approach of Liou (42) to find an upperbound of the error. The exact expression for y(t) is

$$\begin{aligned} y(t) &= Ce^{At}B \\ &= C\left\{\sum_{i=0}^{\infty} \frac{A^i t^i}{i!}\right\}B \\ &= C\left\{\sum_{i=0}^{K-1} \frac{A^i t^i}{i!}\right\}B + C\left\{\sum_{i=K}^{\infty} \frac{A^i t^i}{i!}\right\}B \qquad (5.57) \end{aligned}$$

We are approximating y(t) by the first term of the right hand side of equation (5.57)

$$y(t) \simeq C\left\{\sum_{i=0}^{K-1} \frac{A^i t^i}{i!}\right\}B \qquad (5.58)$$

The error $\varepsilon(t)$ is given by

$$\varepsilon(t) = C \sum_{i=K}^{\infty} \frac{A^i t^i}{i!} B \qquad (5.59)$$

$$= CRB$$

where
$$R = \sum_{i=K}^{\infty} \frac{A^i t^i}{i!} \qquad (5.60)$$

R is a nxn matrix.

If r_{max} is an upper bound of each element r_{ij} of R so that

$$|r_{ij}| \leqslant r_{max} \qquad (5.61)$$

we obtain from equation (5.59)

$$|\varepsilon(t)| \leqslant r_{max} \left(\sum_{i=1}^{n} |b_i| \right) \qquad (5.62)$$

An upperbound r_{max} can be found by defining a suitable norm for the matrix A. Liou suggests the norm as

$$||A|| = \sum_{i,j=1}^{n} |\alpha_{ij}| \qquad (5.63)$$

where α_{ij} is the $(i,j)^{th}$ element of A matrix (42). Later he has modified this definition of $||A||$ (46) as,

$$||A|| = \max_{i} \left(\sum_{j=1}^{n} |\alpha_{ij}| \right) \qquad (5.64)$$

It is found that in most cases K, the number of terms estimated by using the definition of $||A||$ in equations (5.63) and (5.64) is much larger than what actually is required. Hence we make use of a modified definition of the norm of A as

$$||A|| = \max_{j} \left(\sum_{i=1}^{n} |\alpha_{ij}| \right) \qquad (5.65)$$

According to this definition, the norm of A matrix of equation (5.9) is

$$||A|| = \max_{i=1,n-1} (|a_n|, 1 + |a_i|) \qquad (5.66)$$

If $|a_i| >> 1$, then the norm of A is equal or approximately equal to the largest magnitude of a_i, $i = 1,2,\ldots,n$. We now prove that the definition of equation (5.65) satisfies all the conditions of a norm. These conditions are

1) $A \neq 0$ implies $||A|| > 0$
2) $||\alpha A|| = |\alpha| \, ||A||$, α is a scalar
3) $||A_x + A_y|| \leqslant ||A_x|| + ||A_y||$

where A_x and A_y are any two (nxn) matrices.

By definition $(a_i)_{i=1}^{n}$ are not zeros. Hence condition 1 is always satisfied.

For Condition 2, let

$$||A|| = \sum_{i=1}^{n} |\alpha_{ip}|$$

Then

$$||\alpha A|| = \sum_{i=1}^{n} |\alpha\alpha_{ip}|$$

$$= |\alpha| \sum_{i=1}^{n} |\alpha_{ip}|$$

$$= |\alpha| \; ||A|| \qquad (5.68)$$

This proves Condition 2.

Condition 3 is proved as follows. Let α_{ij} and β_{ij} be the $(i,j)^{th}$ element of the matrices A_x and A_y respectively. Then

$$||A_x + A_y|| = \max_{j=1,n}\left(\sum_{i=1}^{n} |\alpha_{ij} + \beta_{ij}|\right)$$

$$\leqslant \max_{j=1,n}\left(\sum_{i=1}^{n} |\alpha_{ij}| + \sum_{i=1}^{n} |\beta_{ij}|\right)$$

$$\leqslant \max_{j=1,n}\left(\sum_{i=1}^{n} |\alpha_{ij}|\right) + \max_{j=1,n}\left(\sum_{i=1}^{n} |\beta_{ij}|\right)$$

$$= ||A_x|| + ||A_y||$$

Or $||A_x + A_y|| \leqslant ||A_x|| + ||A_y||$ (5.69)

A fourth condition necessary in finding an upperbound of r_{ij} is

$$||A^k|| \leqslant ||A||^k \tag{5.70}$$

Let α_{ij}^{k+1} and α_{ij}^{k} be the (i, j)th elements of the matrices A^{k+1} and A^k respectively. Then

$$\alpha_{ij}^{k+1} = \sum_{p=1}^{n} (\alpha_{ip}^{k}\alpha_{pj}) \tag{5.71}$$

Hence

$$\begin{aligned} |\alpha_{ij}^{k+1}| &\leqslant \sum_{p=1}^{n} |\alpha_{ip}^{k}\alpha_{pj}| \\ &= \sum_{p=1}^{n} |\alpha_{ip}^{k}|\ |\alpha_{pj}| \end{aligned} \tag{5.72}$$

Therefore

$$\begin{aligned} \sum_{i=1}^{n} |\alpha_{ij}^{k+1}| &\leqslant \sum_{i=1}^{n}\left(\sum_{p=1}^{n} |\alpha_{ip}^{k}|\ |\alpha_{pj}|\right) \\ &= \sum_{p=1}^{n}\ \sum_{i=1}^{n}\ |\alpha_{ip}^{k}|\ |\alpha_{pj}| \\ &\leqslant \sum_{p=1}^{n}\ ||A^k||\ |\alpha_{pj}| \\ &= ||A^k|| \sum_{p=1}^{n} |\alpha_{pj}| \\ &\leqslant ||A^k||\ ||A|| \end{aligned} \tag{5.73}$$

Equation (5.73) is true for any column of A^{k+1} and hence we get

$$||A^{k+1}|| \leqslant ||A^{k}||\ ||A||$$

$$\leqslant ||A||^{k+1} \qquad (5.74)$$

Applying the above properties of the norm to equation (5.60) we obtain

$$|r_{ij}| \leqslant ||R|| \leqslant \sum_{i=K}^{\infty} \frac{||A||^{i}t^{i}}{i!}$$

$$= \frac{||A||^{K}t^{K}}{K!}\left(1 + \frac{||A||t}{(K+1)} + \frac{||A||^{2}t^{2}}{(K+1)(K+2)} + \ldots\right) \qquad (5.75)$$

Let $\frac{||A||t}{K+1} = x \qquad (5.76)$

Then $\frac{||A||^{2}t^{2}}{(K+1)(K+2)} < x^{2}$

Hence we get

$$|r_{ij}| \leqslant \frac{||A||^{K}t^{K}}{K!}(1 + x + x^{2} + \ldots\ldots) \qquad (5.77)$$

If $|x| < 1$,

$$|r_{ij}| \leqslant \frac{||A||^{K}t^{K}}{K!}\ \frac{1}{1 - x} \qquad (5.78)$$

Thus we get an upperbound on $\varepsilon(t)$ as

$$|\varepsilon(t)| \leqslant \left(\frac{||A||^K t^K}{K!} \frac{1}{1 - x}\right)\left(\sum_{i=1}^{n} |b_i|\right) \qquad (5.79)$$

By making use of equation (5.79) it is possible to choose K such that the error is within allowable limits.

This method of computing the transient response has many advantages. It is not necessary to compute e^{AT} matrix and the initial vector as required in Liou's method (42). The method suggested by Valand involves the manipulation of n series while the new method involves only one series for all values of n. The distribution of the eigenvalues of the system does not affect the method. For example, even if two poles are identical or very close to each other the method is capable of giving the response without any further modification. Since y(t) is expressed as a power series in t the derivatives at any value of t are easily obtainable. The round off errors of computation do not propagate with increasing values of t. This is because y(t) is evaluated at each t by the power series and hence is exact.

However, when $||A||$ is very large and it is required to compute y(t) at a large value of t some difficulties are encountered. Under such circumstances it is necessary to consider a large number of terms. One method of solving this problem is to split Y(s) into smaller factors. The denominator of these factors of Y(s) can be formed from the poles of the system. The numerator coefficients of the factors of Y(s) may be found by solving the n simultaneous equations obtained by substituting n values of s. But a better way of doing this is to form n simultaneous equations by using the initial value of y(t) and its first (n - 1) derivatives. These values are easily obtained by finding the first n coefficients of y(t) as discussed in these chapters. These coefficients are equated to the corresponding coefficients of the factors to obtain the n simultaneous linear equations. The solution of these equations determine the above factors. As $||A||$ of each of these factors has smaller value, the number of terms K necessary in each case is also smaller. The transient response y(t) is then obtained as the sum of the transient responses of the individual factors of Y(s).

CHAPTER 6

COMPUTER ALGORITHMS AND NUMERICAL EXAMPLES

INTRODUCTION

In this chapter we first discuss the computational techniques employed to find the best pole positions by minimizing the ISE. Instead of employing any of the iterative techniques discussed in Chapter 3, we rely on efficient numerical methods of minimizing functions of several variables without calculating derivatives (47, 48). This avoids the complicated filtering operations required in all the exact methods of minimizing ISE reviewed in Chapter 3. The minimization of average error, with constraints on ISE, is done by using the penalty function approach (51). Finally the method is applied to a specific example which is an ideal low pass filter.

COMPUTATION OF ISE

The concept of the complimentary filter is used to compute the ISE for any set of pole positions. The expression for the ISE, as given in Chapter 2, is

$$\text{ISE} = I = \int_{-\infty}^{0} |\bar{a}(t)|^2 dt \qquad (6.1)$$

where $\bar{a}(t)$ is the output of the complimentary filter when v(t), the reversed f(t), is applied as an input. This relation was discussed in detail in Chapter 2. The complimentary filter G(s) is given by

$$G(s) = \prod_{i=1}^{N} \frac{s + s_i}{s - s_i} = \frac{s^N - a_1 s^{N-1} + \ldots + (-1)^N a_N}{s^N + a_1 s^{N-1} + \ldots + a_N} \qquad (6.2)$$

The scheme for evaluating the ISE is shown in Fig. 2.1. The given function f(t) is time reversed and applied to the filter G(s). Since any physical signal vanishes for some $t \geqslant T_0$, the reversed signal v(t) may be considered as starting at $t = 0$ and extending to $t = T_0$. The expression for the ISE then becomes

$$\text{ISE} = I = \int_0^{T_0} |\bar{a}(t)|^2 dt \qquad (6.3)$$

The output $\bar{a}(t)$ of the complimentary filter is computed by using equation (5.36). This equation gives the general expression for the output of any filter G(s) which has the form given by equation (6.2). Thus the expression for $\bar{a}(t)$ is

$$\bar{a}(t) = \left[\sum_{i=1}^{N} \{(-1)^i - 1\} a_i y_{N-i+1}(t) \right] + v(t) \qquad (6.4)$$

Equation (6.4) involves the evaluation of the state variables $(y_i)_{i=1}^{N}$. These variables are defined by equations (5.31 - 5.34). The solution of equation (5.34) gives the values of $(y_i)_{i=1}^N$. This is accomplished by the Runge-Kutta method which is written up as a standard subroutine in the IBM Scientific Subroutine Package Library. The initial values of y_i, are zeros. Once $\bar{a}(t)$ is known the ISE is computed by evaluating the integral of equation (6.3). The integration uses the Five point Quadrature formula. A subroutine ERROR is written in Fortran IV. This subroutine accepts a set of coefficients $(a_i)_{i=1}^N$, the reversed signal v(t) and computes the corresponding ISE of least square representation. This program is given in Appendix I.

MINIMIZATION OF ISE

The general approach employed in minimizing the ISE is to consider the ISE as a function of the parameters of the complimentary filter and to find these parameters such that the ISE is minimized. This has the advantage of reducing the original 2N variables of Aigrain and Williams equations to N variables of the complimentary filter. Another advantage of this approach is that, if $(a_i)_{i=1}^N$ are considered as the variables, the operations involving complex numbers can be completely avoided.

A direct method of minimization is to use any one of the gradient techniques. Unfortunately, as reported in Chapter 3, the ISE is very insensitive to pole positions over a wide range near the optimum point. This makes the gradient techniques very inefficient in minimizing the ISE. Since we want to avoid the complicated filtering operations necessary for the various iterative schemes, a new method of minimization was tried and found to be very useful. This method does not require the computation of the gradients. The basic approach employed is due to Powell (47). Powell uses the method of conjugate directions to minimize functions of several variables. Zangwill has pointed out (48) certain drawbacks in Powell's method and has suggested a new algorithm. This algorithm incorporates all the basic features of Powell's method but does not suffer from the difficulties pointed out by Zangwill. We have made use of the Zangwill algorithm with some modifications, to minimize ISE. The basic theory of these algorithms is now discussed.

FUNCTION MINIMIZATION WITHOUT CALCULATING DERIVATIVES

(1) Powell's Method

The minimization of a function of several variables without calculating derivatives has attracted the attention of several investigators (47-51). Among these the method of Powell (47) is considered the best as this will minimize a quadratic function of n variables in n iterations. Powell considers the minimization of a quadratic function Q(X), of a n-vector X.

Let

$$Q(X) = X^tPX + B^tX + c \qquad (6.5)$$

where P is a nxn matrix, B- is a n- vector and c a scalar constant. Let $(\xi_i)_{i=1}^n$ be n conjugate directions such that

$$\xi_i^tP\xi_j = 0, \quad i \neq j \qquad (6.6)$$

Let X_0 be an initial point. Any point X may be expressed in terms of these conjugate directions.

Let

$$X = X_0 + \sum_{i=1}^n \alpha_i\xi_i$$

$$Q(X) = Q(X_0 + \sum_{i=1}^n \alpha_i\xi_i)$$

$$= (X_0 + \sum_{i=1}^n \alpha_i\xi_i)^tP(X_0 + \sum_{i=1}^n \alpha_i\xi_i)$$

$$+ B^t(X_0 + \sum_{i=1}^n \alpha_i\xi_i) + c$$

Simplifying by using equation (6.6), we obtain

$$Q(X) = Q(X_0) + \sum_{i=1}^n \alpha_i(X_0^tP\xi_i + \xi_i^tPX_0)$$

$$+ \sum_{i=1}^n \alpha_iB^t\xi_i + \sum_{i=1}^n \alpha_i^2\xi_i^tP\xi_i \qquad (6.7)$$

Examining equation (6.7) it is found that minimization of $Q(X)$ by searching in one of the conjugate directions is independent of the other conjugate directions. Hence minimum of $Q(X)$ may be found by searching along each of the conjugate directions only once. The problem of minimizing a quadratic function is thus reduced to the problem of finding n conjugate directions. Powell suggests a method of determining these conjugate directions. Let X_0 and X_1 be two points such that the quadratic $Q(X)$ is minimum at X_0 and X_1 along a direction η. This means

$$\frac{\partial}{\partial\alpha}Q(X_0 + \alpha\eta) = 0 \quad \text{at } \alpha = 0$$

and

$$\frac{\partial}{\partial\alpha}Q(X_1 + \alpha\eta) = 0 \quad \text{at } \alpha = 0 \tag{6.8}$$

$$\begin{aligned} Q(X_0 + \alpha\eta) &= (X_0 + \alpha\eta)^t P(X_0 + \alpha\eta) + B^t(X_0 + \alpha\eta) + c \\ &= X_0^t P X_0 + \alpha\eta^t P X_0 + \alpha X_0^t P\eta + \alpha^2\eta^t P\eta \\ &\quad + B^t(X_0 + \alpha\eta) + c \end{aligned}$$

$$\frac{\partial}{\partial\alpha}Q(X_0 + \alpha\eta) = \eta^t P X_0 + X_0^t P\eta + 2\alpha\eta^t P\eta + B^t\eta \tag{6.9}$$

and

$$\frac{\partial}{\partial\alpha}Q(X_1 + \alpha\eta) = \eta^t P X_1 + X_1^t P\eta + 2\alpha\eta^t P\eta + B^t\eta \tag{6.10}$$

Substituting $\alpha = 0$ and using equation (6.8) we obtain

$$\eta^t P(X_1 - X_0) + (X_1 - X_0)^t P\eta = 0$$

Hence

$$\eta^t P(X_1 - X_0) = 0 \qquad (6.11)$$

By equation (6.11) the direction $(X_1 - X_0)$ is P conjugate to η. Based on this theory Powell suggests an algorithm to minimize Q(X).

Let X_0 be an initial point. The initial directions of search $(\xi_i)_{i=1}^n$ are chosen as the co-ordinate directions even though they are not P conjugate directions. At the end of n searches along these directions a point X_1 is obtained. A new direction ξ_{n+1} is chosen such that

$$\xi_{n+1} = X_1 - X_0 \qquad (6.12)$$

The function is minimized along ξ_{n+1} to yield a new starting point X_0. The directions are rearranged according to the rule

$$\xi_i = \xi_{i+1}, \quad i \leqslant n \qquad (6.13)$$

The iteration is repeated until the function values does not change.

The function Q(X) will be minimized in n iterations is proved as follows. At the end of n linear searches of each iteration a new direction is chosen by using the relation (6.12). The function $Q(X)$ is minimum at X_1 and X_0 along the direction ξ_n. Hence the new direction chosen at the end of any iteration is P conjugate to ξ_n. Thus at the end of n iterations the function $Q(X)$ has been minimized along n mutually conjugate directions and so the minimum of $Q(X)$ has been obtained.

One drawback of the method is that the new directions chosen need not be linearly independent. The n direction generated need not span the n dimensional space, in which case the method will not give the minimum of $Q(X)$. Zangwill has shown with a counter example that Powell's method does not converge in the case of a particular function which is strictly convex, quadratic and has a unique minimum (48). This led him to suggest a new algorithm. This new algorithm with some modifications, is now discussed.

(2) Zangwill's Method

This method incorporates the quadratic convergence properties of Powell's method and guarantees that in the case of a quadratic function the new directions chosen are linearly independent. This is achieved by ensuring that any new direction chosen can never be zero. The algorithm, as modified to suit our problem, is as follows.

Let $(r_i)_{i=1}^n$ be the n normalized co-ordinate directions and $(\xi_i^K)_{i=1}^n$ be the n normalized conjugate directions used in the K^{th} iteration. Initial conjugate directions are chosen as the co-ordinate directions. X_0 is the initial point and $||X|| = (\sum_{i=1}^{n} x_i^2)^{\frac{1}{2}}$ is the usual Eucledian norm. It is assumed that P matrix is positive definite. (X_i^K) denotes a point obtained by minimizing along i^{th} conjugate direction in K^{th} iteration and (X_0^K) denotes the starting point of K^{th} iteration.

INITIALIZATION

Let us define X_{n+1}^0 and X_0^1 as

$$X_{n+1}^0 = X_0^1 = X_0, \text{ initial point.}$$

$$\xi_i^1 = r_i, \; i = 1,2,\ldots n \tag{6.14}$$

and $K = 1$
$j = 1$, where K and j are indeces.

ITERATION K

(a) Minimization along conjugate directions.

For i = 1,2,...,n compute α_i such that $Q(X_{i-1}^K + \alpha_i \xi_i^K)$ is minimized and define

$$X_i^K = X_{i-1}^K + \alpha_i \xi_i^K \tag{6.15}$$

At the end of the above n searches choose a new direction such that

$$\xi^K_{n+1} = (X^K_n - X^{K-1}_{n+1})/||X^K_n - X^{K-1}_{n+1}|| \qquad (6.16)$$

Find α_{n+1} so that $Q(X^K_n + \alpha_{n+1}\xi^K_{n+1})$ is minimum and define

$$X^K_{n+1} = X^K_n + \alpha_{n+1}\xi^K_{n+1} \qquad (6.17)$$

The new conjugate directions for the next iteration are chosen as

$$\xi^{K+1}_i = \xi^K_{i+1}, \ i = 1,2,\ldots,n \qquad (6.18)$$

(b) Minimization along co-ordinate directions.

