THE OPTIMIZATION OF A CLASS OF NON-LINEAR FILTERS.

Ph.D. Thesis by

J.K. LUBBOCK, B.Sc. (Eng).

5909084

ProQuest Number: 10803814

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.

ProQuest 10803814

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ABSTRACT

Some time invariant non-linear filters of Zadeh's class η , are optimized. A method is proposed for the physical realization of these filters as multipath structures, which consist of, in general, an orthonormal set of non-linear zero memory polynomial functions, each of which is followed by a linear memory network in cascade. The use of orthonormal polynomial functions instead of, for instance, power law functions simplifies the optimization. An almost routine optimization procedure is proposed and found to work in most practical cases. The theory is extended where necessary to enable its application to the following problems: Noise filtering; prediction; systems analysis; non-linear compensation of a feedback control system. Examples are given.

An expansion of second probability density functions in terms of the same orthonormal polynomials which are chosen for the non-linear filters is discussed. Methods of obtaining the form of the polynomials and the expansion coefficients both analytically and experimentally are proposed. The information, in the form required for the optimization procedure, is shown to consist of certain of the expansion coefficients of both the second probability density of the input and the joint probability density of the input and desired output.

Some special classes of random processes are defined and their properties are derived. A theorem is proved, showing that an optimum filter of class η_1 becomes linear when the input and desired output processes belong to a broad class, which includes Gaussian and many other special processes. Examples of processes of each classification are given and used in the optimization problems; expansion coefficients are worked out for these cases.

PREFACE

This thesis describes the research performed by the author during the period 1952-59, although the majority of this work has been done since the beginning of 1956. This thesis has not been submitted before as a thesis for a degree of any University, but some of the results have already been published in a paper³⁶ of the same title in the proceedings of the Institution of Electrical Engineers monograph No. 344E.

Originality is claimed for a substantial part of the work contained in chapters 3, 5, 6, 7, 8, 9, 10 and 11 as well as appendices 13.1 and 13.3, but in order to maintain continuity, work of other authors is included; an effort has been made to acknowledge such work in the text or by providing a reference.

The use of orthonormal polynomials^{*} in the multipath filter synthesis is original, and leads to some simplification of the optimization, which is reduced to an almost routine procedure which engineers can use. Zadeh in his paper entitled, "Optimum Non-linear Filters"² fails to

* The polynomials are orthogonal in the weaker sense that

$$\langle \theta_n[x(t)] \theta_m[x(t)] \rangle = 0 \quad n \neq m$$

They are not orthogonal in the wider sense that

$$\langle \theta_n[x(t_1)] \theta_m[x(t_2)] \rangle = 0 \quad n \neq m \text{ for all } t_1 \text{ and } t_2$$

(see references 3 and 4 on orthogonal representation of non-linear functionals.

solve a self-chosen example due to complexity and, as far as the author is aware, the examples in chapters 6, 7, 8 and 10 represent the first non-linear memory filters to have been optimized. A number of non-linear memory filter optimizations are performed to illustrate applications in noise filtering, prediction, system analysis and non-linear compensation; these applications, which are original, may be of significant practical value in the control field. It is considered that the collection, simplification and integration of information from a number of sources, in this work also represents some contribution to the advancement of the subject.

The author acknowledges with gratitude, the assistance of Mrs E.M. Chambers of the Engineering Laboratory, Cambridge, for doing a large part of the numerical computation, the assistance of Mr P.J. Lawrence also at the Engineering Laboratory, Cambridge, for preparing and running the programme on EDSAC and the assistance and generous hospitality of Dr Bernard Murphy and his staff at the European Computation Centre, Brussels, during the running of programmes on PACE. Acknowledgements are also due to the Director of the Mathematical Laboratory for permission to use E.D.S.A.C., Mr J.F. Coales, O.B.E., Reader in Engineering at Cambridge University and Dr d'Ombraïn, the

author's former supervisor at Battersea College of Technology and many colleagues, for their suggestions, encouragement and useful discussions.

J.K. Lubbock

J.K. Lubbock

20. 11. 59

CONTENTS

1.0 Introduction

2.0 Mathematical Preliminaries.

- 2.1 Probability density functions.
- 2.2 Classification of random processes.
- 2.3 Expectation.
- 2.4 The characteristic function.
- 2.5 Spectral densities of stationary signals.

3.0 An Expansion of Second Probability Density Functions.

- 3.1 A class of orthonormal polynomials.
- 3.2 Analytical and experimental determination of the orthonormal polynomials.
- 3.3 The expansion of second probability density functions.
- 3.4 Analytical and experimental determination of the expansion coefficients.

4.0 The Classification of Time-Invariant Non-linear Filters.

- 4.1 Characterization of the past.
  - 4.1.1 Wiener's method.
  - 4.1.2 Singleton's method.
- 4.2 The non-linear operator.
- 4.3 Bose's method.

4.3.1 Expansion of the filter function.

4.3.2 The optimization procedure.

4.3.3 Conclusion.

4.4 Zadeh's classification.

4.4.1 The class of attenuators.

4.4.2 The linear class.

4.4.3 The zero memory class.

4.4.4 Multipath zero memory class.

4.4.5 Multipath memory class.

5.0 The Optimization of some Filters of class γ_i .

5.1 Filters of class γ_i .

5.2 Zero memory non-linear filters.

5.3 Multipath zero-memory filters.

5.4 Multipath memory filters.

5.5 Linear attenuators.

5.6 Linear memory filters.

5.7 Applications.

6.0 Some special classes of Second Probability Density

Functions.

6.1 Class Λ .

6.1.1 The Gaussian Process.

6.1.2 An asymmetrical probability density of
class Λ .

6.2 Separable random processes and class Δ .

6.2.1 Some properties of separable processes.

6.2.2 A property of the expansion coefficients.

6.2.3 Linear transformation.

6.2.4 Class Δ .

6.3 Semi-separable random processes.

6.3.1 Some properties of semi-separable processes.

6.3.2 Classes Δ_1 and Δ_2 .

6.3.3 An example of a semi-separable process.

6.4 A theorem on optimum multipath filters.

7.0 Non-linear Noise Filtering.

7.1 Additive signal and noise.

7.2 Multiplicative signal and noise.

7.3 Problem 1. The optimization of some non-linear noise filters for a random telegraph signal and Gaussian noise with identical spectra.

7.3.1 A zero memory filter.

7.3.2 Multipath zero memory filters.

7.3.3 A multipath memory filter.

7.4 Problem 2. The optimization of some non-linear noise filters for a continuous input signal and Gaussian noise with identical spectra.

7.4.1 A multipath zero memory filter.

7.4.2 A multipath memory filter.

7.5 Conclusions.

8.0 Non-linear Prediction.

8.1 Problem 3. The optimization of some non-linear predictors for a smoothed random telegraph signal input.

8.1.1 The optimum linear predictor.

8.1.2 An optimum multipath zero memory predictor.

8.1.3 An optimum multipath memory predictor.

8.2 Conclusions.

9.0 The Analysis of Non-linear Systems.

9.1 Systems with a single input.

9.1.1 Best approximate attenuators.

9.1.2 Best approximate linear filters.

9.1.3 Problem 4. The best approximate linear filter for a zero memory non-linear system.

9.1.4 Best approximate non-linear filters.

9.2 Systems with two inputs.

9.2.1 Best approximate attenuators.

9.2.2 Problem 5. The best approximate attenuators for the zero memory filters of problems 1 and 2.

9.2.3 A linear memory filter and an attenuator.

9.2.4 Two linear filters.

9.2.5 A zero memory filter and an attenuator.

9.3 Application to feedback system analysis.

9.3.1 Problem 6. Linearization of a non-linear limiter.

10.0 The Non-linear Compensation of a Feedback Control System.

10.1 The Theory.

10.2 Problem 7.

11.0 Limitations and Extensions of the Theory.

11.1 Available information.

11.2 Non-stationary signals.

11.3 Other performance criteria.

11.4 Classification of processes.

11.5 Generalizations.

11.6 Generalization of the orthonormal polynomials.

11.7 Generalizations of the probability density
expansion.

11.7.1 The first generalization.

11.7.2 The second generalization.

11.8 Higher Order filters.

11.8.1 The first canonical form.

11.8.2 The second canonical form.

11.8.3 Some multi-input filters of class η_1 .

11.8.4 Wiener's filter using generalized
orthonormal polynomials.

12.0 References.

13.0 Appendices.

13.1 Analytical and experimental determination of the orthonormal polynomial for a clipped Gaussian signal.

13.1.1 The analytical method.

13.1.2 The experimental method.

13.1.3 The variance of the measured values of certain joint moments over finite samples.

13.2 Lampard's method for measuring correlation coefficients.

13.2.1 Laguerre polynomials.

13.2.2 Another set of orthonormal functions.

13.3 Solutions of the Wiener Hopf equation and simultaneous integral equations of Wiener Hopf type.

13.3.1 The solution of the Wiener Hopf equation.

13.3.2 The method of undetermined coefficients.

13.3.3 Solution in the time domain.

13.4 Some mathematical Notes.

13.4.1 Bilateral transforms.

13.4.2 Higher order delta functions.

LIST OF FIGURES.

- Fig.1 The measurement of the expansion coefficient of a joint second probability density function
- Fig.2 Singleton's delay line filter.
- Fig.3 Bose's experimental procedure for the determination of optimum non-linear memory filters.
- Fig.4 Synthesis of Bose's optimum non-linear memory filter.
- Fig.5 A multipath zero memory filter of degree R .
- Fig.6 A multipath memory filter of degree R .
- Fig.7 The cross-correlation function is invariant under non-linear transformation of the first variable.
- Fig.8 Separability is preserved under linear transformation of the second variable.
- Fig.9 Zero memory filter functions for problem 1.
- Fig.10 The performance index plotted against number of paths for problem 1.
- Fig.11 The optimum multipath memory filter for problem 1.
- Fig.12 The zero memory filter function for problem 2.
- Fig.13 The performance index plotted against number of paths for problem 2.
- Fig.14 Block diagram of the two-path memory filter for problem 2.
- Fig.15 Block diagram of the three-path memory filter for problem 2.
- Fig.16 The zero memory predictor function for problem 3.

- Fig.17 The performance index plotted against number of paths for the predictor of problem 3.
- Fig.18 Predictor mean-squared error as a function of prediction time T .
- Fig.19 An equivalent linear filter for a zero memory non-linear system of problem 4.
- Fig.20 Linearization of a non-linear feedback loop.
- Fig.21 A feedback system incorporating a limiter.
- Fig.22 The simulated system of fig.21 and the measuring circuits.
- Fig.23 Key to principal symbols for the analogue computing circuits.
- Fig.24 The filter functions for problem 6.
- Fig.25 Simulation of the approximate system.
- Fig.26)
- Fig.27) Pen recordings for the filter of problem 7.
- Fig.28)
- Fig.29)
- Fig.30 An optimum linear feedback system.
- Fig.31 Improvement of the performance of an optimum linear system by the addition of non-linear feedforward compensation.
- Fig.32 Suppression of the input $x(t)$ to $H(p)$.
- Fig.33 The final system with non-linear feedforward compensation.

Fig.34 Experimental measurement of various correlation coefficients needed in the design of the optimum compensating network C. (Case 1).

Fig.35

Fig.36

Fig.37 The circuit for generating a clipped Gaussian signal and the measurement of its moments.

Fig.38 Curves of $-\frac{100}{3}\theta_1$; $-17.323\theta_2$; $-12.072\theta_3$; $10.3988\theta_4$; $10.3600\theta_5$ taken on a variplotter.

Fig.39 Signal samples of $-\frac{100}{3}\theta_1$; $-17.323\theta_2$; $-12.072\theta_3$; $10.3988\theta_4$; and $10.3600\theta_5$.

Fig.40 Measurement of correlation coefficients by Lampard's method.

Fig.41 Synthesis of transfer functions $\frac{\alpha p^{n-1}}{(p+\alpha)^n}$ and $\frac{\alpha/2 (p-\alpha/2)^{n-1}}{(p+\alpha/2)^n}$

LIST OF PROBLEMS.

- Problem 1. Noise filter. Signal: random telegraph signal.
Noise: Gaussian.
- Problem 2. Noise filter. Signal: output of a square law
device with Gaussian input.
Noise: Gaussian.
- Problem 3. Predictor. Input: Smoothed random telegraph
signal.
- Problem 4. Best approximate linear filter for a non-linear
system.
- Problem 5. Best approximate attenuators for the memory
filters of problems 1 and 2.
- Problem 6. Experimental linearization of a feedback system.
- Problem 7. Non-linear compensation of a linear feedback
system.

1.0 INTRODUCTION

In communication engineering the term 'filter' is normally used to mean 'a linear frequency selective device'. Filters may be used to transmit certain frequency bands, whilst highly attenuating others; in this way it is possible to transmit unattenuated a desired component of input, whilst rejecting an unwanted component, providing their frequency spectra are essentially different (i.e. non-overlapping).

Even Although the input components may have identical spectra, selection of a desired component ^{from} amongst them can be performed, with perfection in favourable circumstances, by the identification of other essential dissimilarities in their characteristics, such as amplitude dissimilarities. However, a non-linear filter is required for this purpose since a linear filter can operate only on the basis of spectral differences.

Non-linear physically realizable filters can be logically (though perhaps not exhaustively) classified in a number of ways (see chapter 4). One of the most useful classifications is due to Zadeh^{1,2} who defines a hierarchy of classes each more general than its predecessor and each being more complicated to realize physically and ^{to} optimize than its predecessor. The simplest class in this hierarchy, known as class $\mathcal{N}_1$, is just one stage more complicated than the linear class; it is for this reason that it has been chosen as the object of investigations described in this thesis.