Choose α to minimize $Q(X^K_{n+1} + \alpha r_j)$. If $\alpha = 0$ and $j = n$, repeat (b) with $j = 1$. If $\alpha = 0$ and $j \neq n$, repeat (b) with $j = j + 1$. If (b) is repeated n times in succession X^K_{n+1} is a minimum point. If $\alpha \neq 0$, choose

$$X^{K+1}_0 = X^K_{n+1} + \alpha r_j$$

$$j = j + 1$$

and $$K = K + 1 \qquad (6.19)$$

Parts (a) and (b) of iteration K are repeated in succession until the minimum point is obtained.

In the above algorithm the function is first minimized along the n conjugate directions and then along one of the co-ordinate directions. If the function does not change along this co-ordinate direction the next co-ordinate direction is tried. If all the n co-ordinate directions are tried in succession and there is no change in Q(X), Grad(Q) at this point is zero and hence a local minimum is obtained. The major improvement of this method over Powell's method is in the choice of the new direction ξ_{n+1}^{K} in step (a) of iteration K as given by equation (6.16). This ensures that ξ_{n+1}^{K} is always different from zero, except for iteration 1. If $\xi_{n+1}^{1} = [0]$, it means that the initial point is a minimum point and the search can be immediately terminated. For $K > 1$ we note that

$$Q(X_n^K) \leqslant Q(X_0^K) < Q(X_{n+1}^{K-1}) \tag{6.20}$$

and hence $X_n^K \neq X_{n+1}^{K-1}$. The point X_n^K is a minimum along ξ_n^K and the point X_{n+1}^{K-1} is also a minimum along the same direction. Therefore ξ_{n+1}^{K} is P conjugate to ξ_n^K. Since ξ_{n+1}^{K} and ξ_n^K are non zero vectors and the matrix P is positive definite, it follows that these directions are also linearly independent. Hence the quadratic function Q(X) can be minimized in n iterations.

It can also be proved that if the function Q(X) is strictly convex, is continuously differentiable for all X and has a greatest lower bound, the method converges to an optimal point. Consider the sequence of points $(X_{n+1}^{K})_{K=1}^{\infty}$ generated during each iteration. For convenience let us define

$$\bar{X}^{K} = X_{n+1}^{K}, \; K = 1,2,\ldots\infty.$$

The method guarantees that

$$Q(\bar{X}^{K+1}) \leqslant Q(\bar{X}^{K}) \qquad (6.21)$$

Since Q(X) has a greatest lower bound, the relation (6.21) gives

$$\lim_{K\to\infty} Q(\bar{X}^{K+1}) = \lim_{K\to\infty} Q(\bar{X}^{K}) \qquad (6.22)$$

Let us define

$$\lim_{K\to\infty} Q(\bar{X}^{K+1}) = Q(\bar{X}^{\infty+1})$$

and

$$\lim_{K\to\infty} Q(\bar{X}^{K}) = Q(\bar{X}^{\infty}).$$

We are always selecting $\bar{X}^{K+1}$ such that

$$\bar{X}^{K+1} = \bar{X}^{K} + \alpha_K \eta^{K} \qquad (6.23)$$

where η^{K} is some conjugate direction and $Q(\bar{X}^{K} + \alpha_K\eta^{K})$ is minimum along η^{K}. Since Q(X) is continuously differentiable we have in the limit

$$Q(\bar{X}^{\infty+1}) = Q(\bar{X}^{\infty} + \alpha_{\infty}\eta^{\infty}) \leqslant Q(\bar{X}^{\infty} + \beta\eta^{\infty}) \qquad (6.24)$$

where β is any scalar.

Since $Q(X)$ is strictly convex there should be at least one point between $\bar{X}^{\infty}$ and $\bar{X}^{\infty+1}$, along η, such that at this point $\bar{X}$, $Q(x)$ is less than $Q(\bar{X}^{\infty})$ or $Q(\bar{X}^{\infty+1})$. As this is not possible,

$$\bar{X}^{\infty+1} = \bar{X}^{\infty} \qquad (6.25)$$

The point $\bar{X}^{\infty+1}$ is arrived at, after competing step (b) of iteration K of the algorithm. Hence

$$Q(\bar{X}^{\infty+1}) \leqslant Q(\bar{X}^{\infty} + \beta r_i),\ i = 1,2,\ldots,n \qquad (6.26)$$

Equation (6.26) ensures that the final point $\bar{X}^{\infty}$ is an optimum point.

The original method of Zangwill requires the availability of n normalized directions different from co-ordinate directions to start with (48). The method as given here does not require this. The algorithm has been suitably modified so that the co-ordinate directions alone are sufficient to start with.

A digital computer program ZAGMIN is written in FORTRAN IV which will minimize any function of several variables using the above method. It is found that the criteria for terminating the minimization as suggested in the algorithm is too strong. Hence the program is also made to terminate when the change in function value in successive iterations is less than a preassigned value. The initial step of the linear search is to be specified. This step is doubled or halved at each search point until the minimum is passed. Three points at equal intervals between which the minimum lies are found and the minimum point is computed by quadratic interpolation (47, 51). If the function value at this point is not less than that at the middle point, the computed point is ignored and the middle point is taken as the minimum.

The ISE, I, is considered as a function of the variables $(a_i)_{i=1}^N$ and is minimized for different functions using the above program. It is found that the new method is more efficient than the original method of Powell in minimizing the ISE. The values of the parameters $(a_i)_{i=1}^N$ at the minimum ISE give the best pole positions. These poles are used throughout the subsequent analysis and design.

Once the poles are determined, the orthonormal functions $\Phi_k(s)$, $k = 1,2,...N$ can be found. The order of the poles is not important in forming these orthonormal functions. In the examples considered the real poles are taken first, followed by the complex pole pairs. A digital computer program for computing the numerator and denominator coefficients of the orthonormal functions $\Phi_k(s)$, $k = 1,2,...N$, is given in Appendix I. The inverse of these functions $\phi_k(t)$, $k = 1,2,...N$ are computed using the methods discussed in Chapter 5. The subroutine SERIES takes in the numerator and denominator coefficients of any rational function $\Phi_k(s)$ and computes the coefficients of the series expansion of $\phi_k(t)$ in powers of t. The maximum value of t at which $\phi_k(t)$ is required is also given as an input to SERIES. If this maximum value $t_\ell > 1$, the series expansion is made with respect to a new variable τ such that

$$\tau = t/t_\ell \qquad (6.27)$$

The total number of terms necessary can be calculated by using the relation (5.79). Another method of determining the number of terms is to make use of the condition that the absolute value of n^{th} term of the series expansion tends to zero as n tends to infinity. As all computations are carried out in double precision arithmetic the computation of the coefficients is terminated when the terms become less than 10^{-16}. The relation (5.79) is used as a general guide to find the total number of terms required.

The function subprogram VAL uses these coefficients and maximum value of t to compute the inverse at any point t. The subroutine SERIES and function subprogram VAL are given in Appendix I.

MINIMIZATION OF AVERAGE ERROR

It has been proved in Chapter 4 that the frequency response of H(s) can be improved by minimizing the average error. Hence the approximation

$$h(t) = \sum_{i=1}^{N} C_i\phi_i(t) \tag{6.28}$$

is to be so chosen such that the average error γ is minimized subject to the condition that ISE is less than a preassigned value μ. This can be achieved if the minimization of γ is done such that $(C_i)_{i=1}^{N}$ are always chosen to satisfy equation (4.22). This equation is

$$\sum_{i=1}^{N} (C_i - \tilde{C}_i)^2 \leqslant (\mu - \tilde{I}) = \delta^2 \tag{6.29}$$

where

$$\tilde{C}_i = \int_0^\infty f(t)\phi_i(t)dt$$

and $\tilde{I}$ is the minimum ISE.

This constraint on the choice of $(C_i)_{i=1}^{N}$ can be incorporated into the expression of average error by the penalty function approach (51). The expression for the average error is modified as follows

$$\gamma_n = \int_0^\infty |f(t) - \sum_{i=1}^{N} C_i\phi_i(t)|dt + w\rho|\sum_{i=1}^{N}(C_i - \tilde{C}_i)^2 - \delta^2|^2 \tag{6.30}$$

where w is a weight factor and ρ is a constant defined as

$$\rho = 0 \text{ when } \sum_{i=1}^{N}(C_i - \tilde{C}_i)^2 \leqslant \delta^2$$

$$\text{and} \qquad \rho = 1 \text{ when } \sum_{i=1}^{N}(C_i - \tilde{C}_i)^2 > \delta^2 \tag{6.31}$$

The value of the weight factor w may be changed as desired (51). The choice of ρ as in equation (6.31) ensures that when $(C_i)_{i=1}^{N}$ is inside or on the sphere defined by equation (6.29), γ_n and γ are the same and whenever the search point goes outside this sphere, γ_n is made greater than γ. The actual value of γ_n is controlled by a proper choice of w. It has been proved in Chapter 5, that γ has a minimum value in the sphere defined by equation (6.29). Hence as the minimization of γ_n is continued the points of search approach this sphere and finally converge to a point within or on this sphere.

Any of the gradient techniques may be used to minimize γ_n.

$$\text{Grad }(\gamma_n) = \text{Grad }(\gamma) + w\rho\text{Grad}\left|\sum_{i=1}^{N}(C_i - \tilde{C}_i)^2 - \delta^2\right|^2 \tag{6.32}$$

The k th component of Grad (γ_n) is given as

$$\frac{\partial\gamma_n}{\partial C_k} = \frac{\partial\gamma}{\partial C_k} + 4w\rho(C_k - \tilde{C}_k)\left\{\sum_{i=1}^{N}(C_i - \tilde{C}_i)^2 - \delta^2\right\} \tag{6.33}$$

The term $\frac{\partial\gamma}{\partial C_k}$ is computed as discussed in Chapter 4. This is given as,

$$\frac{\partial\gamma}{\partial C_k} = \int_0^{T_0}\text{Sgn}\{e(t)\}\{-\phi_i(t)\}dt \tag{6.34}$$

where $e(t) = f(t) - \sum_{i=1}^{N} C_i\phi_i(t)$

The interval $[0, T_0]$ is divided into small intervals and the integral in each interval is evaluated using Gauss quadrature formula. The error involved in evaluation of the gradient due to the discontinuity of Sgn(e(t)) can be made negligible by choosing the subinterval to be very small. This avoids the determination of the zero crossings of e(t) for each set of coefficients $(C_i)_{i=1}^{N}$. Similarly the time of computation of average error and gradient is considerably saved by computing and storing the values of the functions $\phi_i(t)$, i = 1,2,...,N at those points as required by Gauss quadrature formula. This need be done only once at the beginning of the minimization. The subroutine MDQG makes use of these values and computes γ_n and Grad (γ_n) for any point $(C_i)_{i=1}^{N}$. In the following numerical examples the actual minimization of average error was carried out by using Fletcher-Powell algorithm (52).

The subroutine MDQG and a program for minimizing the average error are given in Appendix 1. The minimization of average error makes use of the standard Fletcher-Powell algorithm available in the IBM Scientific Subroutine Package Library.

NUMERICAL EXAMPLES

The method developed in this dissertation was used to approximate the ideal low pass filter. The reason for choosing this example is that it can be made use of to check the comparative merits and demerits of the new method.

The frequency response of an ideal low pass is defined as

$$\begin{aligned} F(j\omega) &= e^{-j\omega t_0}, \quad |\omega| < 1 \\ &= 0 \qquad , \quad |\omega| > 1 \end{aligned} \tag{6.35}$$

The corresponding time function f(t) is given as

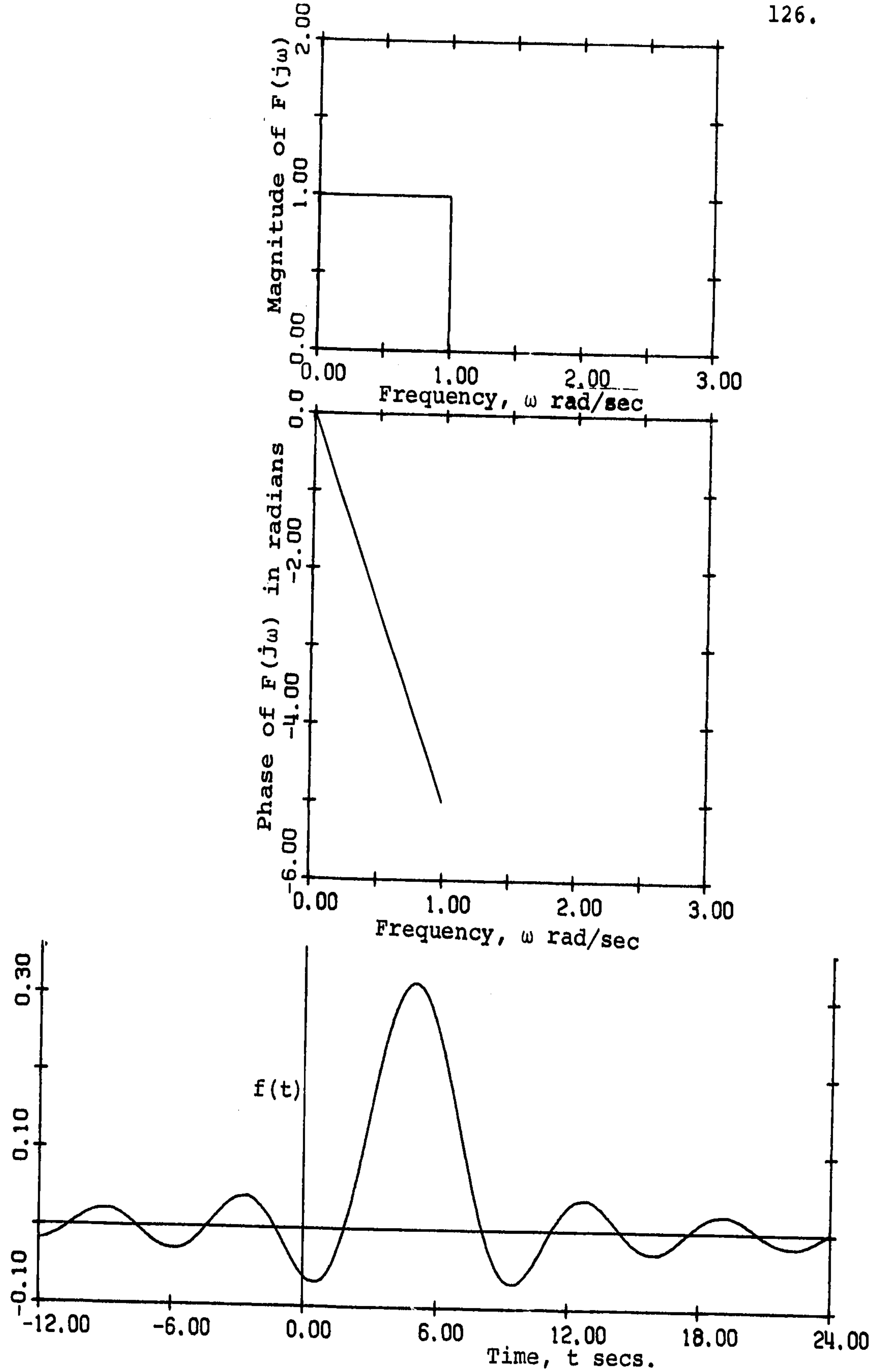


Figure 6.1. Frequency and Impulse responses of ideal low-pass filter. Delay=5 secs.

$$f(t) = \frac{1}{\pi} \frac{\sin(t - t_0)}{(t - t_0)} \tag{6.36}$$

f(t) is obtained by delaying the sint/t function by t_0.

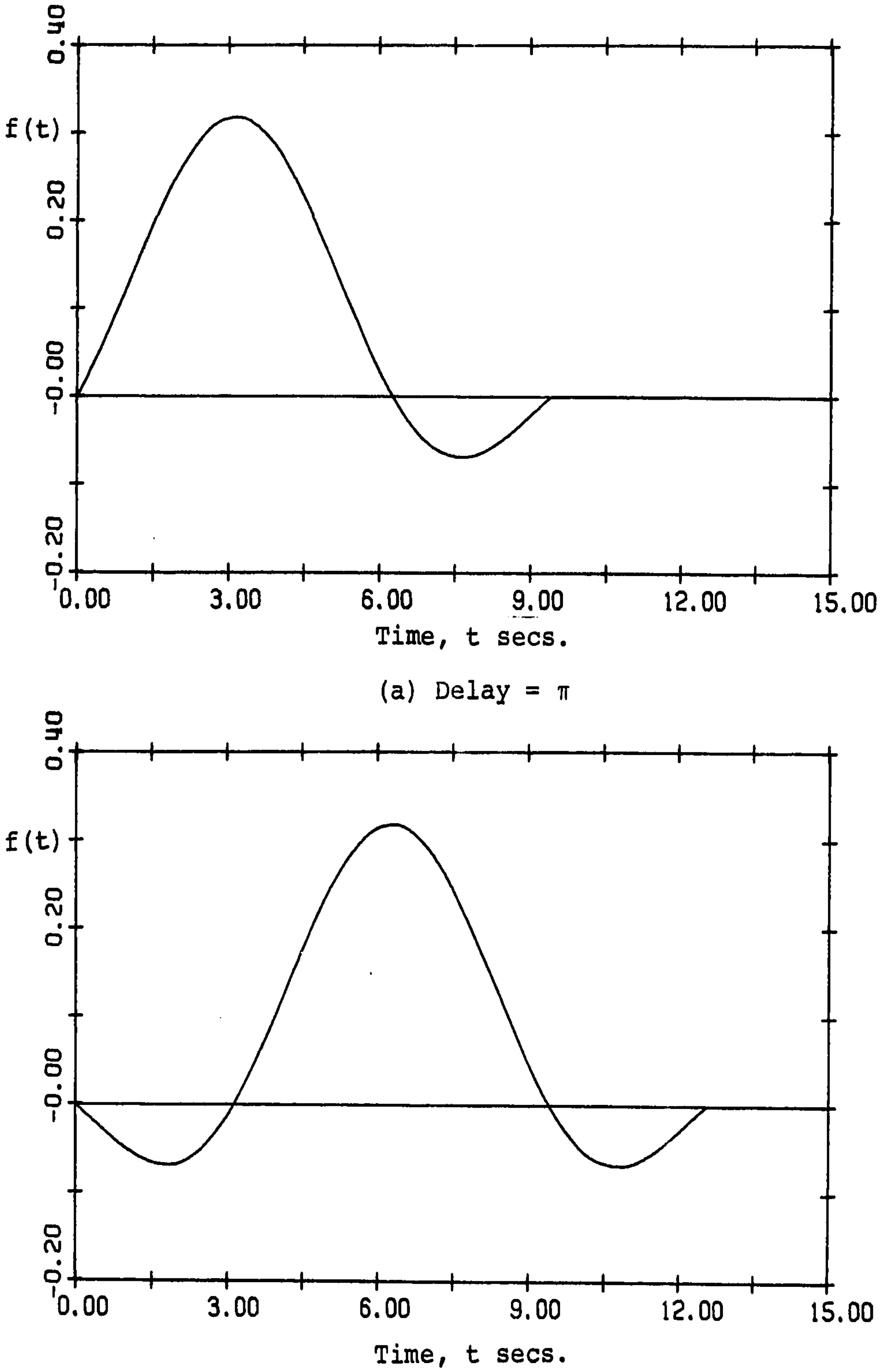
These functions are shown in Fig. 6.1.

Theoretically F(jω) is not realizable. This is because f(t) exists for negative values of t. The function f(t) is not absolutely integrable. Hence f(t) does not satisfy the assumptions made in Chapter 4. But even though F(jω) is not physically realizable, it is possible to obtain a satisfactory approximation of F(jω) by suitably truncating f(t). Two different versions of f(t) are used to approximate F(jω). They are

$$\begin{aligned} 1) \quad f(t) &= \frac{1}{\pi}\frac{\sin(t - \pi)}{(t - \pi)}, \quad t\varepsilon[0,\ 3\pi] \\ &= 0 \qquad\qquad\quad, \quad t\not\varepsilon[0,\ 3\pi] \end{aligned} \tag{6.37}$$

$$\begin{aligned} 2) \quad f(t) &= \frac{1}{\pi}\frac{\sin(t - 2\pi)}{(t - 2\pi)}, \quad t\varepsilon[0,\ 4\pi] \\ &= 0 \qquad\qquad\quad, \quad t\not\varepsilon[0,\ 4\pi] \end{aligned} \tag{6.38}$$

f(t) as defined by equations (6.37) and (6.38) are physically realizable. These functions are shown in Fig. 6.2a and Fig. 6.2b respectively. Each of these functions is approximated by 5th order and 8th order filters. These approximations are now given.