The term 'filter' shall be interpreted in a much broader sense, for this work, than the sense implied in current

communication literature; herein a filter is neither necessarily linear nor frequency sensitive. Furthermore, we do not imply that its function is restricted to the separation of input components, a device which performs that function shall be termed 'a noise filter'; by 'filter' we mean any device which operates upon its input in a physically realizable manner so as to produce a different output; thus even an attenuator and any non-linear device is a filter; we exclude only boxes with straight through connections. The following is a list of some devices which, for the purpose of this thesis, are classified as filters:-

Noise filters; predictors; compensation networks used, for example, in feedback control loops; encoders; decoders.

The last two of these devices will of necessity be non-linear whilst the others may be either linear or non-linear.

An optimum or best filter within some specified class is one which gives ^{the} least mean squared error of all filters of its class under actual working conditions. The mean squared error is given by the equation

$$\overline{e^2} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [z(t) - y(t)]^2 dt \quad (1.1)$$

where $z(t)$ is the desired output

$y(t)$ is the actual system output.

An optimum non-linear filter belonging to class η_1 ,

or one of its subclasses, provided the subclass is more general than the linear class, will give a mean squared error which is as small, or smaller, than the optimum linear filter. Practical application can be envisaged where this advantage will justify the additional complexity of the optimization procedure and physical form of the non-linear filter.

In chapter 2 some well-known statistical properties of random processes are defined; this chapter, which is intended primarily to introduce the notation, may be omitted by the reader who is well versed in statistical methods. Chapter 3 deals with an expansion of second probability density functions and shows how the coefficients of the expansion may be determined both analytically and by analogue measurements on long samples of the signal; these results are used in the optimization of multipath filters in chapter 5. Chapter 4 reviews other contributions in the field of non-linear filter classification and optimization. In chapter 5, some filters of class η_1 are optimized. In chapter 6 some special classes of second probability density functions are defined, their properties are investigated and some examples are given. These properties and examples are used in the filter applications of subsequent chapters. An interesting theorem is proved which shows that if the input and desired output processes

belong to two fairly general classes, then the optimum multipath filter is linear. Chapters 7,8,9 and 10 deal entirely with applications including additional theory and examples.

2.0 MATHEMATICAL PRELIMINARIES.

To quote from the introduction to this thesis:

"An optimum filter within some specified class is one which gives ^{the} least mean squared error of all filters of its class, under the actual input working conditions"; a filter which is optimum for one input signal will not necessarily be optimum for another different input signal. For this reason it is necessary to carry out an analysis on the actual working input signal in order to obtain enough information to enable the filter to be optimized.

Signals in practice may consist of varying voltages, currents, pressures, levels, flows, temperatures, viscosities, etc. but all such signals can be divided into two main classes:-

- i. Deterministic.
- ii. Non-deterministic or random.

A deterministic signal can be exactly specified a priori as a function of time, but for a random signal we can state only the probability of it having a value lying within a specified range at some particular time, thus whilst a deterministic signal can be completely defined a priori, a random signal can be only incompletely characterized a priori by probability density functions or statistical averages and there may be an infinite number of different signal samples which fit the statistical properties. For the sake of generality and because random signals occur abundantly in nature, signal samples are more conveniently

analysed in terms of their statistical properties; furthermore all the information, about the input signal and desired output, used in the filter optimization procedure proposed herein may be conveniently expressed in the form of statistical averages. In this and the following chapter, some statistical properties are defined and it is shown how they may be obtained from long signal samples by analogue measurements.

2.1 Probability Density Functions.

Statistical properties, whether they be probability density functions or statistical averages must be defined over an 'ensemble' of signal samples, each member of which may be considered as emanating from a number of identical sources working under identical external conditions. If the sources are time-dependent (involve some parameter which is a function of time) we can reasonably expect the statistical properties of the signal to change with time. The ensemble of signals collectively referred to as a random process will usually be impossible to generate in practice; but for stationary signals, by invoking the ergodic hypothesis, an ensemble can be built up by collecting a large number of samples from the same source at different times.

Let $x(t)$ be a signal sample belonging to a continuous⁺ random process $X(t)$ then:

+ The definitions which follow need some modification to cover the case of discrete time series.

The first probability function $F_1(x, t)$ is defined as the probability that $x \leq x_1$ at $t = t_1$

The first probability density function $p_1(x, t)$ is defined by the equation

$$p_1(x, t) = \frac{d}{dx} F_1(x, t) \quad (2.1)$$

Therefore the probability that $x_1 < x \leq x_1 + \delta x_1$ at $t = t_1$ is given by:

$$\int_{x_1}^{x_1 + \delta x_1} p_1(x, t) dx \simeq p_1(x, t) \delta x_1 \quad (2.2)$$

The second probability function $F_2(x_1, x_2; t_1, t_2)$ is defined as the probability that $x \leq x_1$ at $t = t_1$ and $x \leq x_2$ at $t = t_2$

The second probability density function $p_2(x, x'; t, t')$ is defined by the equation:

$$p_2(x, x'; t, t') = \frac{\partial^2 F_2(x, x'; t, t')}{\partial x \partial x'} \quad (2.3)$$

Therefore the probability that $x_1 < x \leq x_1 + \delta x_1$ when $t = t_1$ and $x_2 < x \leq x_2 + \delta x_2$ when $t = t_2$

is given by:

$$\int_{x_1}^{x_1 + \delta x_1} dx_1 \int_{x_2}^{x_2 + \delta x_2} dx_2' p_2(x, x'; t_1, t_2) \simeq p_2(x_1, x_2; t_1, t_2) \delta x_1 \delta x_2 \quad (2.4)$$

The joint second probability function between two random variables $x(t)$ and $z(t)$ belonging to two random processes X and Z i.e. $F_2(x_1, z_1; t_1, t_2)$ is defined as the probability that $x \leq x_1$ at $t = t_1$ and $z \leq z_2$ at $t = t_2$

Similarly the joint second probability density function

$p_2(x, z; t, t')$ is defined by the equation:

$$p_2(x, z; t, t') = \frac{\partial^2 F_2(x, z; t, t')}{\partial x \partial z} \quad (2.5)$$

If the two processes X and Z are statistically independent then

$$p_2(x, z; t, t') = p_1(x, t) p_1(z, t') \quad (2.6)$$

Higher probability functions and probability density functions may be defined in a similar manner.

The following property of probability density functions may be proved from their definition.

$$p_k(x_1, x_2, \dots, x_k) = \underbrace{\int \dots \int}_{n=k} dx_{k+1} \dots dx_n p_n(x_1, x_2, \dots, x_n) \quad (2.7)$$

2.2 Classification of random processes.

The probability functions of random processes defined so far, are not distinct in the sense that both probability functions and probability density functions of order k

can be derived from the corresponding function of order n where $n > k$ (see equation 2.7). Thus we assume that each probability function of order n is more general than, and conveys more descriptive information about the process than one of order k ($n > k$). However, there are classes of process for which this assumption is untrue.