(a) Delay = π

(b) Delay = 2π

Figure 6.2. Truncated Impulse responses of ideal lowpass filter.

The following notations are used in the results.

$\tilde{H}(s)$ - Least square filter.

$H(s)$ - Filter obtained by minimizing average error γ with constraints on ISE.

$(\tilde{b}_i)_{i=1}^{N}$ - The numerator parameters of $\tilde{H}(s)$.

$(b_i)_{i=1}^{N}$ - The numerator parameters of $H(s)$

$(a_i)_{i=1}^{N}$ - The denominator parameters of $\tilde{H}(s)$ and $H(s)$.

$(\tilde{C}_i)$ - The coefficients of orthonormal functions at minimum ISE.

$(C_i)_{i=1}^{N}$ - The coefficients of orthonormal functions at minimum average error with constraints on ISE.

$(s_i)_{i=1}^{N}$ - The poles of $H(s)$ and $\tilde{H}(s)$.

$\tilde{h}(t)$ and $h(t)$ are the time functions of $\tilde{H}(s)$ and $H(s)$ respectively.

$\tilde{e}(t)$ and $e(t)$ are the corresponding error functions.

The orthonormal functions for each case are given in Appendix II.

Example 1

$$f(t) = \frac{1}{\pi} \frac{\sin(t - \pi)}{(t - \pi)} , \; t\varepsilon[0, 3\pi]$$

The function f(t) is shown in Fig. 6.2a.

a) Order of filter, N = 5

$$\tilde{H}(s) = \frac{\sum_{i=1}^{5} \tilde{b}_i s^{5-i}}{s^5 + \sum_{i=1}^{5} a_i s^{5-i}} \text{, Minimum ISE} = 0.00021$$

$$H(s) = \frac{\sum_{i=1}^{5} b_i s^{5-i}}{s^5 + \sum_{i=1}^{5} a_i s^{5-i}} \text{, Minimum } \gamma = 0.06163, \quad \text{ISE} \leqslant 0.0005$$

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} 2.04708 \\ 3.36562 \\ 3.27161 \\ 1.92788 \\ 0.6517 \end{bmatrix} ; \quad \begin{matrix} s_1 = -0.79076 \\ s_2, s_3 = -0.39496 \pm j\ 0.64967 \\ s_4, s_5 = -0.2332 \pm j\ 1.17103 \end{matrix}$$

$$\begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \tilde{b}_3 \\ \tilde{b}_4 \\ \tilde{b}_5 \end{bmatrix} = \begin{bmatrix} 0.00355 \\ 0.07107 \\ 0.37711 \\ 0.21931 \\ 0.68641 \end{bmatrix} ; \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} = \begin{bmatrix} 0.00807 \\ 0.06382 \\ 0.38919 \\ 0.21292 \\ 0.68873 \end{bmatrix}$$

$$\begin{bmatrix} \tilde{c}_1 \\ \tilde{c}_2 \\ \tilde{c}_3 \\ \tilde{c}_4 \\ \tilde{c}_5 \end{bmatrix} = \begin{bmatrix} 0.20352 \\ -0.16866 \\ -0.4689 \\ -0.04183 \\ 0.05637 \end{bmatrix} ; \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix} = \begin{bmatrix} 0.20407 \\ -0.16752 \\ -0.46892 \\ -0.03934 \\ 0.05305 \end{bmatrix}$$

The frequency responses of $\tilde{H}(j\omega)$ and $H(j\omega)$ are given in Fig. 6.3 and Fig. 6.4 respectively. The least square approximation $\tilde{h}(t)$ and the error function $\tilde{e}(t)$ are given in Fig. 6.5. Fig. 6.6 gives the function $h(t)$ and the corresponding error function $e(t)$ for minimum γ with constraints on ISE.

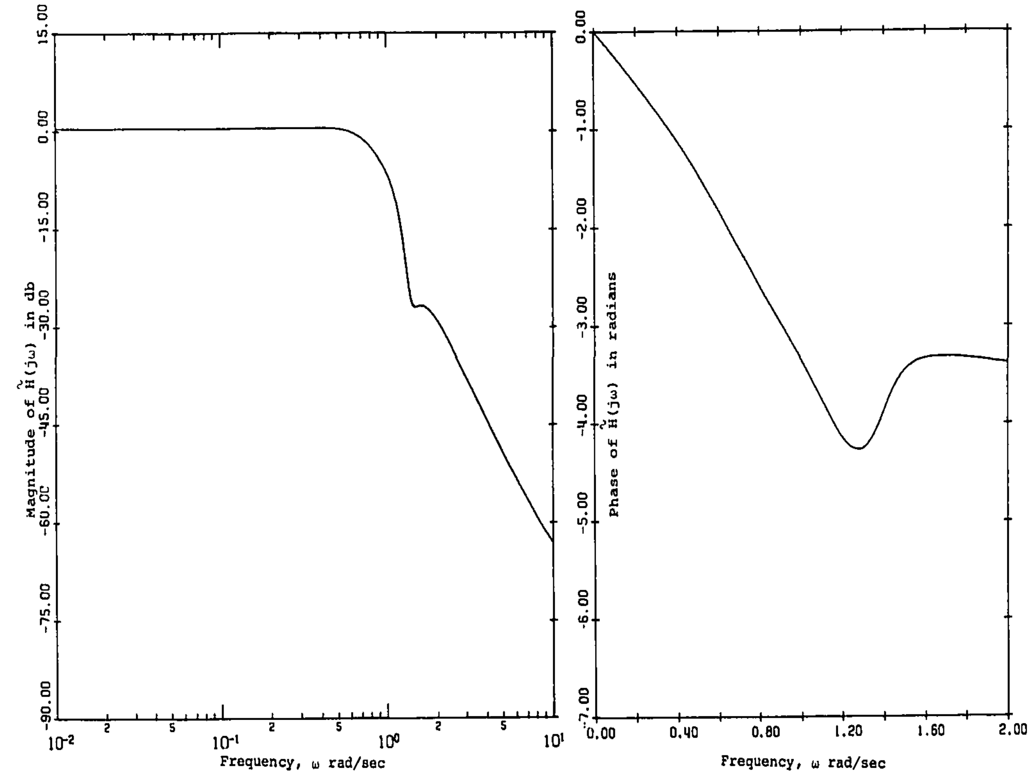


Fig. 6.3 Magnitude and Phase responses of $\tilde{H}(j\omega)$, N=5, Delay=π, Minimum ISE=0.00021.

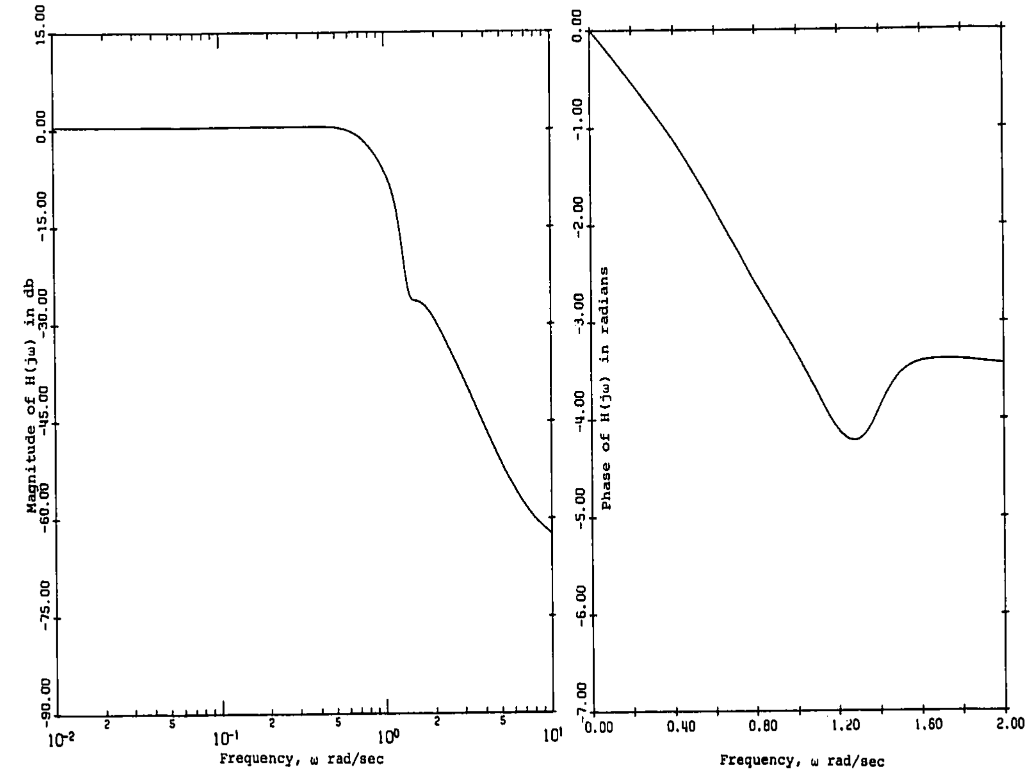


Fig. 6.4 Magnitude and Phase responses of H(jω), N=5, Delay=π, Minimum γ=0.06163, ISE≤0.0005.

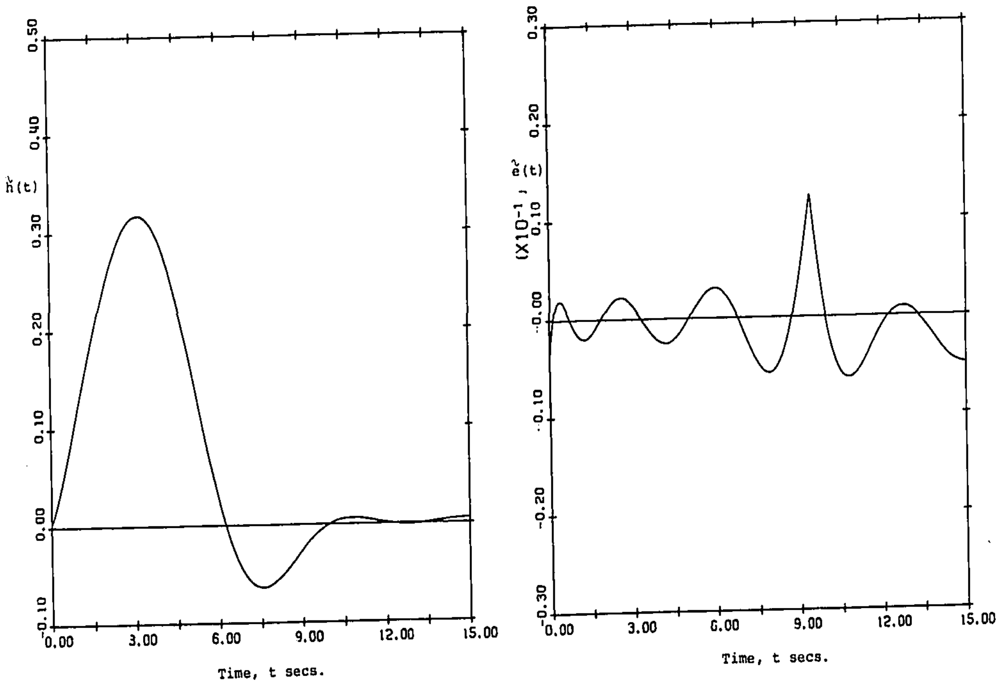


Figure 6.5. The least square approximation $\tilde{h}(t)$ and error $\tilde{e}(t)$. N=5, Delay=π, Minimum ISE=0.00021

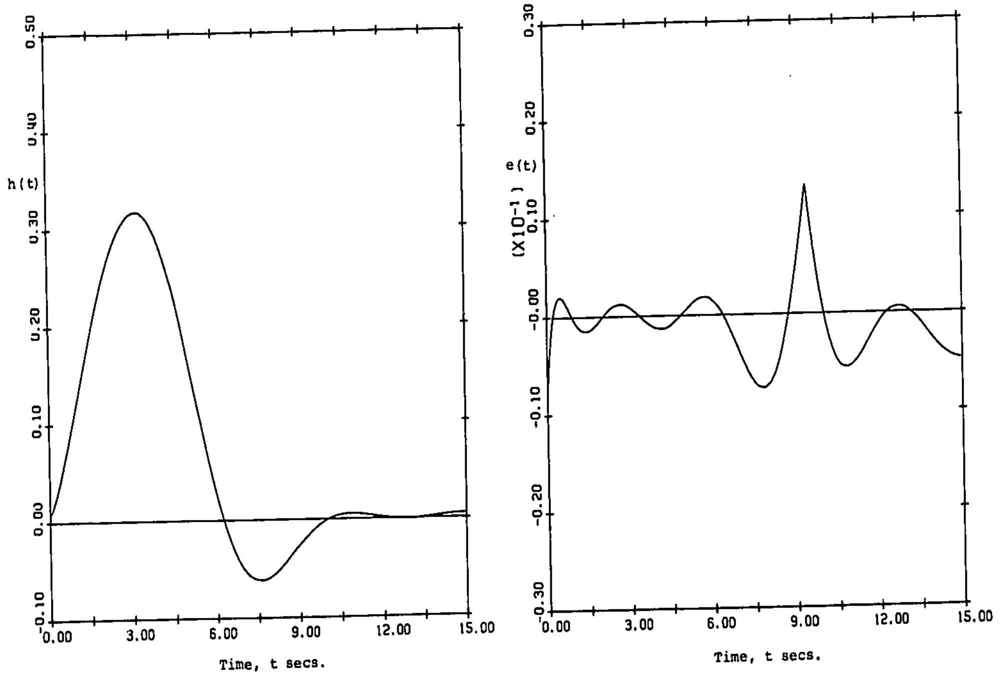


Figure 6.6. The approximation h(t) and error e(t) for minimum average error. N=5, Delay=π, Minimum γ=0.06163, ISE$\leq$0.0005

b) Order of filter, N = 8

$$\tilde{H}(s) = \frac{\sum_{i=1}^{8} \tilde{b}_i s^{8-i}}{s^8 + \sum_{i=1}^{8} a_i s^{8-i}} \; ; \text{ Minimum ISE} = 0.000057$$

$$H(s) = \frac{\sum_{i=1}^{8} b_i s^{8-i}}{s^8 + \sum_{i=1}^{8} a_i s^{8-i}} \; ; \begin{array}{l} \text{Minimum } \gamma = 0.02058 \\ \quad\text{ISE} \leqslant 0.0005 \end{array}$$

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{bmatrix} = \begin{bmatrix} 4.20363 \\ 10.54344 \\ 20.47894 \\ 24.75407 \\ 23.68999 \\ 12.88635 \\ 4.44377 \\ 0.04039 \end{bmatrix} ; \quad \begin{array}{rl} s_1 &= -0.00934 \\ s_2 &= -2.10856 \\ s_3, s_4 &= -0.37017 \pm j\,1.09699 \\ s_5, s_6 &= -0.2006 \pm j\,1.67774 \\ s_7, s_8 &= -0.47209 \pm j\,0.55961 \end{array}$$

$$\begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \tilde{b}_3 \\ \tilde{b}_4 \\ \tilde{b}_5 \\ \tilde{b}_6 \\ \tilde{b}_7 \\ \tilde{b}_8 \end{bmatrix} = \begin{bmatrix} -0.01818 \\ 0.08441 \\ 0.60015 \\ 0.90200 \\ 3.49263 \\ 1.63849 \\ 4.53701 \\ 0.04189 \end{bmatrix} ; \quad \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_8 \end{bmatrix} = \begin{bmatrix} -0.01526 \\ 0.07433 \\ 0.62983 \\ 0.85798 \\ 3.54706 \\ 1.60465 \\ 4.55149 \\ 0.03984 \end{bmatrix}$$

$$\begin{bmatrix} \tilde{c}_1 \\ \tilde{c}_2 \\ \tilde{c}_3 \\ \tilde{c}_4 \\ \tilde{c}_5 \\ \tilde{c}_6 \\ \tilde{c}_7 \\ \tilde{c}_8 \end{bmatrix} = \begin{bmatrix} 0.13891 \\ 0.03703 \\ 0.1587 \\ -0.26183 \\ 0.01719 \\ -0.18327 \\ -0.2203 \\ -0.31303 \end{bmatrix} ; \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \\ c_8 \end{bmatrix} = \begin{bmatrix} 0.13577 \\ 0.03763 \\ 0.15864 \\ -0.26248 \\ 0.01604 \\ -0.18415 \\ -0.21827 \\ -0.31454 \end{bmatrix}$$

The frequency responses of $\tilde{H}(j\omega)$ and $H(j\omega)$ are given in Fig. 6.7 and Fig. 6.8 respectively. The least square approximation $\tilde{h}(t)$ and the error function $\tilde{e}(t)$ are given in Fig. 6.9. Fig. 6.10 gives the function $h(t)$ and the corresponding error function $e(t)$ for minimum γ, with constraint on ISE.

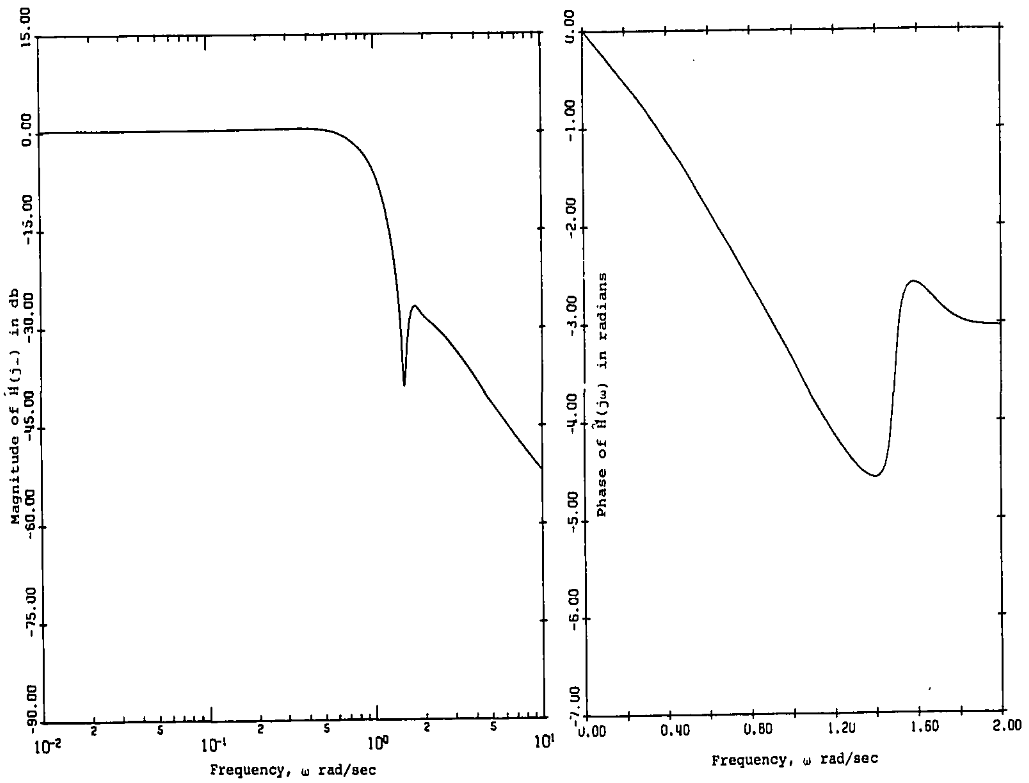


Figure 6.7 Magnitude and Phase responses of $\tilde{H}(j\omega)$, N=8, Delay=π, Minimum ISE=0.000057.

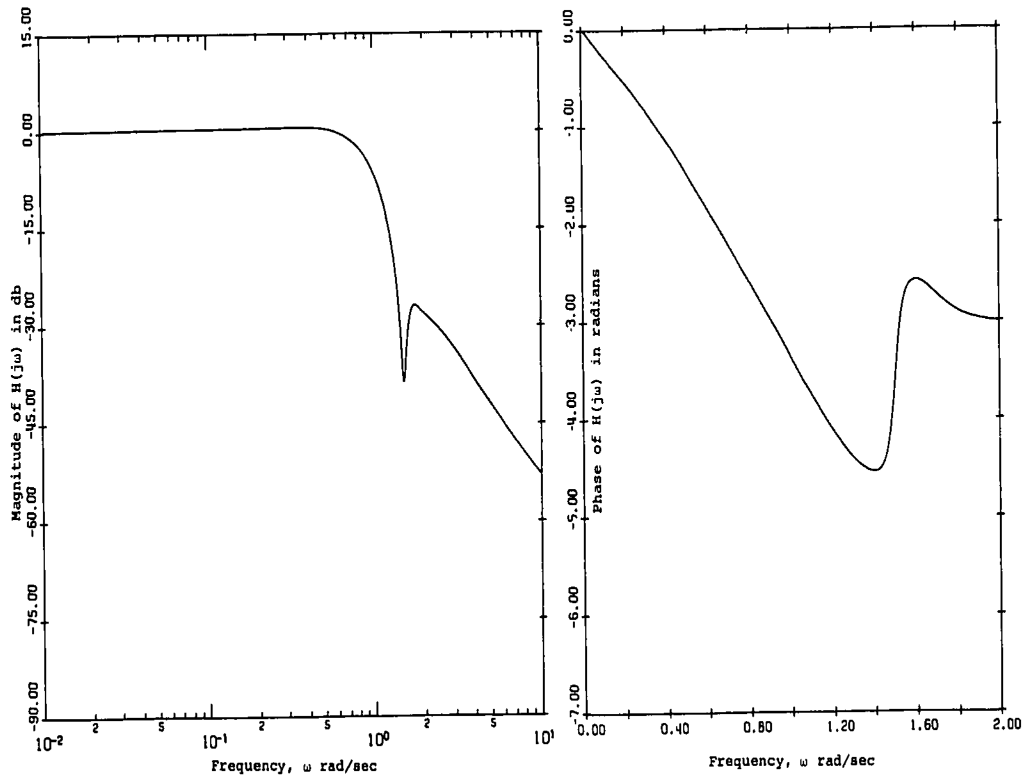


Figure 6.8 Magnitude and Phase responses of $H(j\omega)$, N=8, Delay=π, Minimum γ=0.02058, ISE≤0.0005.

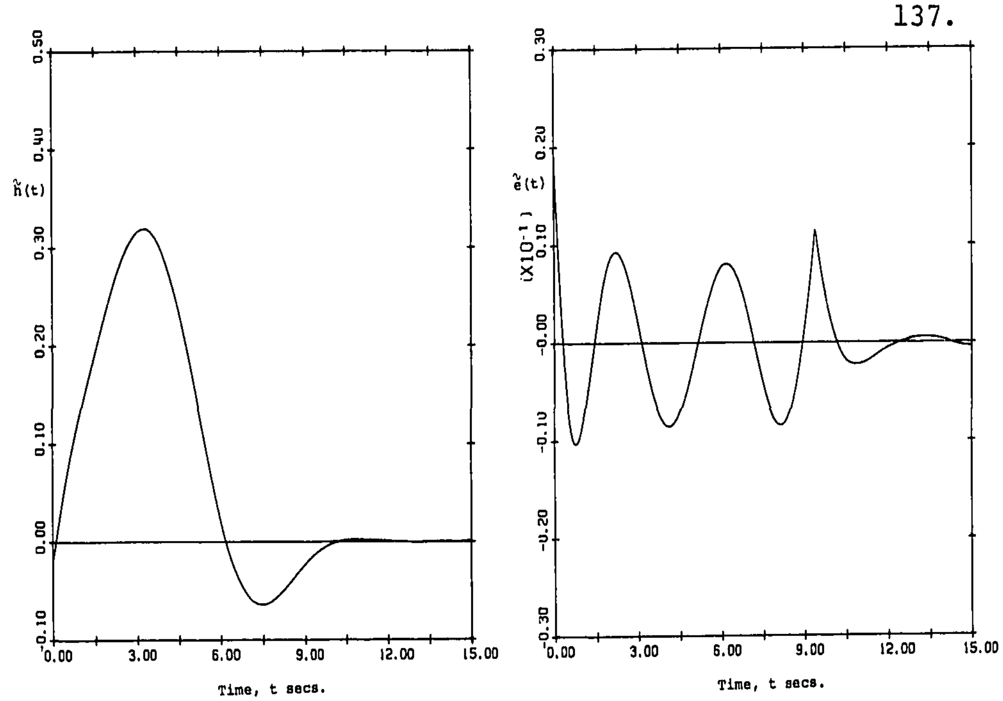


Figure 6.9. The least square approximation $\tilde{h}(t)$ and error $\tilde{e}(t)$. N=8, Delay=π, Minimum ISE=0.000057

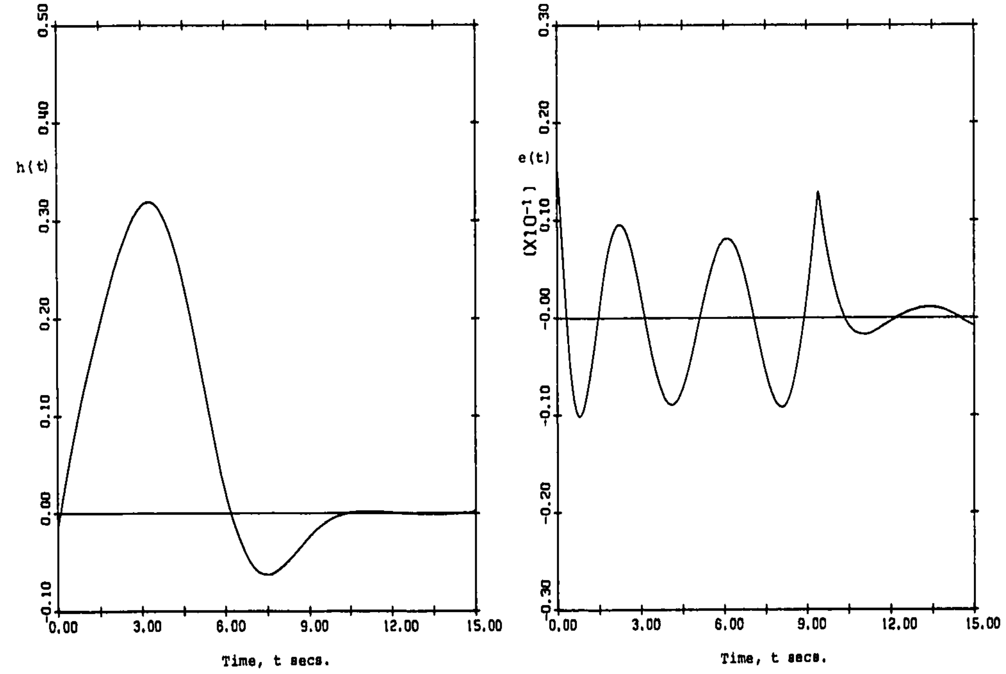


Figure 6.10. Thu approximation h(t) and error e(t) for minimum average error. N=8, Delay=π, Minimum γ=0.02058, ISE≤0.0005

Example 2

$$f(t) = \frac{1}{\pi}\,\frac{\sin(t - 2\pi)}{(t - 2\pi)}\ ,\ t\varepsilon[0,\ 4\pi]$$

The function f(t) is given in Fig. 6.2b.

a) Order of filter, N = 5.

$$\tilde{H}(s) = \frac{\sum\limits_{i=1}^{5} b_i s^{5-i}}{s^5 + \sum\limits_{i=1}^{5} a_i s^{5-i}}\ ;\ \text{Minimum ISE} = 0.00077$$

$$H(s) = \frac{\sum\limits_{i=1}^{5} b_i s^{5-i}}{s^5 + \sum\limits_{i=1}^{5} a_i s^{5-i}}\ ;\ \begin{array}{l}\text{Minimum } \gamma = 0.25125\\ \qquad \text{ISE} \leqslant 0.035\end{array}$$

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_4 \end{bmatrix} = \begin{bmatrix} 1.44778 \\ 1.99895 \\ 1.39288 \\ 0.65976 \\ 0.15055 \end{bmatrix}\ ;\ \begin{array}{r l} s_1 &= -0.43095 \\ s_2,\ s_3 &= -0.30148 \pm j\ 0.53016 \\ s_4,\ s_5 &= -0.20693 \pm j\ 0.94678 \end{array}$$

$$\begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \tilde{b}_3 \\ \tilde{b}_4 \\ \tilde{b}_5 \end{bmatrix} = \begin{bmatrix} 0.01695 \\ -0.12350 \\ 0.06757 \\ -0.28401 \\ 0.13778 \end{bmatrix}\ ;\ \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} = \begin{bmatrix} 0.01308 \\ -0.06371 \\ 0.05126 \\ -0.17422 \\ 0.13778 \end{bmatrix}$$

$$\begin{bmatrix} \tilde{C}_1 \\ \tilde{C}_2 \\ \tilde{C}_3 \\ \tilde{C}_4 \\ \tilde{C}_5 \end{bmatrix} = \begin{bmatrix} 0.01884 \\ 0.17006 \\ -0.47941 \\ -0.20586 \\ -0.01208 \end{bmatrix} ; \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \end{bmatrix} = \begin{bmatrix} 0.02329 \\ 0.10712 \\ -0.31375 \\ -0.13867 \\ -0.00454 \end{bmatrix}$$

The frequency responses of $\tilde{H}(j\omega)$ and $H(j\omega)$ are given in Fig. 6.11 and Fig. 6.12 respectively. The least square approximation $\tilde{h}(t)$ and the error function $\tilde{e}(t)$ are given in Fig. 6.13. Fig. 6.14 gives the function $h(t)$ and the corresponding error function $e(t)$ for minimum γ, with constraints on ISE.

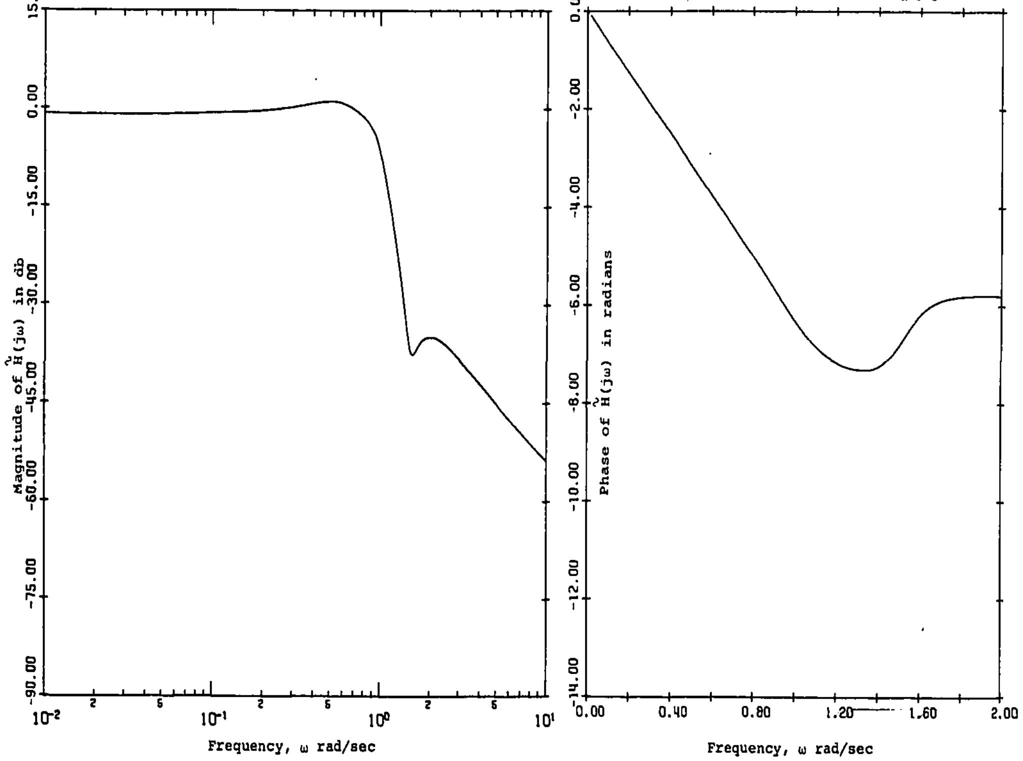


Figure 6.11 Magnitude and Phase responses of $\tilde{H}(j\omega)$, N=5, Delay=2π, Minimum ISE=0.00077.

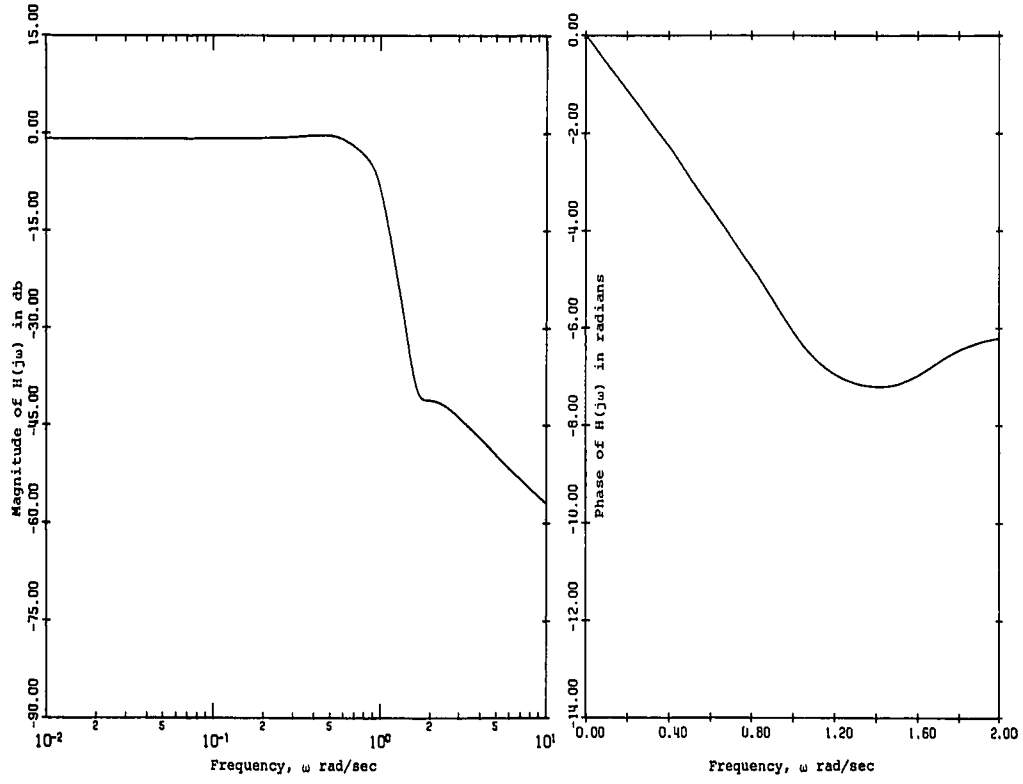


Figure 6.12 Magnitude and Phase responses of H(jω), N=5, Delay=2π, Minimum γ=0.25125, ISE$\leq$0.035.

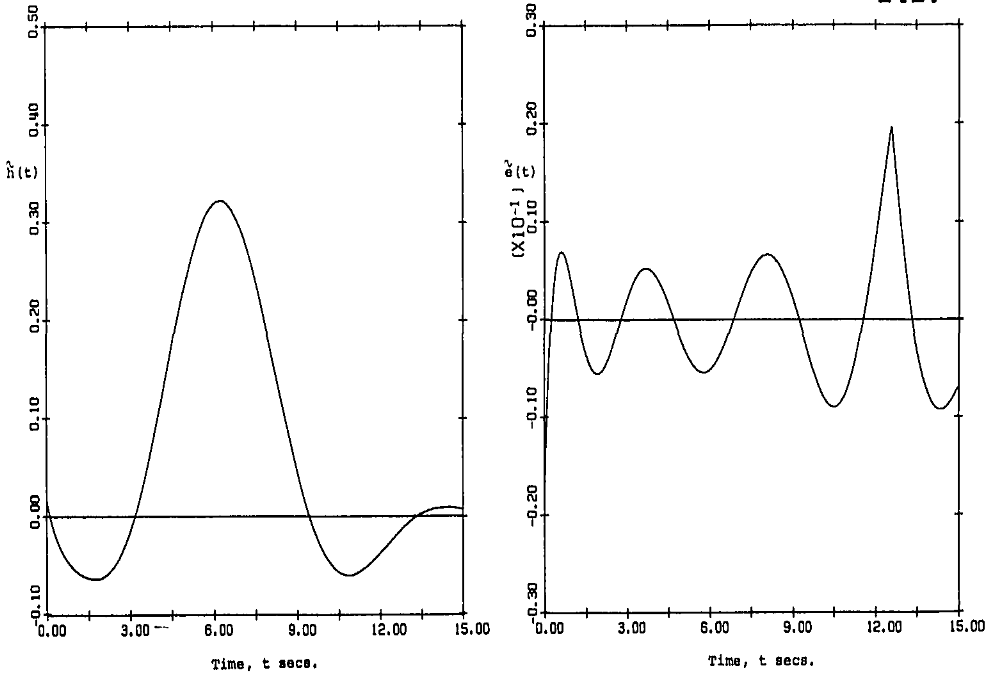


Figure 6.13. The least square approximation $\tilde{h}(t)$ and error $\tilde{e}(t)$. N=5, Delay=2π, Minimum ISE=0.00077

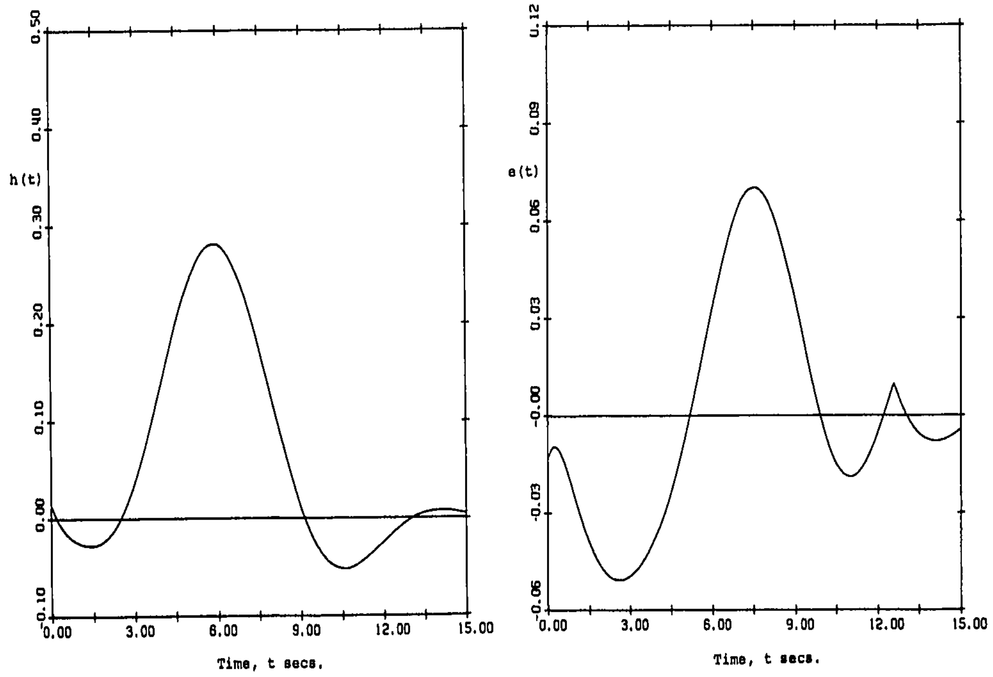


Figure 6.14. The approximation h(t) and error e(t) for minimum average error. N=5, Delay=2π, Minimum γ=0.25125, ISE<0.035

b) Order of filter, $N = 8$

$$\tilde{H}(s) = \frac{\sum_{i=1}^{8} \tilde{b}_i s^{8-i}}{s^8 + \sum_{i=1}^{8} a_i s^{8-i}} \text{ ; Minimum ISE} = 0.00052$$

$$H(s) = \frac{\sum_{i=1}^{8} b_i s^{8-i}}{s^8 + \sum_{i=1}^{8} a_i s^{8-i}} \text{ ; Minimum } \gamma = 0.232 \qquad \text{ISE} \leqslant 0.03$$

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{bmatrix} = \begin{bmatrix} 4.15279 \\ 16.39557 \\ 22.5592 \\ 28.67467 \\ 20.43117 \\ 11.10009 \\ 3.28816 \\ 0.44341 \end{bmatrix}; \quad \begin{aligned} s_1, s_2 &= -0.26537 \pm j\,0.15885 \\ s_3, s_4 &= -0.27231 \pm j\,0.60108 \\ s_5, s_6 &= -0.18275 \pm j\,1.00335 \\ s_7, s_8 &= -1.35596 \pm j\,2.8976 \end{aligned}$$

$$\begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \tilde{b}_3 \\ \tilde{b}_4 \\ \tilde{b}_5 \\ \tilde{b}_6 \\ \tilde{b}_7 \\ \tilde{b}_8 \end{bmatrix} = \begin{bmatrix} 0.0232 \\ -0.08829 \\ 0.00813 \\ -1.49849 \\ -0.03038 \\ -2.33046 \\ 0.54567 \\ 0.39488 \end{bmatrix} ; \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_8 \end{bmatrix} = \begin{bmatrix} -0.00352 \\ -0.02243 \\ -0.09383 \\ -0.5897 \\ -0.10756 \\ -1.28322 \\ 0.36157 \\ 0.39488 \end{bmatrix}$$

$$\begin{bmatrix} \tilde{c}_1 \\ \tilde{c}_2 \\ \tilde{c}_3 \\ \tilde{c}_4 \\ \tilde{c}_5 \\ \tilde{c}_6 \\ \tilde{c}_7 \\ \tilde{c}_8 \end{bmatrix} = \begin{bmatrix} -0.21314 \\ 0.3121 \\ 0.35736 \\ -0.04933 \\ -0.14889 \\ -0.08089 \\ -0.00124 \\ -0.0045 \end{bmatrix} ; \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \\ c_8 \end{bmatrix} = \begin{bmatrix} -0.1409 \\ 0.22583 \\ 0.23957 \\ -0.02398 \\ -0.1167 \\ -0.04878 \\ -0.0037 \\ 0.00062 \end{bmatrix}$$

The frequency responses of $\tilde{H}(j\omega)$ and $H(j\omega)$ are given in Fig. 6.15 and Fig. 6.16 respectively. The least square approximation $\tilde{h}(t)$ and the error function $\tilde{e}(t)$ are given in Fig. 6.17. Fig. 6.18 gives the function $h(t)$ and the corresponding error function $e(t)$ for minimum γ, with constraints on ISE.