A purely random process is one in which successive values of are statistically independent. For a process of this kind, prediction of future values from past history is impossible. It follows that:

$$p_2(x_1, x_2; t_1, t_2) = p_1(x_1, t_1) p_1(x_2, t_2) \quad (2.8)$$

Similarly:

$$\begin{aligned} p_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) \\ = p_1(x_1, t_1) p_1(x_2, t_2) \dots p_1(x_n, t_n) \end{aligned} \quad (2.9)$$

Thus higher probability density functions convey no more statistical descriptive information about the process than the first probability density.

The next more general class of process is called a Markov process. For this class it is possible to express all available statistical descriptive information in the form of the second probability density function, since higher probability densities convey no more information.

$$p_3(x_1, t_1; x_2, t_2; x_3, t_3) = \frac{p_2(x_1, t_1; x_2, t_2) p_2(x_2, t_2; x_3, t_3)}{p_1(x_2, t_2)} \quad (2.10)$$

In terms of conditional probabilities:

$$\begin{aligned} p_n(x_1, t_1; x_2, t_2 \cdots x_{n-1}, t_{n-1} | x_n, t_n) \\ = p_2(x_{n-1}, t_{n-1} | x_n, t_n) \end{aligned} \quad (2.11)$$

where $p_2(x_1, t_1; x_2, t_2 \cdots x_{n-1}, t_{n-1} | x_n, t_n)$ is the probability density of $x = x_n$ at $t = t_n$ given $x = x_1$ at $t = t_1$, $x = x_2$ at $t = t_2$ ----- $x = x_{n-1}$ at $t = t_{n-1}$.

2.3 Expectation.

The average value of, or expected value of $x(t)$ is defined as :

$$E[x(t)] = \int_{-\infty}^{\infty} x p_1(x, t) dx = \langle x(t) \rangle \quad (2.12)$$

Similarly:

$$E[x^2] = \int_{-\infty}^{\infty} x^2 p_1(x, t) dx = \langle x^2(t) \rangle \quad (2.13)$$

$$E[x^n] = \int_{-\infty}^{\infty} x^n p_1(x, t) dx = \langle x^n(t) \rangle \quad (2.14)$$

and $E[g(x)] = \int_{-\infty}^{\infty} g(x) p_1(x, t) dx = \langle g(x) \rangle$ (2.15)

The variance is defined as:-

$$E[(x - \langle x \rangle)^2] = \int_{-\infty}^{\infty} (x - \langle x \rangle)^2 p_1(x, t) dx = \langle x^2(t) \rangle - \langle x(t) \rangle^2 \quad (2.16)$$

The standard deviation is defined as:

$$\sqrt{E[(x - \langle x \rangle)^2]} = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \sigma \quad (2.17)$$

All of these averaged values are functions of time for the general non-stationary case.

For the two processes X and Z we may define the joint moments:

$$\alpha_{nm}(t_1, t_2) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dz, x^n z^m p_2(x, z; t_1, t_2) = \langle x^n z^m \rangle \quad (2.18)$$

where suffixes of x and z indicate instants in time

$$\text{i.e. } \langle x^n z^m \rangle = \langle x^n(t_1) z^m(t_2) \rangle$$

whence

$$\begin{aligned} \alpha_{n0}(t) &= \langle x^n \rangle & \alpha_{0m} &= \langle z^m \rangle \\ \alpha_{10}(t) &= \langle x \rangle & \alpha_{01} &= \langle z \rangle \end{aligned} \quad (2.19)$$

Central moments are defined by:

$$\mu_{nm}(t_1, t_2) = E[(x - \alpha_{10})^n (z - \alpha_{01})^m] \quad (2.20)$$

whence the variances of $x(t)$ and $z(t)$ are given by:

$$\mu_{20}(t) = \sigma_x^2(t) \quad \text{and} \quad \mu_{02}(t) = \sigma_z^2(t) \quad (2.21)$$

also

$$\mu_{11}(t_1, t_2) = E[(x_1 - d_{10})(z_2 - d_{01})] = \alpha_{11}(t_1, t_2) - d_{10}(t_1)d_{01}(t_2) \quad (2.22)$$

which is known as the covariance or cross-correlation of x and z , whilst if x and z belong to the same process μ_{11} is then known as an auto-correlation coefficient^{*}. For correlation coefficients the symbol ϕ is commonly used.

It can be shown that the auto-correlation coefficient of the sum of a number of random signals is equal to the sum of all their auto and cross-correlation coefficients.

$$\phi_{yy}(t_1, t_2) = \sum_{n=1}^N \sum_{m=1}^N \phi_{x_n x_m}(t_1, t_2) \quad (2.23)$$

where

$$y(t) = x_1(t) + x_2(t) + \dots + x_N(t)$$

The normalized correlation coefficient is often given the symbol ρ

$$\rho(t_1, t_2) = \frac{\mu_{11}(t_1, t_2)}{\sqrt{\mu_{20}(t_1)\mu_{02}(t_2)}} \quad (2.24)$$

All these joint moments for two processes x and z are functions of t_1 and t_2 in the general case of non-stationary processes.

---

* $\alpha_{11} = E[xz]$ is commonly referred to as the "correlation function".

If the processes are ergodic, then all the above expectations can be redefined in terms of time averages; the moments can be defined in general terms by:

$$\begin{aligned} \mu_{nm}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \left[x(t) - \overline{x(t)} \right]^n \left[z(t+\tau) - \overline{z(t)} \right]^m dt \\ &= \overline{\left[x(t) - \overline{x(t)} \right]^n \left[z(t+\tau) - \overline{z(t)} \right]^m} \end{aligned} \quad (2.25)$$

where the bar is used to denote a time average.

Some useful properties of auto and cross-correlation coefficients of ergodic signals are listed below:

1. $\phi_{xx}(\tau) = \phi_{xx}(-\tau)$
2. $\phi_{xx}(0) = \sigma_x^2$
3. $|\phi_{xx}(\tau)| \leq \phi_{xx}(0)$
4. $\phi_{xy}(\tau) = \phi_{yx}(-\tau)$

2.4 The characteristic function.

The characteristic function of a random variable is defined by:-

$$\psi(u, t) = E[e^{jux}] = \int_{-\infty}^{\infty} e^{jux} p(x, t) dx \quad (2.26)$$

From the properties of Fourier transforms it follows that:

$$p(x, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-jux} \psi(u, t) du \quad (2.27)$$

Again from the properties of Fourier transforms it is easily shown that:

$$\text{if } x(t) = \sigma(t) + u(t) \quad (2.28)$$

$$\text{then } \psi_x(u, t) = \psi_\sigma(u, t) \times \psi_u(u, t)$$

The characteristic function may be expanded in terms of the moments d_n using MacLaurin's series:

$$\psi(u, t) = 1 + \sum_{n=1}^{\infty} \frac{(ju)^n}{n!} d_n(t) \quad (2.29)$$

Thus, provided $\psi(u, t)$ exists, the moments $d_n(t)$ uniquely determine the first probability density function $p_1(u, t)$

The characteristic function of two random variables x and z is defined by:

$$\psi(u_1, u_2; t_1, t_2) = E[e^{j(u_1 x + u_2 z)}] = \iint_{-\infty}^{\infty} e^{j(u_1 x + u_2 z)} p_2(x, z; t_1, t_2) dx dz \quad (2.30)$$

and

$$p(x, z; t_1, t_2) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-j(u_1 x + u_2 z)} \psi(u_1, u_2; t_1, t_2) du_1 du_2 \quad (2.31)$$

Similarly, it can be shown that

$$\psi(u_1, u_2; t_1, t_2) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{\alpha_{nm}(t_1, t_2)}{n! m!} (ju_1)^n (ju_2)^m \quad (2.32)$$

The infinite set of joint moments $\alpha_{nm}(t_1, t_2)$ uniquely determine the joint second probability density function $p(x, z; t_1, t_2)$.