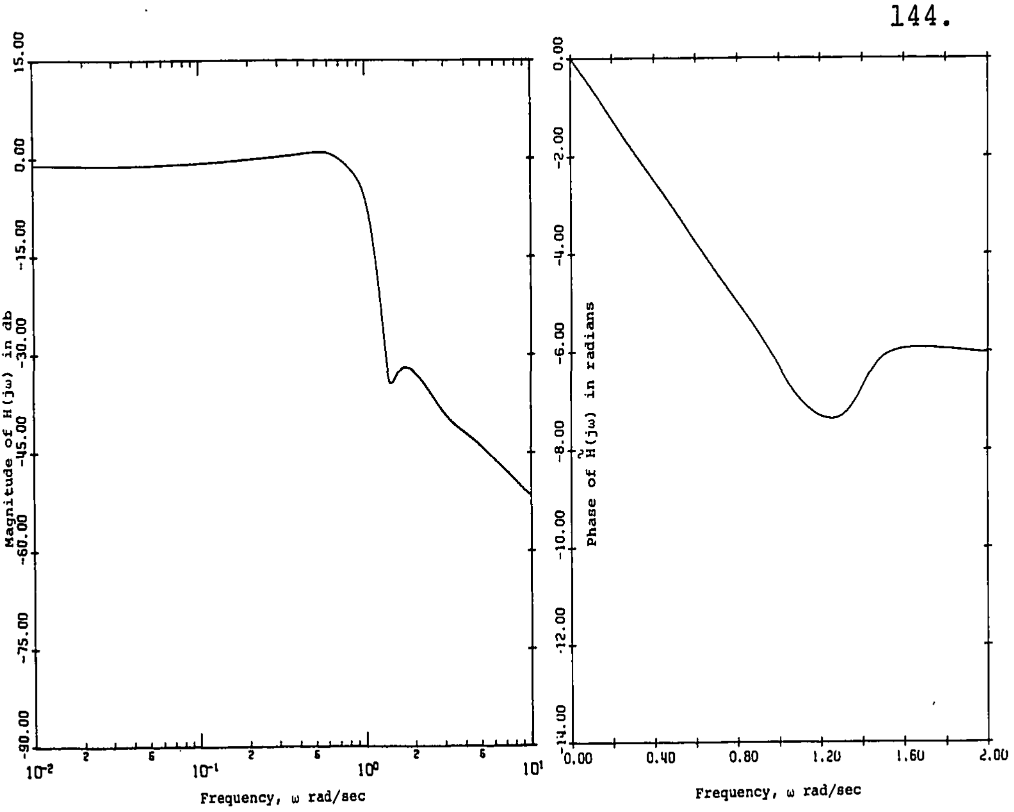


Figure 6.15 Magnitude and Phase responses of $\tilde{H}(j\omega)$, N=8, Delay=2π, Minimum ISE=0.00052.

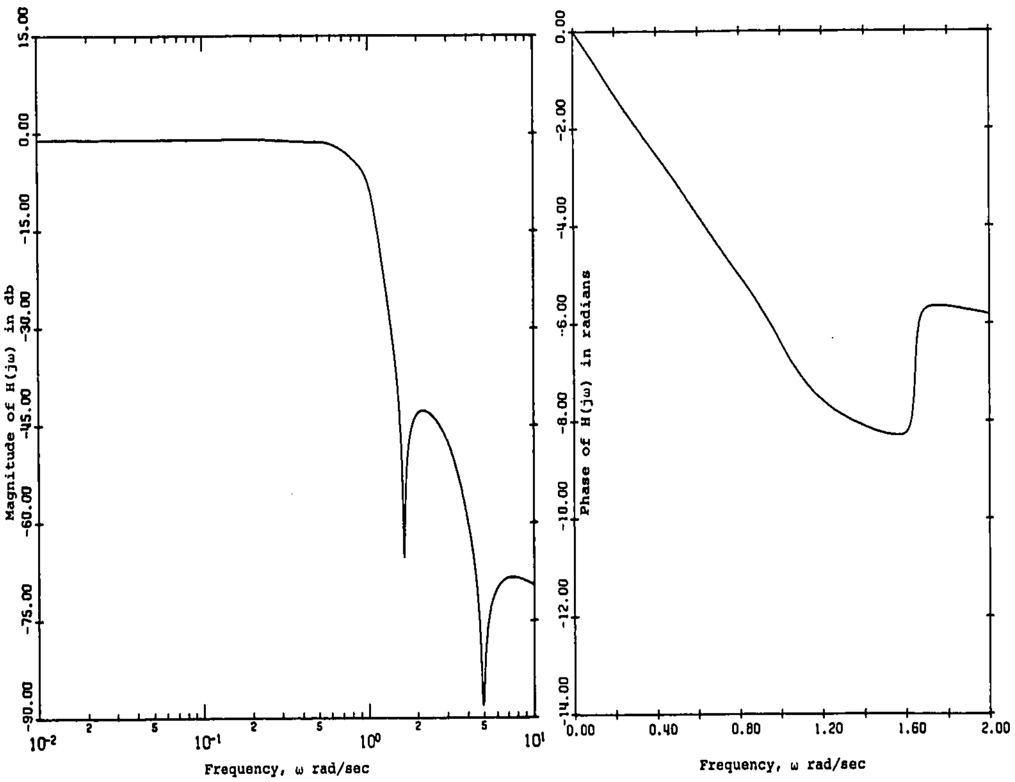


Figure 6.16 Magnitude and Phase responses of $H(j\omega)$, N=8, Delay=2π, Minimum γ=0.232, ISE$\leqslant$0.03.

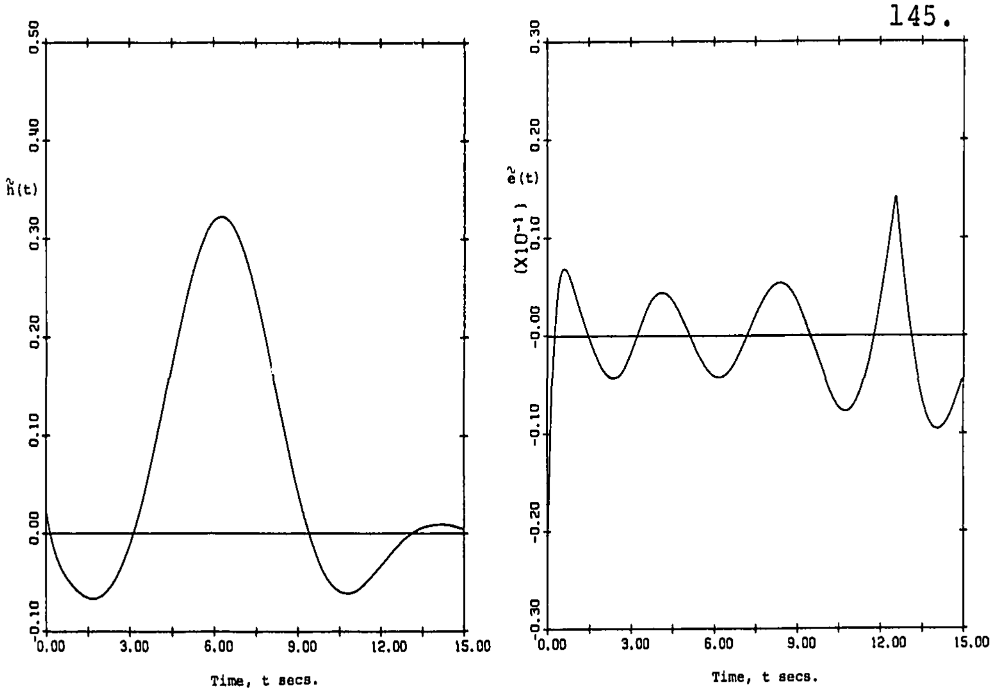


Figure 6.17. The least square approximation $\tilde{h}(t)$ and error $\tilde{e}(t)$. N=8, Delay=2π, Minimum ISE=0.0052

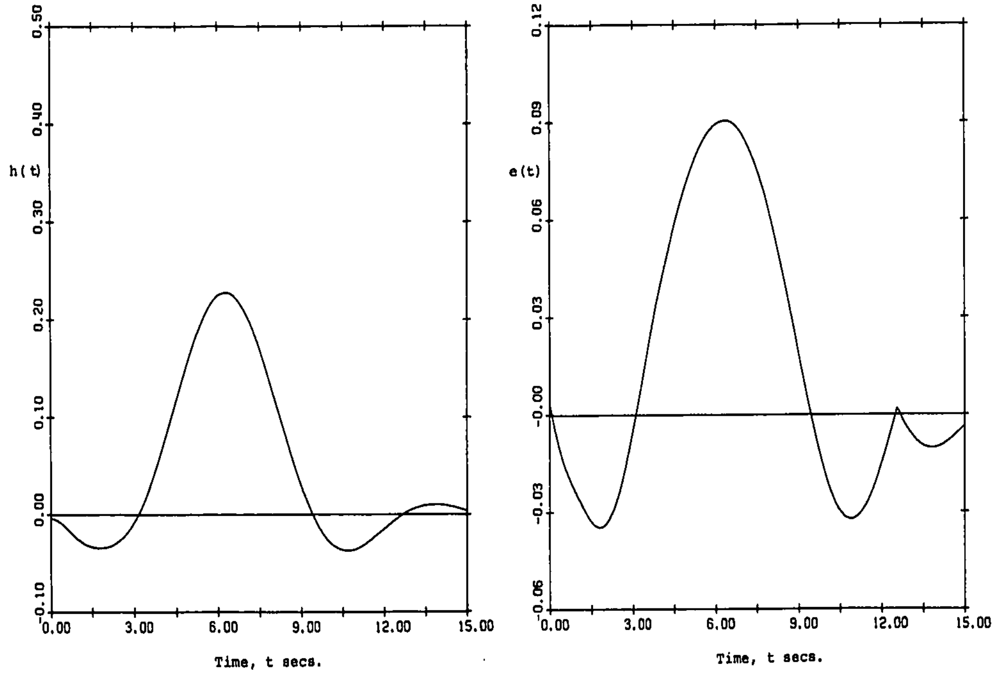


Figure 6.18. The approximation h(t) and error e(t) for minimum average error. N=8, Delay=2π, Minimum γ=0.232, ISE$\leq$0.03

SUMMARY AND CONCLUSIONS OF NUMERICAL EXAMPLES

The numerical examples show that it is possible to improve the frequency response by minimizing the average error. In examples 2a and 2b there is considerable improvement in the frequency response while in examples 1a and 1b this improvement is not significant. In example 1 it was possible to reduce the ISE to a very low value and hence the unconstraint minimum of average error was very close to the centre of the sphere of equation (6.29). Therefore the minimization of average error gave a set of coefficients very close to those at minimum ISE. But in example 2, it was not possible to reduce the ISE to such low values as in the first example. The coefficients giving the unconstrained minimum of average error, were not close to the centre of the sphere of equation (6.29). In this case it was possible to improve the frequency response by minimizing the average error while keeping the ISE less than some suitable value. It was observed that the magnitude characteristics of these improved responses had a constant attenuation at the low frequency ($\omega < 1$). This was due to the relaxing of the condition on ISE. This cculd be overcome by any one of the following three methods.

(1) By reducing the radius of the sphere giving the constraint on ISE; that is, by reducing ISE.

(2) By multiplying H(s) with a suitable constant.

(3) By keeping the constant term of the numerator of $\tilde{H}(s)$ and H(s) to be the same.

In the examples 2a and 2b the frequency responses of H(s) were adjusted by method (3). It was found that this gave the desired response at low frequencies without affecting the improvement obtained at higher frequencies.

The frequency responses of these filters at higher frequencies are inferior to those of the corresponding Butterworth filters. But, these filters have better frequency characteristics in the transition band. For example in case 2b ($N = 8$, Delay $= 2\pi$, γ minimum) the maximum ripple in the pass band has a value of -43db. This magnitude is first attained at $\omega = 1.54$. The Butterworth response of 8th order filter at $\omega = 1.54$ is only -29db (53).

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

The problem of approximating an arbitrary frequency response by a realizable filter has been studied in this research. The general approach employed was to convert the frequency domain approximation into an equivalent time domain approximation. This was done by making use of two important relations between the frequency domain and the time domain errors. These are the equivalence of least square criterion in the frequency domain and in time domain and the relation between the least upper bound of the error in frequency domain and the average error in the time domain. A new method of minimizing the upper bound of the magnitude of the frequency domain error with constraints on integral squared error has been developed. This method made use of a set of orthonormal functions of exponentials.

One important aspect of the problem was the determination of the poles of the filter which would yield the least integral squared error. The concept of complimentary filter was used to evaluate the ISE. This has the advantage that the ISE corresponding to any set of distinct poles could be easily computed using a simple filtering operation. This did not involve computations using complex numbers. An algorithm was developed and implemented to compute the ISE. The minimization of ISE was carried out by using Zangwill algorithm. This algorithm was an improvement on Powell's method of minimizing a function of several variables without calculating derivatives. The ISE is very insensitive to variations in pole positions over a wide range around the optimum point and hence the gradient methods of minimization was not successful in minimizing ISE. This led to the use of Zangwill algorithm. This algorithm with some modifications was implemented. The mathematical basis of this method was also reviewed.

A new method of choosing the zeros of the filter so as to minimize the deviation in the frequency domain was developed. It was proved that the average error in time domain is related to the least upper bound of the magnitude of the deviation in the frequency domain. This gave a convenient way of minimizing the deviation in the frequency domain. However, the minimization of the average error does not mean the minimization of ISE and vice versa. It was found that minimization of one of these errors alone need not give the desired result. This led to the development of a method of minimizing the average error subject to constraints on ISE. This was carried out by using a set of orthonormal functions of exponentials. The mathematical basis of this scheme of minimization was established. It was proved that for every ISE less than or equal to a preassigned value there exists a set of coefficients of the orthonormal expansion such that the average error in time domain at these coefficients is a minimum. The uniqueness properties of this approximation were also discussed. The actual minimization was carried out by using the Fletcher Powell algorithm. The constraints on the ISE were implemented by using the penalty function approach.

In order to implement the above scheme of minimization it was found necessary to compute the inverse of a Laplace transform expressed as a rational function of complex frequency. A new method of numerically computing this inverse was developed. In this method the transient response was expressed as a series expansion in powers of the independent variable t. The coefficients of the series expansion could be easily computed by using the recursive relations established for this purpose. The new method was found to be more efficient than some of the already existing methods of evaluating the transient response (5). This method makes the approximation of an arbitrary frequency response using the time domain approach, more practical.

The results obtained by applying this technique to the specific example of approximating the ideal low pass were very encouraging. The frequency characteristics of the filters obtained by this method were better than those of the least square filters. This improvement in the frequency response is not very marked in the first example considered (Delay = π) while in the second example (Delay = 2π) the improvement was very significant. This was because of the fact that in the first case it was possible to minimize the ISE to a very low value and hence the two points of minimum ISE and minimum average error were close to each other while in the second case this was not true.

RECOMMENDATIONS FOR FUTURE WORK

In the example considered, the time domain function of the desired frequency response was analytically known. In general, when approximating an arbitrary frequency response the corresponding time domain function should first be computed. This can be done by using the Fast Fourier Transform algorithm. It is also possible to check beforehand by using the inverse FFT whether the truncated time function can give a sufficiently close approximation of the desired frequency response. Once the time function is known it should be reversed in time before applying as input to the complimentary filter for evaluating the ISE.

The extension of the method to the design of Recursive Digital filters for arbitrary frequency response offers an interesting field for future work in this direction. A set of discrete orthonormal functions have already been developed by Young and Huggins (54) and Broome (55). The least square error of approximation can be evaluated by using the concept complimentary filter in Z-domain (56). All the theories developed in this dissertation can be extended to the discrete system. The new method of inverting the Laplace transform can be modified to develop a corresponding numerical method inverting Z-transform. The method of computing transient response can also be made use of to find the equivalent Z-transform expression of a Laplace transform (57).

Another field of future work is to use the optimum poles to construct a set of functions satisfying the Harr condition and to carry out the minimization of the average error with respect to the coefficients of these functions. This will improve the quality of the approximation.

This research has shown that it is possible to get a better frequency response by relaxing the condition on ISE. This suggests the possibility of directly minimizing the Chebyshev error in the frequency domain without using the relation between the upperbound in the frequency domain and the average error in time domain. This involves the evaluation of an analytical expression for Chebyshev error in frequency domain.

REFERENCES

1. C. Pottle and J. S. Thorp, "On the Optimum Least-Squares Approximation to the Ideal Low Pass," Proc. 1964 2nd Ann. Allerton Conf. on Circuit and System Theory, pp. 201-219.

2. C. Pottle and J. C. K. Wong, "Optimum Least-Squares Approximations to the Ideal Low-Pass Filter," IEEE Transactions on Circuit Theory (correspondence), CT-17, May 1970, pp. 282-284.

3. Meredith S. Ulstad, "Time Domain Approximations and an Active Network Realization of Transfer Functions Derived from Ideal Filters," CT-15, No. 3, pp. 205-210, September 1968.