2.5 Spectral densities of stationary signals.

The Fourier transform of an auto-correlation coefficient is termed the ~~power~~ spectral density and is defined by:

$$\Phi_{xx}(j\omega) = \int_{-\infty}^{\infty} \phi_{xx}(\tau) e^{-j\omega\tau} d\tau \quad (2.33)$$

It follows that:

$$\phi_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{xx}(j\omega) e^{j\omega\tau} d\omega \quad (2.34)$$

Since $\phi_{xx}(\tau)$ is an even function of τ , then $\Phi_{xx}(j\omega)$ is a real function of frequency.

The Fourier transform of a cross-correlation function is termed the cross-spectral density and is defined by:

$$\Phi_{xy}(j\omega) = \int_{-\infty}^{\infty} \phi_{xy}(\tau) e^{-j\omega\tau} d\tau \quad (2.35)$$

Likewise:

$$\phi_{xy}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{xy}(j\omega) e^{j\omega\tau} d\omega \quad (2.36)$$

Since $\phi_{xy}(\tau)$ is not necessarily an even function of τ
 $\Phi_{xy}(j\omega)$ may be a complex function of ω

3.0 AN EXPANSION OF SECOND PROBABILITY DENSITY FUNCTIONS.

It can be shown (see chapter 5) that sufficient information, needed for the optimization of the class of non-linear filters considered by this paper, is contained in the second probability density of the input and the joint second probability density of the input and desired output. In practice, samples of the input and desired output may be available but their statistical description in the form of second probability densities will not usually be known. The coefficients of an expansion of the second probability density can be obtained by analogue measurements on long signal samples; this expansion method of expressing the required information is most convenient for the optimization of the particular canonical form of filter (the multipath filter) under consideration.

3.1 A class of orthonormal polynomials.

First we define a class of orthonormal polynomials used in the series expansion of second probability density functions.

Let $x(t)$ be a random function of time belonging to the ensemble X for which the first probability density function is $p(x, t)$.

We may define a set of polynomials $\theta_n^{\wedge}[x, t]$ which are orthonormal, with respect to $p(x, t)$ as weighting function, over the whole range of x :

i.e.*

$$\int \theta_n [x, t] \theta_m [x, t] p(x, t) dx = \delta_{nm}$$

where $\delta_{nm} = 1 \quad n = m$
 $= 0 \quad n \neq m$

(3.1)

The first few of the polynomials are easily identified since:

$$\left. \begin{aligned} \int p(x, t) dx &= 1 \quad ; \quad \int x p(x, t) dx = \langle x(t) \rangle \\ \text{where } \langle x(t) \rangle &\text{ is the ensemble average of } x \text{ at time } t. \\ \text{and } \int (x - \langle x \rangle)^2 p(x, t) dx &= \sigma_x^2(t) \\ \text{where } \sigma_x^2(t) &\text{ is the variance of } x \text{ at time } t. \end{aligned} \right\} \quad (3.2)$$

Thence we may write:

$$\begin{aligned} \int 1 \cdot 1 p(x, t) dx &= 1 \\ \int \frac{x - \langle x \rangle}{\sigma_x} \cdot 1 p(x, t) dx &= 0 \\ \int \frac{(x - \langle x \rangle)^2}{\sigma_x^2} p(x, t) dx &= 1 \end{aligned} \quad (3.3)$$

---

* Throughout this thesis, for each integral written without limits, it is implied that integration takes place over the whole range of the variable of integration.

thus we conclude that the first two polynomials are:

$$\theta_0[x, t] = 1$$

$$\theta_1[x, t] = \frac{x(t) - \langle x(t) \rangle}{\sigma_x(t)} \quad (3.4)$$

In order to obtain an expression for $\theta_2[x, t]$ we write it in the form:

$$\theta_2[x, t] = a_{22}(t)x^2(t) - a_{21}(t)\theta_1[x, t] - a_{20}(t)$$

in which the variable t is omitted,

The following three simultaneous equations, may then be solved to obtain the unknowns a_{22} , a_{21} and a_{20} :

$$\int \theta_2(x) \theta_1(x) p(x) dx = \int [a_{22}x^2 - a_{21}\theta_1(x) - a_{20}] \theta_1(x) p(x) dx = 0$$

$$\int \theta_2(x) \theta_0(x) p(x) dx = \int [a_{22}x^2 - a_{21}\theta_1(x) - a_{20}] p(x) dx = 0$$

and

$$\int \theta_2^2(x) p(x) dx = \int [a_{22}x^2 - a_{21}\theta_1(x) - a_{20}]^2 p(x) dx = 1$$

whence

$$\theta_2(x) = \frac{x^2 - \langle x^2 \theta_1(x) \rangle \theta_1(x) - \langle x^2 \rangle}{[\langle x^4 \rangle - \langle x^2 \theta_1(x) \rangle^2 - \langle x^2 \rangle^2]}^{1/2} \quad (3.5)$$

By proceeding in this way, successive polynomials may be obtained. The recurrence relation for the general case becomes:

$$\theta_n(x) = \frac{x^n - \sum_{r=0}^{n-1} \langle x^n \theta_r(x) \rangle \theta_r(x)}{\left[\langle x^{2n} \rangle - \sum_{r=0}^{n-1} \langle x^n \theta_r(x) \rangle^2 \right]^{1/2}} \quad (3.7)$$

For non-stationary signals θ_n and all the ensemble averages are functions of time.

It appears therefore that the set of orthonormal polynomials always exists if the integrals $\int x^n p(x) dx$ converge for all integer values of $n > 0$, this is not a very severe restriction in practice.