4. Howard D. Helms, "Nonrecursive Digital Filters: Design Methods for Achieving Specifications on Frequency Response," IEEE Transactions on Audio and Electroacoustics, AU-16, No. 3, pp. 336-342, September 1968.

5. K. A. Muraleedharan and K. A. Stromsmoe, "On Computation of Transient Response," Proc. IEEE (Lett.), Vol. 60, p. 995, August 1972.

6. John B. Thomas, An Introduction to Statistical Communication Theory, John Wiley and Sons, New York, 1969.

7. David F. Tuttle, Jr., Network Synthesis, Vol. 1, John

Wiley and Sons, Inc., New York, 1958.

8. Louis Weinberg, Network Analysis and Synthesis, McGraw Hill, New York, 1962.

9. E. A. Guillemin, Synthesis of Passive Networks, John Wiley and Sons, Inc., New York, 1957.

10. L. S. W. Kendall, Time-Domain Synthesis of Linear Networks, Prentice-Hall, Inc., New Jersey, 1971.

11. Lewis Franks, Signal Theory, Prentice-Hall, Inc., New Jersey, 1961.

12. R. Courant and D. Hilbert, Methods of Mathematical Physics, Vol. I, Interscience, New York, 1953.

13. P. R. Aigrain and E. M. Williams, "Synthesis of n-reactance networks for desired transient response," Journal of Applied Physics, Vol. 20, 6, pp. 597-600, June 1949.

14. G. Sansone, Orthogonal Functions, Interscience Publishers, Inc., New York, 1959.

15. Y. W. Lee, Statistical Theory of Communicaiton, John Wiley and Sons, Inc., New York, 1960.

16. E. C. Titchmarsh, Theory of Fourier Integrals, Oxford University Press, London, 1937.

17. William H. Kautz, "Transient Synthesis in the Time Domain," IRE Transactions on Circuit Theory, CT-1, No. 3, pp. 29-39, September 1954.

18. T. Y. Young and W. H. Huggins, "Complimentary Signals and Orthogonalized Exponentials," IRE Transactions on Circuit Theory, CT-9, No. 4, pp. 362-370, December 1962.

19. T. Y. Young and W. H. Huggins, "On the Representation of Electrocardiograms," IEEE Transactions on Bio-Medical Electronics, BME-10, No. 3, pp. 86-95, July 1963.

20. T. Y. Young, "Signal Theory and Electrocardiography," Dr. Eng dissertation, The Johns Hopkins University, 1962.

21. Dan C. Ross, "Orthonormal Exponentials," Proceedings of National Electronic Conference, V-18, pp. 838-849, October 1962.

22. Lawrence A. O'Neill, "The Representation of Continuous Speech with a Periodically Sampled Orthogonal Basis," IEEE Transactions on Audio and Electroacoustics, Vol. AU-17, No. 1, pp. 14-21, March 1969.

23. Moyett T. Clark, "Word Recognition by Means of Orthogonal Functions," IEEE Transactions on Audio and Electroacoustics, Vol. AU-18, No. 3, pp. 304-312, September 1970.

24. Balth Van Der Pol and H. Bremmer, Operational Calculus based on the Two-Sided Laplace Integral, Univer-

sity Press, Cambridge, 1955.

25. Ruel V. Churchill, Complex Variables and Applications, McGraw-Hill Book Company, New York, 1960.

26. R. N. McDonough, "Matched Exponents for the Representation of Signals," Dr. Eng dissertation, The Johns Hopkins University, 1963.

27. William C. Yengst, "Approximation to a Specified Time Response," IRE Transactions on Circuit Theory, CT-9, No. 2, pp. 152-162, June 1962.

28. George A. Perdikaris and Gladwyn V. Lago, "Synthesis of Digital and/or Continuous Networks in Time Domain," Swieeco Record of Technical Papers, pp. 372-376, April 22-24, 1970, Dallas, Texas.

29. Benjamin C. Kuo, Analysis and Synthesis of Sampled-Data Control Systems, Prentice-Hall, Inc., New Jersey.

30. Cristofor G. Vasiliu, "A Practical Method for Time-Domain Network Synthesis," IEEE Transactions on Circuit Theory, CT-12, pp. 234-241, June 1965.

31. R. N. McDonough and W. H. Huggins, "Best Least-Squares Representation of Signals by Exponentials," IEEE Transactions on Automatic Control, AC-13, No. 4, pp. 408-412, August 1968.

32. Raymond W. Sears, Jr., "Digital Optimization of Exponential Representations of Signals," pp. 44-55,

IEEE Convention Record 1967.

33. R. W. Sears, Jr., "Digital Optimization of Exponential Representations of Signals," Ph.D. dissertation, The John Hopkins University.

34. L. E. McBridge, Jr., H. W. Schaefgen and K. Steiglitz, "Time-Domain Approximation by Iterative Methods," IEEE Transactions on Circuit Theory, CT-13, No. 4, pp. 381-387, December 1966.

35. Gerry Miller, "An Iterative Solution to the Equations of Aigrain and Williams," IEEE Transactions on Circuit Theory (Correspondence), CT Vol. 17, pp. 155-158, February 1970.

36. Gerry Miller, "Least-Squares Approximation of Functions by Exponentials," Ph.D. dissertation, The John Hopkins University, 1969.

37. John R. Rice, The Approximation of Functions, Vol. 1, Addison-Wesley Publishing Company, Inc., Massachusetts, 1964.

38. E. W. Cheney, Introduction to Approximation Theory, McGraw-Hill Book Company, New York, 1966.

39. T. M. Apostol, Mathematical Analysis, Addison-Wesley Publishing Company, Inc., Massachusetts, 1957.

40. Murlan S. Corrington, "Simplified Calculation of Transient Response," Proc. IEEE, Vol. 53, pp. 287-292, March 1965.

41. M. R. Aaron and J. F. Kaiser, "On the Calculation of Transient Response," Proc. IEEE (Correspondence), Vol. 53, pp. 1269, September 1965.

42. M. L. Liou, "A Novel Method of Evaluating Transient Response," Proc. IEEE, Vol. 54, pp. 20-23, January 1966.

43. W. E. Thomson, "Evaluation of Transient Response," Proc. IEEE (Lett.), Vol. 54, p. 1584, November 1966.

44. J. Valant, "Calculation of Transient Response," Electron. Lett., Vol. 4, p. 261, June 1968.

45. Katsuhiko Ogata, "State Space Analysis of Control Systems, "Prentice-Hall, Inc., New Jersey, 1967.

46. F. F. Kuo and J.F. Kaiser, System Analysis By Digital Computer, John Wiley and Sons, Inc., New York, 1966.

47. M. J. D. Powell, "An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives," The Computer Journal, Vol. 7, pp. 155-162, July 1964.

48. Willard I. Zangwill, "Minimizing a Function Without Calculating Derivatives," The Computer Journal, 10, pp. 293-296, November 1967.

49. H. H. Rosenbrock, "An Automatic Method for Finding the Greatest or Least Value of a Function," The Computer Journal, Vol. 3, pp. 175-184, October 1960.

50. H. A. Spang, "A Review of Minimization Techniques for Nonlinear Functions," S.I.A.M. Review, Vol. 4, pp. 343-365, October 1962.

51. Donald A. Pierre, Optimization Theory with Applications, John Wiley and Sons, Inc., New York, 1969.

52. R. Fletcher and M. J. D. Powell, "A Rapidly Convergent Descent Method for Minimization," The Computer Journal, Vol. 6, pp. 163-168, July 1963.

53. A Handbook on Electrical Filters, White Electromagnetics, Inc., Rockville, 1963.

54. T. Y. Young and W. H. Huggins, "Discrete Orthonormal Exponentials," Proceedings of the National Electronic Conference, V-18, pp. 10-18, October 1962.

55. Paul W. Broome, "Discrete Orthonormal Sequences," Journal of the Association for Computing Machinery, Vol. 12, No. 2, pp. 151-168, April 1965.

56. M. T. Clark, "Optimization of the Representation of Sampled Data Signals on Orthonormal Bases," Electronic and Aerospace Systems, Eascon '68 Convention Record, pp. 414-421, September 1968.

57. Prentiss N. Robinson and Guner S. Robinson, "A Computer Method for Obtaining Z-transforms," IEEE Transaction on Audio and Electroacoustics (Correspondence), AU-20, No. 1, pp. 98-99, March 1972.

APPENDIX I

```
      SUBROUTINE ERROR(NDIM,AC,VALUE,Y,DERY,AUX,RFX)
      EXTERNAL FCT,OUTP,RFX
C  SUBROUTINE ERROR(NDIM,AC,VALUE,Y,DERY,AUX,RFX)
C  THIS SUBROUTINE COMPUTES THE INTEGRAL SQUARED ERROR
C  (ISE) OF EXPONENTIAL REPRESENTATION BY THE METHOD OF
C  COMPLIMENTARY FILTER.IT USES THE RKGS SUBROUTINE IN
C  SSPLIB OF IBM.THE ARGUMENT LIST OF RKGS IS CHANGED BY
C  ADDING A NEW PARAMETER AC OF N VECTOR.AC IS THE VECTOR
C  CONTAINING THE COEFFICIENTS OF THE COMPLIMENTARY FIL
C  TER.A(1) IS THE COEFFICIENT OF S**(N-1) OF DENOMINATOR
C  OF COMPLIMENTARY FILTER.
C  THE COMMON STATEMENT IS USED WITH SUBROUTINES ERROR,
C  ZAGMIN, FCT,AND OUTP
C NDIM # OF THE COEFFICIENTS
C AC-NDIM VECTOR CONTAINING THE COEFFICIENTS OF DR OF
C  COMPLIMENTARY FILTER
C  VALUE-THIS GIVES THE ISE ON RETURN
C  Y-VECTOR OF DIMENSION NDIM. CONTAINING THE STATE
C  VARIABLES.INPUT VALUES OF THIS VECTOR ARE ZEROS.
C   DERY-VECTOR OF DIMENSION NDIM . INPUT VALUES ARE
C   SPECIFIED BY RKGS.
C  AUX-AS SPECIFIED BY RKGS
C  RFX- A FUNCTION SUBPROGRAMME GIVING THE TIME REVERSED
C  SIGNAL
C  DELT-INTERVAL OF INTEGRATION.ERR- ISE. V-VECTOR OF
C  DIMENSION 5,A WORKSPACE.
C  BV,S,KOU ARE WORKSPACES.
C  TL-DURATION OF SIGNAL
      REAL PRMT(6),Y(10),DERY(10),AUX(8,10),AC(10),V(5)
      COMMON DELT,ERR,BV,V,S,TL,KOU
      ERR=0.0
      KOU=0
      BV=0.
      S=4.*DELT
      PRMT(1)=0.
      PRMT(2)=TL
      PRMT(3)=DELT
      PRMT(4)=1.E-4
      PRMT(6)=DELT
      X=NDIM
      XX=1./X
      DO 7 I=1,NDIM
      Y(1)=0.
    7 DERY(I)=XX
      CALL RKGS(PRMT,Y,DERY,NDIM,IHLF,FCT,OUTP,AUX,AC)
      VALUE=ERR
      RETURN
      END

      SUBROUTINE FCT(X,Y,DERY,NDIM,AC,RFX)
C  THIS SUBROUTINE COMPUTES THE DERIVATIVES OF STATE
C  VARIABLES OF COMPLIMENTARY FILTER AT X.
C  ALL PARAMETERS ARE AS GIVEN IN ERROR.
      EXTERNAL RFX
      REAL Y(10),DERY(10),AC(10),V(5)
      COMMON DELT,ERR,BV,V,S,TL,KOU
      INTEGER KOU
      N=NDIM-1
      DO 7 I=1,N
      II=I+1
    7 DERY(I)-Y(II)
      DERY(NDIM)=RFX(X)
      DO 8 I=1,NDIM
      L=NDIM-I+1
    8 DERY(NDIM)=DERY(NDIM)-AC(I)*Y(L)
      RETURN
      END

      SUBROUTINE OUTP(X,Y,DERY,IHLF,NDIM,PRMT,AC,RFX)
C     THIS SUBROUTINE COMPUTES THE ISE FROM THE OUTPUT
C     VALUES OF RKGS.
C  ALL PARAMETERS ARE AS GIVEN BY ERROR AND RKGS.
      EXTERNAL RFX
      REAL PRMT(6),Y(10),DERY(10),AUX(8,10),AC(10),V(5)
      COMMON DELT,ERR,BV,V,S,TL,KOU
      INTEGER KOU
      IF(IHLF.GE.11)WRITE(6,101)IHLF,X
  101 FORMAT('0','IHLF=',13,2X,F12.7)
      IF ((PRMT(6)-X).LT.(1.E-5))GO TO 17
      RETURN
   17 KOU=KOU+1
      IF(KOU.EQ.5)GO TO 27
      IF(KOU.GE.6)GO TO 30
      V(KOU)=RFX(X)
   11 DO 7 I=1,NDIM,2
      L=NDIM-I+1
    7 V(KOU)=V(KOU)-(AC(I)*Y(L)+AC(I)*Y(L))
      PRMT(6)=PRMT(6)+DELT
      RETURN
 27   V(KOU)=RFX(X)
   13 DO 8 I=1,NDIM,2
      L=NDIM-I+1
    8 V(KOU)=V(KOU)-(AC(I)*Y(L)+AC(I)*Y(L))
```

```
      V1=V(5)**2
      ERR=ERR+(BV+3.875*(V(1)**2+V(4)**2)+2.625*(V(2)**2
      +V(3)**2)+V1)
      IF((PRMT(2)-PRMT(6)).LT.(1.E-6))GO TO 29
      PRMT(6)=PRMT(6)+DELT
      IF((PRMT(2)-PRMT(6)).GE.S)GO TO 31
      KOU=5
      BV=V1
      ERR=ERR*(DELT/3.)
      RETURN
   31 KOU=0
      BV=V1
      RETURN
   29 ERR=ERR*(DELT/3.)
      PRMT(5)=0.
      RETURN
 30   V1=RFX(X)
   15 DO 9 I=1,NDIM,2
      L=NDIM-I+1
    9 V1=V1-(AC(I)*Y(L)+AC(I)*Y(L))
      V1=V1**2
      ERR=ERR+(DELT/2.)*(BV+V1)
      IF((PRMT(2)-PRMT(6)).LT.(1.E-6))GO TO 32
      PRMT(6)=PRMT(6)+DELT
      BV=V1
      RETURN
   32 PRMT(5)=0.
      RETURN
      END

      FUNCTION RFX(X)
C  THIS FUNCTION COMPUTES THE VALUE OF TIME REVERSED SIN
C  (X)/X FUNCTION OF DELAY (2*PI)AND DURACTION (4*PI)
      DOUBLE PRECISION T,TD,TU,PI,FFX
      TD=2.DO*PI
      TU=4.DO*PI
      PI=3.1415927DO
      T=TU-X
      RFX=FFX(T,TD,TU)
      RETURN
      END

      FUNCTION FFX(T,TD,TU)
C  THIS FUNCTION COMPUTS THE VALUE OF SIN(T)/T FUNCTION
C  FOR DELAY (TD) AND DURATION (TU)
      DOUBLE PRECISION T,TD,PI,TX,TU,FFX
      IF(T.GE.TU) GO TO 3
      PI=3.1415927DO
      TX=T-TD
      IF(DABS(TX).LT.1.D-7) GO TO 1
      FFX=DSIN(TX)/TX
      GO TO 2
1     FFX=1.DO
2     FFX=FFX/PI
      RETURN
3     FFX=0.DO
      RETURN
      END
      SUBROUTINE ZAGMIN(N,A,E,ALFA,TALF,EXS,KOUNT,A0,A2,
     P,Y,DERY,AUX)
C THIS SUBROUTINE MINIMISES A FUNCTION OF N VARIABLES
C WITHOUT COMPUTING THE DERIVATIVES.THIS USES THE ZANGWILL
C ALGORITHM.
C SUBROUTINE ZAGMIN(N,A,E,ALFA,TALF,EXS,KOUNT,A0,A2,P,Y,
C DERY,AUX)
C N-#OF VARIABLES-INPUT
C A-VECTOR OF DIMENSION N,CONTAINING THE INITIAL VALUES
C OF THE VARIABLES.
C ON RETURN A CONTAINS THE FINAL VALUES.
C E-FUNCTION VALUE
C ALFA-INITIAL INCREMENT OF SEARCH,INPUT.
C TALF-A SMALL VALUE FOR TESTING THE MINIMUM.IF FUNCTION
C DOES NOT CHANGE FOR ANY  DISTANCE LARGER THAN TALF THE
C POINT IS MINIMUM.
C EXS-A SMALL VALUE OF FUNCTION TO TEST MINIMUM.IF
C FUNCTION CHANGE IS LESS THAN EXS IN ONE ITERATION
C MINIMUM IS FOUND
C KOUNT-MAXIMUM # OF ITERATIONS.
C A0,A2,Y,DERY - EACH OF DIMENSION N IS WORKSPACE
C P-ARRAY OF DIMENSION (N,N) WORKSPACE
C AUX-ARRAY OF (8,N).WORK SPACE
C THIS SUBROUTINE WILL WRITE THE VECTOR A AND FUNCTION
C VALUE AFTER EACH ONE DIMENSIONAL SEARCH.
      REAL A(10),A0(10),A2(10),Y(10),DERY(10),AUX(8,10),
     V(5)
      REAL P(10,10)
      COMMON DELT,ERR,BV,V,S,TL,KOU
C INITIALISATION
      N1=N+1
      N2=N-1
      SALF=ALFA
      IC=0
      KO=0
      DO 101 II=1,N
      DO 101 JJ=1,N
      P(II,JJ)=0.
      IF(II.EQ.JJ)P(II,JJ)=1.
  101 CONTINUE
      DO 102 K=1,N
      AO(K)=A(K)
  102 A2(K)=A(K)
      CALL ERROR(N,A,VALUE,Y,DERY,AUX)
      WRITE(6,500)N,ERR,(A(IZAG),IZAG=1,N)
  500 FORMAT('0','ORDER',I2,1X,'INITIAL ISE',F12.7,1X,
     'A=',(8F12.7))
      E2=ERR
      OER=ERR
C  MINIMISATION ALONG CONJUGATE DIRECTIONS.
    1 J=0
      I=0
    2 J=J+1
      IF(J.EQ.N1) GO TO 4
      GO TO 5
    4 J=N
      I=1
    5 DO 103 K=1,N
  103 A(K)=A2(K)+ALFA*P(K,J)
      WRITE(6,900)J,ALFA
      CALL ERROR(N,A,VALUE,Y,DERY,AUX)
      E3=ERR
      IF(ERR.LT.E2) GO TO 11
      DO 104 K=1,N
  104 A(K)=A2(K)-ALFA*P(K,J)
      CALL ERROR(N,A,VALUE,Y,DERY,AUX)
      E1=ERR
      IF(ERR.LT.E2) GO TO 8
      IF(ALFA.LT.TALF) GO TO 10
      ALFA=ALFA/2.
      GO TO 5
    8 ALFA=ALFA
      GO TO 11
   10 DO 105 K=1,N
  105 A(K)=A2(K)
      ERR=E2
      GO TO 15
   11 E1=E2
      E2=ERR
      ALFA=ALFA+ALFA
      DO 106 K=1,N
  106 A(K)=A(K)+ALFA*P(K,J)
      WRITE(6,800)ALFA
      CALL ERROR(N,A,VALUE,Y,DERY,AUX)
      IF(ERR.LT.E2) GO TO 11
      E3=ERR
      ALFA=ALFA/2.
      DO 107 K=1,N
  107 A(K)=A(K)-ALFA*P(K,J)
      CALL ERROR(N,A,VALUE,Y,DERY,AUX)
      IF(ERR.LT.E2) GO TO 12
      E3=ERR
      DO 707 K=1,N
  707 A2(K)=A(K)-ALFA*P(K,J)
      DFA=(ALFA/2.)*(((-3.)*E1+(4.)*Ew-E3)/(-E1+(2.)*E2-
      E3))
      ALFA=DFA-ALFA-ALFA
      DO 108 K=1,N
  108 A(K)=A(K)+ALFA*P(K,J)
      CALL ERROR(N,A,VALUE,Y,DERY,AUX)
      IF(ERR.LE.E2) GO TO 14
      DO 109 K=1,N
  109 A(K)=A2(K)
      ERR=E2
      GO TO 15
   14 DO 110 K=1,N
  110 A2(K)=A(K)
      E2=ERR
      GO TO 15
   12 E1=E2
      E2=ERR
      DFA=(ALFA/2.)*(((-3.)*E1+(4.)*E2-E3)/(-E1+(2.)*E2-
      E3))
      ALFA=DFA-ALFA
      DO 111 K=1,N
  111 A2(K)=A(K)+ALFA*P(K,J)
      CALL ERROR(N,A2,VALUE,Y,DERY,AUX)
      IF(ERR.LT.E2) GO TO 13
      DO 112 K=1,N
  112 A2(K)=A(K)
      ERR=E2
      GO TO 15
   13 E2=ERR
      DO 113 K=1,N
  113 A(K)=A2(K)
   15 IF(J.EQ.N) GO TO 16
   19 I=0
      ALFA=SALF
      WRITE(6,300)J,ERR,(A(IZAG),IZAG=1,N)
  300 FORMAT('0','CONJ DIR.',I2,1X,'ISE',F12.7,1X,'A=',
      (8F12.7))
      GO TO 2
   16 IF(I.EQ.O)GO TO 17
      IF((ABS(OER-ERR)).LE.(EXS)) GO TO 29
   33 KO=KO+1
      WRITE(6,600)KO,ERR,(A(IZAG),IZAG=1,N)
  600 FORMAT('0','ITERATION',I3,1X,'ISE',F12.7,1X,'A=',
      (8F12.7))
      IF(KO.GT.KOUNT) GO TO 30
      IC=IC+1
      ALFA=SALF
      GO TO 18
C   SEARCH ALONG N CONJUGATE DIRECTIONS ARE OVER.NEW CON
C   JUGATE DIRECTIONS ARE CHOSEN.
```

```
   17 DO 114 D=1,N
      DO 114 JJ=1,N2
      JJ1=JJ+1
  114 P(K,JJ)=P(K,JJ1)
      PL=0.
      DO 115 K=1,N
      P(K,N)=A(K)-AO(K)
  115 PL=PL+P(K,N)*P(K,N)
      PL=SQRT(PL)
      IF(PL.LE.1.E-6) GO TO 33
      DO 116 K=1,N
  116 P(K,N)=P(K,N)/PL
      GO TO 19
C   MINIMISATION ALONG THE CO-ORDINATE DIRECTIONS.
   18 I=0
      DO 117 K=1,N
  117 AO(K)=A(K)
   20 I=I+1
      IF(IC.EQ.N1)IC=1
      AIC=A(IC)
   21 A(IC)=AIC+ALFA
      WRITE(6,900)IC,ALFA
  900 FORMAT('0',I5,2X,'ALFA=',F12.7)
      CALL ERROR(N,A,VALUE,Y,DERY,AUX)
      E3=ERR
      IF(ERR.LT.F2) GO TO 24
      A(IC)=AIC-ALFA
      CALL ERROR(N,A,VALUE,Y,DERY,AUX)
      E1=ERR
      IF(ERR.LT.E2) GO TO 23
      IF(ALFA.LE.TALF) GO TO 22
      ALFA=ALFA/2.
      GO TO 21
   22 A(IC)=AIC
      WRITE(6,400)IC,E2,(A(IZAG),IZAG=1,N)
      IF(I.EQ.N) GO TO 31
      IC=IC+1
      ALFA=SALF
      GO TO 20
   23 ALFA=-ALFA
   24 E1=E2
      E2=ERR
      ALFA=ALFA+ALFA
      A(IC)=A(IC)+ALFA
      WRITE(6,800)ALFA
  800 FORMAT('0','ALFA=',F12.7)
      CALL ERROR(N,A,VALUE,Y,DERY,AUX)
      IF(ERR.LT.E2) GO TO 24
      E3=ERR
      ALFA=ALFA/2.
      A(IC)=A(IC)=ALFA
      CALL ERROR(N,A,VALUE,Y,DERY,AUX)
      IF(ERR.LT.E2) GO TO 26
      E3=ERR
      DFA=(ALFA/2.)*(((-3)*E1+(4.)*E2-E3)/(-E1+(2.)*E2-E3))
      AIC=A(IC)
      A(IC)=A(IC)-ALFA-ALFA+DFA
      CALL ERROR(N,A,VALUE,Y,DERY,AUX)
      IF(ERR.LT.E2) GO TO 25
      A(IC)=AIC-ALFA
      ERR=E2
      GO TO 28
   25 E2=ERR
      GO TO 28
   26 E1=E2
      E2=ERR
      DFA=(ALFA/2.)*(((-3.)*E1+(4.)*E2=E3)/(-E1+(2.)*E2-E3))
      AIC=A(IC)
      A(IC)=A(IC)-ALFA+DFA
      CALL ERROR(N,A,VALUE,Y,DERY,AUX)
      IF(ERR.LT.E2) GO TO 27
      A(IC)=AIC
      ERR=E2
      GO TO 28
   27 E2=ERR
   28 ALFA=SALF
      OER=ERR
      A2(IC)=A(IC)
      WRITE(6,400)IC,ERR,(A(IZAG),IZAG=1,N)
  400 FORMAT('0','COOR DIR',I2,1X,'ISE',F12.7,1X,'A=',
       (8F12.7))
      GO TO 1
C  FUNCTION CHANGE IS LESS THAN EXS,ITERATION IS TERMINATED.
   29 WRITE(6,118)EXS
  118 FORMAT('0','ERROR CHANGE IS LESS THAN',F12.7)
      E=ERR
      RETURN
   30 WRITE(6,119)
  119 FORMAT('0','KOUNT IS EXCEEDED')
      E=ERR
      RETURN
C  A MINIMUM IS OBTAINED.
   31 WRITE(6,120)
  120 FORMAT('0','CONVERGENCE IS OBTAINED')
      E=E2
      RETURN
      END
      SUBROUTINE SERIES(N,A,B,TL,F,X,NT)
C  THIS SUBROUTINE COMPUTES THE COEFFICIENTS OF THE
C  SERIES EXPANSION OF THE INVERSE OF A RATIONAL LAPLACE
C  TRANSFORM
C  N-DEGREE OF DENOMINATOR
C  A-VECTOR OF DIMENSION N.THIS CONTAINS THE DENOMINATOR
C  COEFFICIENTS OF LAPLACE TRANSFORM.A(1)-COEFFICIENT OF
C  S**(N-1) ETC.
C  B-COEFFICIENT VECTOR OF NUMERATOR,SIMILAR TO A.
C  F-A VECTOR TO STORE THE COEFFICIENTS OF SERIES EXPAN
C  SION.AN ARBITRARY DIMENSION OF F(300) IS USED.TL-
C  MAXIMUM VALUE OF TIME AT WHICH INVERSE IS REQUIRED.
C  NT-NUMBER OF TERMS OF SERIES EXPANSION
      DIMENSION A(10),B(10),X(10),F(300)
      DOUBLE PRECISION A,B,F,X,FACT,SUM,TL
      IF(TL.LE.1.DO)TL=1.DO
      FACT=TL
      NN=N-1
      NNN=N-2
      DO 1 I=1,NNN
 1    X(I)=0.DO
      X(NN)=TL
      X(N)=-A(1)*TL
      F(1)=B(1)
      DO 6 I=2,300
      SUM=0.DO
      DO 2 J=1,N
      JJ=N-J+1
 2    SUM=SUM+X(J)*B(JJ)
      F(I)=SUM
      IF((I.GT.N).AND.(DABS(F(I)).LE.1.D-16))GO TO 7
      FACT=TL/I
      DO 3 J=1,N
 3    X(J)=X(J)*FACT
      SUM=0.DO
      DO 4 J=1,N
      JJ=N-J+1
 4    SUM=SUM-A(JJ)*X(J)
      DO 5 J=1,NN
      JJ=J+1
 5    X(J)=X(JJ)
      X(N)=SUM
 6    CONTINUE
 7    NT=I
      RETURN
      END
      FUNCTION VAL(NT,F,TL,T)
C   THIS FUNCTION COMPUTES THE INVERSE OF A LAPLACE TRANS
C   FORM USING THE OUTPUT VALUES OF SUBROUTINE SERIES.
      DIMENSION F(300)
      DOUBLE PRECISION F,VAL,X,T,TL
      IF(TL.LE.1.DO)TL=1.DO
      X=T/TL
      VAL=F(NT)
      NTT=NT-1
      DO 1 I=1,NTT
      J=NTT-1+1
      VAL=VAL*X+F(J)
 1    CONTINUE
      RETURN
      END
```

```
C  COMPUTATION OF THE NRS AND DRS OF THE ORTHONORMAL
C  FUNCTIONS FROM THE POLE VALUES,N IS SYSTEM ORDER.IR-
C  NO.OF REAL POLES.VECTOR R HAS THE POLE VALUES,MAGNI
C  TUDE OF THE REAL POLES FIRST,FOLLOWED BY COMPLEX POLES,
C  MAGNITUDE OF REAL PART FIRST, FOLLOWED BY THAT OF THE
C  IMAGINARY PART
C  COEFFICIENTS OF NR AND DR OF THE FUNCTIONS ARE PRINTED
C  IN ORDER FROM LARGEST TO SMALLEST POWER.ON RETURN R
C  VECTOR HAS VALUES AS FOLLOWS:MAGNITUDE OF REAL POLES
C  FIRST FOLLOWED BY P1,Q1,P2,Q2 ETC.
C   PMPY IS THE SUBROUTINE FOR POLYNOMIAL OPERATIONS GIVEN
C  IN SSPLIB OF IBM.
       REAL R(10),XD(10),XN(10),Y(3),B(10),BB(10),Z(10)
       REAL A(10)
       READ(5,1)N
 1     FORMAT(I2)
       READ(5,1)IR
       READ(5,2)(R(I),I=1,N)
 2     FORMAT(10F20.7)
       WRITE(6,100)(R(I),I=1,N)
 100   FORMAT('0','POLE VALUES',10F12.7)
       XD(1)=1.
       XN(1)=1.
       IXD=1
       IXN=1
       Y(1)=1.
       IY=1
       IF(IR.EQ.0)GO TO 11
       Y(2)=1.
       IY=2
       DO 7 I=1,IR
       Y(1)=R(I)
       CALL PMPY(Z,IZ,XD,IXD,Y,IY)
       DO 3 K=1,IZ
 3     XD(K)=Z(K)
       IXD=IZ
       IF(I.EQ.1) GO TO 5
       II=I-1
       Y(1)=-R(II)
       CALL PMPY(Z,IZ,XN,IXN,Y,IY)
       IXN=IZ
       DO 4 K=1,IZ
 4     XN(K)=Z(K)
 5     DO 6 K=1,IXN
 6     B(K)=SQRT(R(I)+R(I))*XN(K)
       DO 21 KJI=1,IXN
       KJII=IXN-KJI+1
 21    A(KJI)=XD(KJII)
       WRITE(6,101)I,(B(K),K=1,IXN)
```

```
101   FORMAT('0','NR OF',I2,1X,10F12.7)
      WRITE(6,102)I,(A(K),K=1,IXN)
101   FORMAT('0','DR OF',I2,1X,10F12.7)
7     CONTINUE
      Y(1)=-R(IR)
      Y(2)=1.
      IY=2
11    IC=N-IR
      IC=IC/2
      DO 10 I=1,IC
      IN=IR+2*I-1
      INN=IN+1
      R(INN)=R(INN)*R(INN)+R(IN)*R(IN)
      R(IN)=R(IN)+R(IN)
      CALL PMPY(Z,IZ,XN,IXN,Y,IY)
      IXN=IZ
      SQRIN=SQRT(2.*(R(IN)))
      SQRINN=SQRT(R(INN))
      DO 8 J=1,IZ
      XN(J)=Z(J)
      K=IZ-J+1
      JJ=J+1
      B(J)=SQRIN*Z(K)
8     BB(JJ)=SQRINN*B(J)
      WRITE(6,203)(XN(KK),KK=1,IXN)
203   FORMAT('0','XN',10F12.7)
      IZ=IZ+1
      B(IZ)=0.
      BB(1)=0.
      WRITE(6,103)IN,(B(K),K=1,IZ)
103   FORMAT('0','B OF',I2,1X,10F12.7)
      WRITE(6,104)INN,(BB(K),K=1,IZ)
104   FORMAT('0','BB OF',I2,1X,10F12.7)
      IY=3
      Y(1)=R(INN)
      Y(2)=R(IN)
      Y(3)=1.
      CALL PMPY(Z,IZ,XD,IXD,Y,IY)
      IXD=IZ
      DO 9 K=1,IZ
9     XD(K)=Z(K)
      WRITE(6,105)IN,(XD(K),K=1,IXD)
105   FORMAT('0','XD OF',I2,10F12.7)
      Y(2)=-Y(2)
10    CONTINUE
      WRITE(6,100)(R(K),K=1,N)
      STOP
      END
```

```
C  THIS PROGRAMME COMPUTES THE ORTHONORMAL COEFFICIENTS
C  AT MINIMUM ISE
C  INTEGRATION IS DONE BY USING THE SUBROUTINE DQG32 OF
C  SSPLIB.
C  FCTX IS A FUNCTION SUBPROGRAMME GIVING THE FUNCTION TO
C  BE INTEGRATED
C  THIS PROGRAMME WRITES OUT THE ORTHONORMAL COEFFICIENTS
C  IN ORDER.
C  TL-DURATION OF FUNCTION.THIS SHOULD BE SPECIFIED
C  COMMON STATEMENT SHOULD BE USED WITH FUNCTION FCTX
C  INPUT DATA IS AS FOLLOWS
C  N-ORDER OF ORTHONORMAL FUNCTION
C  A-DR OF ORTHONORMAL FUNCTION,A(1) IS THE COEFFICIENT
C  OF S**(N-1)
C  B-NR COEFFICIENTS SIMILAR TO A
C  ORTHONORMAL FUNCTIONS ARE ARRANGED FROM SMALLEST TO
C  LARGEST.
C  LAST CARD IS BLANK.THIS TERMINATES THE PROGRAMME.
      DIMENSION F(300),X(10),A(10),B(10)
      DOUBLE PRECISION PI,XL,XU,FCTX,Y,V,T,F,TL,X,A,B
      COMMON F,TL,NT
      EXTERNAL FCTX
      PI=3.1415927DO
      T=PI/2.DO
      TL=4.DO*PI
      NL=TL/T
      NI=0
      SUM=0.DO
 6    READ(5,1)N
      IF(N.EQ.O) GO TO 7
      NL=TL/T+.5
      READ(5,2)(A(I),I=1,N)
      READ(5,2)(B(I),I=1,N)
 1    FORMAT(I2)
 2    FORMAT(10F20.7)
      WRITE(6,3)(A(I),I=1,N)
 3    FORMAT('0','A=',10F12.7)
      WRITE(6,4)(B(I),I=1,N)
 4    FORMAT('0','B=',10F12.7)
      CALL SERIES(N,A,B,TL,F,X,NT)
      XU=0.DO
      V=0.DO
      DO 5 K=1,NL
      XL=XU
      XU=XU+T
      CALL DQG32(XL,XU,FCTX,Y)
 5    V=V+Y
      NI=NI+1
      WRITE(6,101)NI,V
 101  FORMAT('0','C',I1,'=',F25.16)
      GO TO 6
 7    STOP
      END

      FUNCTION FCTX(X)
      DOUBLE PRECISION FCTX,X,PI,F,TL,VAL,FFX
      DIMENSION F(300)
      COMMON F,TL,NT
      PI=3,1415927DO
      TD=2.DO*PI
      FCTX=VAL(NT,F,TL,X)*FFX(X,TD,TL)
      RETURN
      END

      SUBROUTINE MDQG(N,KOUNT,C,EA,GRAD,M)
C   THIS SUBROUTINE COMPUTES THE AVERAGE ERROR AND GRADIENTS
C   N-ORDER OF SYSTEM.KOUNT-THE ITERATION # TO BE USED BY
C   DEMFP.C- COEFFICIENTS OF ORTHONORMAL FUNCTIONS,A N
C   VECTOR.EA-AVERAGE ERROR,GRAD-GRADIENTS OF EA WITH
C   RESPECT TO THE COEFFICIENTS.M-AN INTEGER,M=0 IF CON
C   STRAINT ON ISE IS SATISFIED.M=1 IF THIS IS NOT TRUE.
C   THIS SUBROUTINE WRITES M,KOUNT,WU,C,EA AND GRAD.
      DOUBLE PRECISION CX(3),XZ(3),FD(10),GRAD(10),C(10),
     *CH(10),E,EK,XU,T,XX,XL,XA,XB,XT,X,R,FFCT,S,EA,SGN,
     *WU,WT,RSQ
      DIMENSION PHI(10,910)
      COMMON WU,WT,RSQ,CH,FD,E EK,XU,T,PHI
      CALL GOC(NG.CS.XZ)
      NLT=(XU/T)+.5DO
      NLT=NLT*NG*2
      XX=0.DO
      EA=0.DO
      IPF=0
      DO 3 K=1,N
 3    GRAD(K)=0.DO
 4    XL=XX
      XX=T+XX
      XA=.5DO*(XX+XL)
      XB=.5DO*T
      DO 7 K=1,NG
      XT=XZ(K)*XB
      DO 6 KI=1,2
      IPF=IPF+1
      XT=-XT
      X=XA+XT
      R=FFCT(N,C,X,IPF)
      S=SGN(R)
      DO 5 KJ=1,N
      GRAD(KJ)=GRAD(KJ)+CX(K)*FD(KJ)*S
```

```
5     CONTINUE
      EA=EA+CX(K)*DAES(E)
6     CONTINUE
7     CONTINUE
      IF(IPF.GE.NLT) GO TO 8
      GO TO 4
8     EA=XB*EA
      DO 9 K=1,N
      GRAD(K)=XB*GRAD(K)
9     CONTINUE
      XX=0.DO
      DO 12 I=1,N
      XX=XX+(C(I)-CH(I))*(C(I)-CH(I))
12    CONTINUE
      XX=XX-RSQ
      IF(XX.LE.O.DO) GO TO 13
      M=1
      IF(KOUNT.EQ.O) GO TO 14
      EA=EA+WU*XX*XX
      XL=4.DO*WU*XX
      DO 15 I=1,N
      GRAD(I)=GRAD(I)+XL*(C(I)-CH(I))
15    CONTINUE
      GO TO 18
14    XT=0.DO
      XL=0.DO
      XA=16.DO*XX*XX
      DO 16 I=1,N
      XT=XT+GRAD(I)*GRAD(I)
      XL=XL+XA*(C(I)-CH(I))*(C(I)-CH(I))
16    CONTINUE
      WU=DSQRT(XT/XL)
      XL=4.DO*WU*XX
      EA=EA+WU*XX*XX
      DO 17 I=1,N
      GRAD(I)=GRAD(I)+XL*(C(I)-CH(I))
17    CONTINUE
      WU=WU/WT
      TO TO 18
13    M=0
18    WRITE(6,10)M,KOUNT,WU,(C(I),I=1,N)
10    FORMAT('0','M=',I1,1X,'KOUNT',I4,1X,'WU',F12.4,1X,
      'C=',(8F12.7))
      WRITE(6,11)EA,(GRAD(I),I=1,N)
11    FORMAT('0','EA=',F12.7,3X,'GRAD=',(8F12.7))
      RETURN
      END
      SUBROUTINE GQC(NG,CX,XZ)
C  THIS SUBROUTINE GIVES THE CONSTANTS OF 6 POINT GUASS
C  QUADRATURE FORMULA.
      DOUBLE PRECISION CX(3),XZ(3)
      NG=3
      XZ(1)=0.93246951420315203DO
      CX(1)=0.17132449237917034DO
      XZ(2)=0.66120938646626451DO
      CX(2)=0.36076157304813861DO
      XZ(3)=0.23861918608319691DO
      CX(3)=0.46791393457269105DO
      RETURN
      END

      FUNCTION FFCT(N,C,X,IPF)
      DOUBLE PRECISION C(10),FD(10),CH(10),X,E,EK,XU,T,
     *FFCT,FX,WU,WT,RSQ
      DIMENSION PHI(10,910)
      COMMON WU,WT,RSQ,CH,FD,E,EK,XU,T,PHI
      FFCT=0.DO
      DO 1 K=1,N
      FD(K)=PHI(K,IPF)
      FFCT=FFCT+C(K)*FD(K)
 1    CONTINUE
      FFCT=FFCT-FX(X)
      RETURN
      END

      FUNCTION SGN(X)
      DOUBLE PRECISION X,SGN
      IF(X.LT.O.DO) GO TO 1
      IF(X,GT.O.DO) GO TO 2
      SGN=0.DO
      RETURN
 1    SGN=-1.DO
      RETURN
 2    SGN=1.DO
      RETURN
      END

      FUNCTION FX(X)
C  THIS FUNCTION GIVES THE VALUE OF DELAYED SIN(X)/X FUNC
C  TION,DELAY=(2*PI),
C  DURATION (4PI)
      DOUBLE PRECISION FX,FFX,PI,X,TD,TL
      PI=3.1415927DO
      TD=2.DO*PI
      TL=4.DO*PI
      FX=FFX(X,TD,TL)
      RETURN
      END
C  THIS PROGRAMME MINIMISES THE AVERAGE ERROR WITH CON
C  STRAINTS ON ISE.THE PENALTY FUNCTION APPROACH IS USED
C  TO SATISFY THE CONSTRAINTS.WU IS A WEIGHT FACTOR WHICH
C  IS INCREASED BY A FACTOR WT AT THE END OF EACH ITERA
C  TION OF DFMFP,IF CONSTRAINT IS NOT SATISFIED.
C  E-LEAST ISE AS OBTAINED BY MINIMISATION OF ISE.EK-
C  ALLOWABLE ISE FOR MINIMISING THE AVERAGE ERROR.CH-
C  COEFFICIENTS OF ORTHONORMAL FUNCTIONS AT LEAST ISE.
C  XU-UPPER LIMIT OF INTEGRATION.T-INTERVAL OF INTEGRA
C  TION.PHI AN ARRAY OF (10,910) TO STORE VALUES OF ORTHO
C  NORMAL FUNCTIONS AT VALUES OF TIME AS REQUIRED BY THE
C  GUASS QUADRATURE FORMULA.C-A VECTOR OF DIMENSION N GIV
C  ING THE INITIAL VALUES OF COEFFICIENTS FOR MINIMISATION
C  OF AVERAGE ERROR.
C  THE COMMON STATEMENT IS USED WITH,SUBROUTINES MDQG,FFCT
C  AND DFMFP.
C  DFMFP IS THE FLETCHER-POWELL MINIMISATION GIVEN IN SSP
C  LIB OF IBM.
C  A VARIABLE EGS IS ADDED TO THE ARGUMENT LIST OF DFMFP.
C  THE MINIMISATION IS TERMINATED WHEN THE SUM OF THE
C  MAGNITUDES OF GRADIENTS IS LESS THAN EGS.
C  THE STATEMENT IF((M.EQ.1).AND.(KOUNT/N*N.EQ.KOUNT))WU=
C  WU*WT IS ADDED AT THE BEGINNING OF THE ITERATION OF
C  DFMFP
C  THIS PROGRAMME SHOULD GIVE THE VALUES OF WT,E,EK,XU,T
C  AND INITIAL WU
C  THIS PROGRAMME IS SET TO WORK FOR N=T,DELAY=2PI AND
C  DURATION=4PI
C  INPUT DATA
C        N-ORDER OF ORTHONORMAL FUNCTION
C        A-DENOMINATOR COEFFICIENTS OF ORTHONORMAL FUNC
C  TION.A(1) IS THE COEFFICIENT OF S**(N-1)
C        B-NUMERATOR COEFFICIENTS OF THE ORTHONORMAL FUNC
C  TIONS.B(1) IS THE COEFFICIENT OF S**(N-1)
C  THE FUNCTIONS ARE ARRANGED FROM THE SMALLEST TO LARGEST
C        BLANK CARD
C        N-ORDER OF APPROXIMATING FILTER.
C        CH-ORTHONORMAL COEFFICIENTS AT MINIMUM ISE
C        C-INITIAL VALUES OF THE COEFFICIENTS
      DOUBLE PRECISION CX(3),XZ(3),FD(10),GRAD(10),C(10),
     *H(60),G(10),A(10),B(10),F(300),XD(10),TL,XV,R,BN,XV,
     *T,XX,XL,XA,XB,XT,V,VAL,XU,WU,WT,RSQ,CH(10),PI
      DIMENSION PHI(10,910)
      COMMON WU,WT,RSQ,CH,FD,E,EK,XU,T,PHI
      TL=XU
      PI=3.1415927DO
      EXTERNAL MDQG
      T=.2DO
      XU=30.DO
      TL=XU
      CALL GQC(NG,CX,XZ)
      NLT=(XU/T)+.5DO
      NLT=NLT*NG*2
      IPHI=0
4     READ(5,1) N
1     FORMAT(I2)
      IF(N.EQ.0)GO TO 9
      IPHI=IPHI+1
      READ(5,2)(A(I),I=1,N)
      READ(5,2)(B(I),I=1,N
2     FORMAT(10F20.7)
      IPF=0
      WRITE(6,202)(B(I),I=1,N)
202   FORMAT('0','B=',10F12.7)
      WRITE(6,201)(A(I),I=1,N)
201   FORMAT('0','A=',10F12.7)
      CALL SERIES(N,A,B,TL,F,XD,NT)
      XX=0.DO
5     XL=XX
      XX=XL+T
      XA=.5DO*(XX+XL)
      XB=.5DO*T
      DO 7 K=1,NG
      XT=XZ(K)*XB
      DO 6 KI=1,2
      XT=-XT
      XV=XA+XT
      IPF=IPF+1
      PHI(IPHI,IPF)=VAL(NT,F,TL,XV)
6     CONTINUE
7     CONTINUE
      IF(IPF.GE.NLT) GO TO 4
      GO TO 5
9     READ(5,1)N
      READ(5,2)(CH(I),I=1,N)
      READ(5,2)(C(I),I=1,N)
      WT=4.DO
      WU=1.DO
      E=.7431E-3
      EK=.035
      EST=0.3
      EPS=1.E-4
      LIMIT=100
      EGS=.001
      RSQ=EK-E
      CALL DEMFP(MDQG,N,C,V,G,EST,EPS,EGS,LIMIT,IER,H)
      WRITE(6,101)(C(I),I=1,N)
101   FORMAT('0','C=',(8F12.7))
      WRITE(6,102)IER,V,G
102   FORMAT('0','IER',I2,1X,'EA',F12.7,1X,'GRAD',(8F12.7))
      STOP
      END
```

```
C  THIS PROGRAMME COMPUTES THE NUMERATOR COEFFICIENTS OF
C  THE APPRCXIMATION
C VECTOR C-COEFFICIENTS OF EXPANSION OF FUNCTIONS 1,2,ETC.
C  VECTOR R-MAGNITUDE OF REAL POLES FOLLOWED BY P1,Q1,P2,
C  Q2 ETC.
C INPUTS X AND Z ARE THE NR.COEFFICIENTS OF THE RESPEC
C  TIVE FUNCTIONS ORDERED FROM SMALLEST TO LARGEST POWER.
C  IN DATA NRS ARE READ FROM LARGEST TO SMALLEST FUNCTION
C  RESPY.
C  INPUT DATA IS ARRANGED AS FOLLOWS.
C  N-ORDER OF APPROXIMATION
C  IR-NUMBER OF REAL POLES
C  C VECTOR AT WHICH NRS OF APPROXIMATION IS DESIRED.
C  R VECTOR.MAGNITUDE OF REAL POLES FIRST FOLLOWED BY P1,
C  Q1,P2,Q2 ETC.
C  IX-ORDER OF ORTHONORMAL FUNCTION
C  X VECTOR-NRS OF ORTHONORMAL FUNCTION IX ORDERED FROM
C  SMALLEST TO LARGEST POWER.
C  Z VECTOR-NRS OF ORTHONORMAL FUNCTION (IX-1) ORDERED FROM
C  SMALLEST TO LARGEST POWER.  THE ORTHONORMAL FUNCTIONS
C  ARE READ FROM LARGEST TO SMALLEST RESPY.
C  SUBROUTINES PMPY AND PADD ARE FROM SSPUB FOR POLYNOMIAL
C  OPERATIONS.
C  THE NUMERATOR COEFFICIENTS ARE PRINTED OUT,B(1) COEF
C  FICIENT OF S**(N-1) ETC.
      DIMENSION  C(10),R(10),X(10),Y(10),Z(10),ZZ(10)
      READ(5,1)N
      READ(5,1)IR
 1    FORMAT(I2)
      READ(5,2)(C(I),I=1,N)
      READ(5,2)(R(I),I=1,N)
 2    FORMAT(10F20.7)
      WRITE(6,103)(C(I),I=1,N)
 103  FORMAT('0','C=',10F12.7)
      WRITE(6,104)(R(I),I=1,N)
 104  FORMAT('0','R=',10F12.7)
      Y(1)=1.
      IY=1
      ZZ(1)=0.
      IZZ=1
      NX=N-IR
      NX=NX/2
      DO 7 K=1,NX
      IC=N-2*K+2
      ICC=IC-1
      READ(5,1)IX
      READ(5,2)(X(I),I=1,IX)
      READ(5,2)(Z(I),I=1,IX)
      DO 3 I=1,IX
3     X(I)=C(IC)*X(I)+C(ICC)*Z(I)
      CALL PMPY(Z,IZ,X,IX,Y,IY)
      DO 4 I=1,IZ
4     X(I)=Z(I)
      IX=IZ
      CALL PADD(Z,IZ,X,IX,ZZ,IZZ)
      DO 5 I=1,IZ
5     ZZ(I)=Z(I)
      IZZ=IZ
      X(1)=R(IC)
      X(2)=R(ICC)
      X(3)=1.
      IX=3
      CALL PMPY(Z,IZ,X,IX,Y,IY)
      DO 6 I=1,IZ
6     Y(I)=Z(I)
      IY=IZ
7     CONTINUE
      IF(IR.EQ.O) GO TO 13
      DO 11 K=1,IR
      IC=IR-K+1
      READ(5,1)IX
      READ(5,2)(X(I),I=1,IX)
      DO 8 I=1,IX
8     X(I)=C(IC)*X(I)
      CALL PMPY(Z,IZ,X,IX,Y,IY)
      DO 9 I=1,IZ
9     X(I)=Z(I)
      IX=IZ
      CALL PADD(Z,IZ,X,IX,ZZ,IZZ)
      DO 10 I=1,IZ
10    ZZ(I)=Z(I)
      IZZ=IZ
      IF(IC.EQ.1)GO TO 13
      IX=2
      X(1)=R(IC)
      X(2)=1.
      CALL PMPY(Z,IZ,X,IX,Y,IY)
      DO 11 I=1,IZ
11    Y(I)=Z(I)
      IY=IZ
12    CONTINUE
13    DO 14 I=1,IZZ
      IK=IZZ-I+1
14    X(I)=ZZ(IK)
      WRITE(6,102)(X(I),I=1,IZZ)
102   FORMAT('0','B=',10F12.7)
      STOP
      END
      SUBROUTINE FRES(N,A,B,X,DB,PHASE)
C   THIS SUBROUTINE COMPUTES THE MAGNITUDE IN DB AND PHASE
C   IN RADIANS OF A RATIONAL APPROXIMATION.N-ORDER OF THE
C   FILTER.A-COEFFICIENTS OF THE DR.
C   A(1) IS THE COEFFICIENT OF THE TERM S**(N-1).B-NR COEF
C   FICIENTS SIMILAR TO A.
C   X-INPUT FREQUENCY IN RADIANS.DB-MAGNITUDE IN DECIBELS.
C   PHASE-ANGLE IN RADIANS
      DIMENSION A(10),B(10)
      DOUBLE PRECISION X,Y,Z,XX,YY,ZZ,A,B,DB,PHASE
 1    XX=X
      YY=0.
      ZZ=0.
      Y=B(N)
      Z=A(N)
      DO 7 K=1,N
      IF(K/2*2.EQ.K)GO TO 6
      IF(K.EQ.N)GO TO 10
      KK=N-K
      YY=YY+B(KK)*XX
      ZZ=ZZ+A(KK)*XX
      XX=XX*X
      GO TO 7
 6    KK=N-K
      XX=-XX
      IF(K.EQ.N)GO TO 10
      Y=Y+B(KK)*XX
      Z=Z+A(KK)*XX
      XX=XX*X
 7    CONTINUE
 10   IF(N/2*2.EQ.N)GO TO 8
      ZZ=ZZ+XX
      GO TO 9
 8    Z=Z+XX
 9    PHASE=DATAN2(YY,Y)
      PHASE=PHASE-DATAN2(ZZ,Z)
      IF(PHASE.GT.O.DO)PHASE=PHASE-2.DO*3.1415927DO
      Y=Y*Y+YY*YY
      Z=Z*Z+ZZ*ZZ
      DB=DSQRT(Y/Z)
      DB=20.DO*DLOG10(DB)
      RETURN
      END
```