3.2 Analytical and experimental determination of the orthonormal polynomials.

Equation 3.6 may be used for the analytical determination of the orthonormal polynomials provided the $2n$ moments

$$\alpha_{r0}(t) = \langle x^r(t) \rangle, \quad r = 0, 1, 2, \dots, 2n, \quad \text{are known,}$$

since on substitution in equation 3.6 for $\theta_r[x, t]$ all the coefficients of x^r , $r < n$, in the polynomial for $\theta_n[x, t]$ become linear summations of the moments $\alpha_{r0}(t)$. These moments can be easily derived from the characteristic function, since from equation 2.29 we get:

$$\alpha_{r0}(t) = (-1)^{-r/2} \left| \frac{d^r}{du^r} \psi(u, t) \right|_{u=0} \quad (3.7)$$

Also they may be obtained directly from equation 2.14 if the integral is easy to evaluate. Thus the analytical procedure for determining the orthonormal polynomials is not difficult provided the moments or the characteristic function are known, ~~and it may also be possible if given the first probability density function of the process.~~

If the process for which samples are obtained is assumed to be ergodic, then ensemble averages may be replaced by time averages and equation 3.6 may be rewritten:

$$\theta_n(x) = \frac{x^n - \sum_{r=0}^{n-1} \overline{x^r \theta_r(x)} \theta_r(x)}{\left[\overline{x^{2n}} - \sum_{r=0}^{n-1} \overline{x^r \theta_r(x)^2} \right]^{1/2}} \quad (3.6)$$

where the bar above indicates a time average. Again, by substitution for $\theta_r(x)$ in the above equation, it is apparent that all the time averages can be expressed in terms of the $2n$ moments $\alpha_{r0} = \overline{x^r} \quad r=0, 1, 2, \dots, 2n$. These moments can be measured on an analogue computer given only a long sample of the signal. The computing equipment consists of:

1. A set of non-linear function generators capable of generating the functions

$$x, x^2, x^3, \dots, x^{2n}.$$

These may be generated conveniently using one servo type multiplier having $(2n + 1)$ potentiometers connected to its output shaft.

2. An integrator.

The procedure needs no description. Some results obtained on the PACE^{*} analogue computer in Brussels for a clipped Gaussian signal are recorded in appendix 13.1 together with the answers obtained analytically. These two sets of results are identical within experimental accuracy.

3.3 The expansion of a second probability density function.

The following expansion for the joint second probability density $p_2(x, z; t_1, t_2)$ is used in the optimization of multipath memory filters (chapter 5).

$$p_2(x, z; t_1, t_2) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} D_{nm}(t_1, t_2) \theta_n[x, t_1] \gamma_m[z, t_2] p(x, t_1) q(z, t_2) \quad (3.8)$$

where x and z are random variables belonging to the ensembles X and Z respectively and $\theta_n[x, t_1]$ and $\gamma_m[z, t_2]$ are sets of orthonormal polynomials with respect to the first probability density functions $p(x, t_1)$ and $q(z, t_2)$ over the whole range of x and z respectively.

The necessary restrictions on $p_2(x, z; t_1, t_2)$ for the expansion to be possible have not been fully investigated but it is expected and, within a limited experience it has been found, that these are not serious

in practice.

Using the orthogonality of $\theta_n [x, t_1]$ and $\gamma_m [z, t_2]$ we may write:

$$D_{nm}(t_1, t_2) = \iint \theta_n [x, t_1] \gamma_m [z, t_2] p_2(x, z; t_1, t_2) dx dz \quad (3.9)$$

From equations 3.4 we can show that:

$$D_{00} = 1 \quad ; \quad D_{0n} = D_{n0} = 0$$

$$\text{and } D_{11}(t_1, t_2) = \frac{\langle (x(t_1) - \langle x(t_1) \rangle)(z(t_2) - \langle z(t_2) \rangle) \rangle}{\sigma_x(t_1) \sigma_z(t_2)} = \rho_{xz}(t_1, t_2) \quad (3.10)$$

which is the normalized correlation coefficient between x and z .

Certain other properties of these coefficients exist if x and z are members of the same ensemble:

$$\begin{aligned} D_{nm}(t_1, t_2) &= D_{mn}(t_1, t_2) \\ D_{nn}(t_1, t_1) &= \delta_{nn} \end{aligned} \quad (3.11)$$

3.4 Analytical and experimental determination of the expansion coefficients.

By repeated substitution of equation 3.6 for $\theta_n [x, t_1]$ and $\gamma_s [z, t_2]$ in equation 3.9, the coefficients $D_{nm}(t_1, t_2)$ can be expressed in terms of joint moments

of the form:

$$\alpha_{rs}(t_1, t_2) = \langle x^r(t_1) z^s(t_2) \rangle = \iint x^r z^s p_2(x, z; t_1, t_2) dx dz \quad (3.12)$$

From equation 2.32

$$\alpha_{rs}(t_1, t_2) = (-1)^{\frac{-(r+s)}{2}} \left[\frac{\partial^r}{\partial u_1^r} \frac{\partial^s}{\partial u_2^s} \psi_{xz}(u_1, u_2; t_1, t_2) \right]_{\substack{u_1=0 \\ u_2=0}} \quad (3.13)$$

Therefore the coefficients $D_{nm}(t_1, t_2)$ can all be determined analytically from the joint moments $\alpha_{rs}(t_1, t_2)$ or the second joint characteristic function $\psi_{xz}(u_1, u_2; t_1, t_2)$ or the second probability density function $p_2(x, z; t_1, t_2)$ if its two dimensional Fourier transform can be found.

Whilst this analytical procedure is rather tedious it is not in any way difficult, whereas direct evaluation of the double integral of equation 3.9 is often impossible except by numerical methods.

The coefficients $D_{nm}(t_1, t_2)$ possess the physical significance that they are in fact the cross correlation coefficients of the outputs of non-linear function generators θ_n and γ_m , having connected at their inputs the signals $x(t)$ and $z(t)$ respectively. However, in order to be able to measure these coefficients experimentally, it is necessary to assume that the process

is ergodic so that ensemble averages can be replaced by time averages and equation 3.9 can be rewritten in the form:

$$D_{nm}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \theta_n[x(t)] \gamma_m[z(t+\tau)] dt$$

$$= \overline{\theta_n[x(t)] \gamma_m[z(t+\tau)]} \quad (3.14)$$

For the measurement of these coefficients, given a long sample of both $x(t)$ and $z(t)$ the following analogue computing equipment is required: (see fig.1).

1. A set of non-linear function generators capable of generating the polynomials $\theta_r[x]$ and $\gamma_s[z]$ $r=1,2,\dots,n$
 $s=1,2,\dots,m$

These can consist of

- a. $(n + m - 2)$ multipliers or one servo multiplier with $(n + m - 2)$ output potentiometers.
  - b. $(n+m-2)$ adding amplifiers.
  - c. A maximum of $\sum_{r=1}^n (r+1) + \sum_{s=1}^m (s+1)$ attenuators.
2. A cross correlator.

In the absence of apparatus capable of generating finite delays, the method of measuring correlation coefficients due to Lampard (appendix 13.2) is recommended. This method can be used with the linear orthogonal functions of Laning and Battin⁶ (see appendix 13.2).

Experimental results to demonstrate the method have not been obtained but since correlation coefficients have

been measured to within a few percent by Lampard's method (see Lampard's paper)⁵ and it has been demonstrated that the orthonormal polynomials can be determined and synthesized on an analogue machine to within a few per cent, it is considered that the expansion coefficients could be obtained with reasonable accuracy (say 5%) given:

- a. A sufficiently large and accurate computer such as PACE.
- b. An operator with sufficient ability and patience.
- c. Sufficient machine time.

To enable the expansion of the second probability density to be written down entirely from experimental measurement, the first probability density must also be obtained; equivalent information can be obtained experimentally in the form of the moments $\alpha_{r_0} = \overline{x^r}$ but the actual probability density function cannot be obtained; the best that can be done is to obtain a series approximation to the density function and the most obvious series seems to be the Gram-Charlier expansion; however, such an expansion is perhaps inconvenient and, since this information is not needed in the filter optimization, it will not be discussed.

4.0 CLASSIFICATIONS OF TIME INVARIANT NON-LINEAR FILTERS.

An obvious method of simplifying the problem of optimization is to put some restriction on the class of filters under consideration. The choice of the class of filters should logically depend upon the available information (see chapter 11) but in this thesis a class of non-linear filters has been chosen for the simplicity of its optimization and physical realization. Wiener⁴, Singleton⁷ and Bose⁸ have considered some other classes of non-linear filters, for which optimization procedures have been devised. Wiener⁴, Singleton⁷ and Zadeh¹² have produced classifications of non-linear filters which consist of logical hierarchies of classes in which each more general class is more difficult to optimize, more complex to synthesize and capable of giving better performance than its predecessor. The salient points of these authors' work is summarized in the first part of this chapter, in the later sections some subclasses of Zadeh's class 7, are defined prior to their optimization in chapter 5.

4.1 Characterization of the past.

The past of the filter input can be characterized in a number of ways. The following three methods can all be used as a basis on which to classify non-linear filters.

4.1.1 Wiener's Method.

The past input is characterized by a set of orthonormal

functions in terms of which it can be expanded. Wiener chose Laguerre functions for this purpose because of the simplicity with which they can be realized as delta function responses of linear networks (see appendix 13.2). The expansion of the past in terms of Laguerre functions becomes:

$$x(-\tau) = \sum_{n=1}^{\infty} u_n L_n(\tau) e^{-\tau/2} \quad \tau \geq 0 \quad (4.1)$$

where u_n are the Laguerre coefficients of the past.

$L_n(t)$ is the n th generalized Laguerre function defined by:

$$L_n(t) = \frac{1}{(n-1)!} e^t \frac{d^{n-1}}{dt^{n-1}} (t^{n-1} e^{-t}) \quad (4.2)$$

Using the orthonormal properties of $L_n(\tau) e^{-\tau/2}$ over the range $0 - \infty$ we obtain:-

$$u_n = \int_0^{\infty} x(-\tau) L_n(\tau) e^{-\tau/2} d\tau \quad (4.3)$$

For a continuous input $x(t)$, the past is continuously characterized by:
 $u_n(t) = \int_0^{\infty} x(t-\tau) L_n(\tau) e^{-\tau/2} d\tau$ which can be generated by a linear network.
 (see appendix 13.2 eqn 13.2.12)

4.1.2 Singleton's method.

An input which is band limited and contains no frequencies higher than W , can be completely characterized by its sampled values $u_1, u_2, \dots, u_s, \dots$ at intervals of $T = \frac{1}{2W}$. This is the well-known theorem of Shannon⁹. Thus in this case the characteristic coefficients become:

$$\begin{aligned}
 u_1 &= x(t - T) \\
 u_2 &= x(t - 2T) \\
 &\vdots \\
 u_s &= x(t - sT)
 \end{aligned}
 \tag{4.4}$$

Even for signals which are not band limited, provided their spectral density tends to zero as frequency tends to infinity (true of all physical signals) a good characterization can be obtained by taking a short enough sampling interval.

4.1.3 The past behaviour of a continuous signal can also be characterized by its present value and higher derivatives.¹⁰

Thus

$$\begin{aligned}
 u_1 &= x(t) \\
 u_2 &= \frac{dx}{dt} \\
 &\vdots \\
 u_s &= \frac{d^s x}{dt^s}
 \end{aligned}
 \tag{4.5}$$

4.2 The non-linear operator.

Any quantity which depends on the past behaviour of a time function can be expressed as a function of a set of coefficients which characterize the past:

$$y(t) = F(u_1, u_2, \dots, u_s, \dots, u_n) \tag{4.6}$$

For physical realizability, the number of coefficients must be finite and then it is easily shown, by minimizing the

mean squared error using the general method described in chapter 5, that the optimum filter function of order n is given by:

$$F(u_1, u_2, \dots, u_n) = \frac{\int_{-\infty}^{\infty} z p_{n+1}(z, u_1, u_2, \dots, u_n) dz}{p_n(u_1, u_2, \dots, u_n)} \quad (4.7)$$

where $z(t)$ is the filter desired output.

This formula has the obvious interpretation that, for a given set of values for u , $F(u_1, u_2, \dots, u_n)$ is equal to the mean of all the possible corresponding values of z .

The simplicity of equation (4.7) is due to the generality of the form of the filter function, but its evaluation is often impossible and to overcome this difficulty, Singleton uses a power series expansion for the filter function F ; the optimum coefficients of the series are then given by the solution of a set of simultaneous equations and the information needed for the optimization procedure involves higher order correlation functions, which are difficult to measure experimentally. A block diagram of the filter is shown in figure 2.

The filter function equation (4.6) defines a class of n th order filters. Successively higher classes are more general and capable of using more information about the input (see equation 4.7) than those of lower order. Thus a hierarchy of classes is defined by both the Wiener and Singleton approaches.

4.3 Bose's method.

Bose adopts essentially the same approach as Wiener and produces an optimization procedure; this consists of obtaining a set of coefficients which characterize the optimum time-invariant non-linear filter and then synthesizing the filter from a knowledge of its characterizing coefficients. An operator relating output to the past input of the filter is defined in such a way that the characterizing coefficients can be evaluated experimentally. The method is confined to those non-linear filters for which the present behaviour depends less and less upon the remote past of the input and only continuous inputs are considered.