APPENDIX II

LAPLACE TRANSFORMS OF ORTHONORMAL FUNCTIONS

$$1)\ f(t) = \frac{1}{\pi}\frac{\sin(t - \pi)}{(t - \pi)}\ ,\ t\varepsilon[\,0,\ 3\pi]$$

a) $N = 5$

$\Phi_1(s) = 1.25759/(s + 0.79076)$

$\Phi_2(s) = (1.25692s^2 - 0.99392s)/(s^3 + 1.58068s^2$
$+ 1.2027s + 0.45711)$

$\Phi_3(s) = (0.95564s - 0.75568)/(s^3 + 1.58068s^2$
$+ 1.2027s + 0.45711)$

$\Phi_4(s) = (0.96581s^4 - 1.52665s^3 + 1.16159s^2 - 0.44148s)$
$/(s^5 + 2.04708s^4 + 3.36562s^3 + 3.27161s^2$
$+ 1.92787s + 0.6517)$

$\Phi_5(s) = (1.1532s^3 - 1.82285s^2 + 1.38696s - 0.52714)$
$/(s^5 + 2.04708s^4 + 3.36562s^3 + 3.27161s^2$
$+ 1.92787s + 0.6517)$

b) $N = 8$

$\Phi_1(s) = 0.13665/(s + 0.00934)$

$$\Phi_2(s) = (2.05356s - 0.01917)/(s^2 + 2.11789s + 0.01969)$$

$$\begin{aligned}\Phi_3(s) = {} & (1.21684s^3 - 2.57713s^2 + 0.02396s)/(s^4 + \\ & 2.85824s^3 + 2.92808s^2 + 2.85343s + 0.02639)\end{aligned}$$

$$\begin{aligned}\Phi_4(s) = {} & (1.40881s^2 - 2.9837s + 0.02774)/(s^4 + 2.85824s^3 \\ & + 2.92808s^2 + 2.85343s + 0.02639)\end{aligned}$$

$$\begin{aligned}\Phi_5(s) = {} & (0.89578s^5 - 2.56035s^4 + 2.6229s^3 - 2.55604s^2 \\ & + 0.02364s)/(s^6 + 3.25945s^5 + 6.92989s^4 \\ & + 12.18866s^3 + 9.53106s^2 + 8.15732s + 0.07534)\end{aligned}$$

$$\begin{aligned}\Phi_6(s) = {} & (1.51359s^4 - 4.3262s^3 + 4.4319s^2 - 4.31892s \\ & + 0.03994)/(s^6 + 3.25945s^5 + 6.92989s^4 \\ & + 12.18866s^3 + 9.53106s^2 + 8.15732s + 0.07534)\end{aligned}$$

$$\begin{aligned}\Phi_7(s) = {} & (1.37417s^7 - 4.47905s^6 + 9.52287s^5 - 16.74933s^4 \\ & + 13.09733s^3 - 11.20957s^2 + 0.10354s)/(s^8 \\ & + 4.20363s^7 + 10.54344s^6 + 20.47894s^5 \\ & + 24.75407s^4 + 23.68999s^3 + 12.88635s^2 \\ & + 4.44377s + 0.04039)\end{aligned}$$

$$\begin{aligned}\Phi_8(s) = {} & (1.00609s^6 - 3.27931s^5 + 6.97211s^4 - 12.26291s^3 \\ & + 9.5891s^2 - 8.20701s + 0.0758)/(s^8 + 4.20363s^7 \\ & + 10.54344s^6 + 20.47894s^5 + 24.75407s^4 \\ & + 23.68999s^3 + 12.88635s^2 + 4.44377s + 0.04039)\end{aligned}$$

2) $f(t) = \frac{1}{\pi} \frac{\sin(t - 2\pi)}{(t - 2\pi)}$, $t \varepsilon [0, 4\pi]$

a) $N = 5$

$\Phi_1(s) = 0.92839/(s + 0.43095)$

$\Phi_2(s) = (1.09815s^2 - 0.47325s)/(s^3 + 1.03391s^2$
$+ 0.63181s + 0.1603)$

$\Phi_3(s) = (0.66974s - 0.28863)/(s^3 + 1.03391s^2$
$+ 0.63181s + 0.1603)$

$\Phi_4(s) = (0.9098s^4 - 0.94065s^3 + 0.57482s^2 - 0.14584s)/$
$(s^5 + 1.44778s^4 + 1.99896s^3 + 1.39288s^2$
$+ 0.65976s + 0.15056)$

$\Phi_5(s) = (0.88173s^3 - 0.91163s^2 + 0.5571s - 0.14134)/$
$(s^5 + 1.44778s^4 + 1.99896s^3 + 1.39288s^2$
$+ 0.65976s + 0.15056)$

b) $N = 8$

$\Phi_1(s) = 1.03028s/(s^2 + 0.53074s + 0.09566)$

$\Phi_2(s) = 0.31865/(s^2 + 0.53074s + 0.09566)$

$\Phi_3(s) = (1.04367s^3 - 0.55392s^2 + 0.09983s)/(s^4$
$+ 1.07537s^3 + 0.82017s^2 + 0.28321s + 0.04165)$

$$\begin{aligned}\Phi_4(s) = {} & (0.68871s^2 - 0.36553s + 0.06588)/(s^4 \\ & + 1.07537s^3 + 0.82017s^2 + 0.28321s + 0.04165)\end{aligned}$$

$$\begin{aligned}\Phi_5(s) = {} & (0.85498s^5 - 0.91942s^4 + 0.70123s^3 - 0.24214s^2 \\ & + 0.03561s)/(s^6 + 1.44087s^5 + 2.25332s^4 \\ & + 1.70148s^3 + 0.99823s^2 + 0.30979s + 0.04332)\end{aligned}$$

$$\begin{aligned}\Phi_6(s) = {} & (0.87196s^4 - 0.93768s^3 + 0.71515s^2 - 0.24695s \\ & + 0.03632)/(s^6 + 1.44087s^5 + 2.25332s^4 \\ & + 1.70148s^3 + 0.99823s^2 + 0.30979s + 0.04332)\end{aligned}$$

$$\begin{aligned}\Phi_7(s) = {} & (2.32891s^7 - 3.35565s^6 + 5.24779s^5 - 3.9626s^4 \\ & + 2.32479s^3 - 0.72148s^2 + 0.1009s)/(s^8 \\ & + 4.15279s^7 + 16.39557s^6 + 22.5592s^5 + 28.67467s^4 \\ & + 20.43117s^3 + 11.10009s^2 + 3.28816s + 0.44341)\end{aligned}$$

$$\begin{aligned}\Phi_8(s) = {} & (7.4506s^6 - 10.73531s^5 + 16.78857s^4 - 12.67705s^3 \\ & + 7.4374s^2 - 2.30815s + 0.3228)/(s^8 + 4.15279s^7 \\ & + 16.39557s^6 + 22.5592s^5 + 28.67467s^4 \\ & + 20.43117s^3 + 11.10009s^2 + 3.28816s + 0.44341)\end{aligned}$$