4.3.1 Expansion of the filter function.

The filter function of equation (4.6) can be expanded in a convenient form in terms of functions which are orthonormal in the following sense: (see chapter 3)

$$\int_{-\infty}^{\infty} \theta_j(u_s) \theta_k(u_s) \rho(u_s, t) du_s = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

The expansion is:

$$y(t) = \lim_{S \rightarrow \infty} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \dots \sum_{h=1}^{\infty} a_{i,j,\dots,h} \theta_i(x_1) \theta_j(x_2) \dots \theta_h(u_s) \quad (4.8)$$

Wiener originally suggested using orthonormal Hermite functions for the $\theta_i, \theta_j, \dots, \theta_h$; however, the optimization procedure was found to be cumbersome except in the case of a

shot noise input, which gives statistically independent Gaussian signals for $u_1, u_2, \dots, u_n, \dots$

To overcome this difficulty, Bose used the following functions which, by definition, are orthogonal in time, with respect to any bounded weighting function $G(t)$, irrespective of their input:

$$\phi_j(u) \begin{cases} = 1 & u_j - \frac{w}{2} \leq u < u_j + \frac{w}{2} & j = (1, 2, \dots, n-1) \\ & u_j - \frac{w}{2} \leq u \leq b & j = n \\ = 0 & \text{for other values} & \\ & \text{of } u. & \end{cases}$$

$$w = \frac{b-a}{n}$$

$$u_j = a + w \left(j - \frac{1}{2} \right)$$

(4.9)

Therefore

$$G(t) \phi_j[u(t)] \phi_k[u(t)] = \begin{cases} 0 & j \neq k \\ G(t) \phi_j^2[u(t)] & j = k \end{cases} \quad (4.10)$$

Any zero memory filter function can be synthesized with any desired accuracy, using a canonical form represented by the following expansion in terms of the functions ϕ_j

$$y(t) = \sum_{j=1}^n a_j \phi_j[x(t)] \quad (4.11)$$

4.3.2 The optimization procedure.

We write $\Phi_\alpha(\alpha)$ to represent the product function $\phi_i(u_1) \phi_j(u_2) \dots \phi_h(u_s)$ and A_α to represent the corresponding coefficient $a_{ij\dots h}$ then the expansion of the filter function (equation 4.6) with only s inputs becomes:

$$y(t) = \sum_{\alpha} A_{\alpha} \Phi_{\alpha}(\alpha) \quad (4.12)$$

Minimization of the weighted mean squared error.

$$\bar{e}^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T G(t) \left\{ z(t) - \sum_{\alpha} A_{\alpha} \Phi_{\alpha}(\alpha) \right\}^2 dt$$

yields the following equation for the coefficient A_{ρ}

$$\overline{G(t) \Phi(\rho) \sum_{\alpha} A_{\alpha} \Phi(\alpha)} = \overline{z(t) G(t) \Phi(\rho)} \quad (4.13)$$

Using the time orthogonality property of the Φ 's we get the following explicit equation for A_{ρ} :

$$A_{\rho} = \frac{\overline{z(t) G(t) \Phi(\rho)}}{\overline{G(t) \Phi(\rho)}} \quad (4.14)$$

The averaged products in the numerator and denominator are easily determined experimentally as shown in fig.3, whilst the filter function of equation (4.12) can be synthesized as shown in fig.4.

For a zero memory filter, equation (4.14) becomes:

$$a_k = \frac{\overline{z(t) G(t) \phi_k [x(t)]}}{\overline{G(t) \phi_k [x(t)]}} \quad (4.15)$$

The filter function is given by equation (4.11).

4.3.3 Conclusions.

The chief merit of Bose's method lies in the simplicity of the optimization procedure, brought about by the choice of the non-linear functions $\phi_j(u)$ which yields an explicit

equation for the coefficients involving information which can be easily measured experimentally. However, a price has to be paid in terms of complexity of equipment, since whatever mathematical merits the $\bar{\Phi}(\rho)$ functions have, they do not possess the merit of being easy to synthesize. Also since the $\phi_j(u)$ functions are discontinuous, it is considered that a larger number may often be needed for the same accuracy than if a continuous function is chosen. Whilst Singleton's method leads to difficulties of experimental measurement of sufficient information for the optimization, Bose's method leads to complexities in realizing the equipment which is needed to make these measurements and to synthesize the final filter.

4.4 Zadeh's classification.

The classification of Zadeh, like that of Singleton, is based upon the memory properties of the filter. A filter of class η_1 is defined by the following relationship between filter output and input:

$$y(t) = \int_0^{\infty} F[x(t-\tau), \tau] d\tau \quad (4.16)$$

Class η_2 is defined by:

$$y(t) = \int_0^{\infty} \int_0^{\infty} F[x(t-\tau_1), x(t-\tau_2); \tau_1, \tau_2] d\tau_1 d\tau_2 \quad (4.17)$$

Likewise class γ_n is defined by:

$$y(t) = \underbrace{\int_0^\infty \dots \int_0^\infty}_n F[x(t-\tau_1) \dots x(t-\tau_n); \tau_1, \tau_2 \dots \tau_n] d\tau_1 \dots d\tau_n \quad (4.18)$$

It can be shown that the optimum filter of class γ_n is given by the solution of the integral equation:

$$\begin{aligned} & \underbrace{\int_0^\infty \dots \int_0^\infty}_n \underbrace{\int \dots \int}_n F[x_1 \dots x_n; \tau_1 \dots \tau_n] p_{2n}[x_1 \dots x_n; x'_1 \dots x'_n; \tau_1 \dots \tau_n; \tau'_1 \dots \tau'_n] \\ & \quad dx_1 \dots dx_n d\tau_1 \dots d\tau_n \\ & = \int z p_{n+1}(z; x'_1 \dots x'_n; \tau'_1 \dots \tau'_n) dz \end{aligned} \quad (4.19)$$

As yet, this equation is too difficult to be solved except in the case of class γ_1 , which has been adopted as the object of investigations in this thesis.

The following subclasses of class γ_1 are filters which can be realized and synthesized by the methods described in this thesis.

4.4.1 The class of attenuators.

An attenuator is a device which gives an output at some instant equal to the sum of a term which is proportional to the a.c. component of its input at the same instant and a dc term. (By this definition we allow that the constant of proportionality may be greater than unity).

$$y(t) = k [x(t) - \overline{x(t)}] + K \quad (4.20)$$

where $x(t)$ is the input

$\overline{x(t)}$ is the mean value of input

$y(t)$ is the output

k and K are constants

$$\therefore \overline{y(t)} = K$$

4.4.2 The linear class.

The output of a linear filter, in general, can be expressed as a continuous linear summation of all past values of the a.c. component of its input, in the form of the convolution integral, plus a term which is constant. The class of attenuators is a subclass of the linear class.

$$y(t) = \int_0^{\infty} [x(t-\tau) - \overline{x(t)}] w(\tau) d\tau + K \quad (4.21)$$

where $w(\tau) + K$ is the delta function response

and $K = \overline{y(t)}$

4.4.3 The zero memory class.

The output of a zero memory filter is expressed by an instantaneous non-linear function of its a.c. component of input. The class of attenuators is a subclass of the zero memory class.

$$y(t) = f[x(t)] = \psi[x(t)] + K \quad (4.22)$$

where $\psi[x]$ is a non-linear function satisfying the condition:

$$\overline{\psi[x(t)]} = 0$$

also $f[x]$ and $\psi[x]$ are arbitrary non-linear functions.

$$\text{Again } K = \overline{y(t)}$$

4.4.4 Multipath zero memory filter.

Fig. 5 shows a multipath zero memory filter of degree R , each path of which consists of a polynomial function generator $\theta_v[x]$ defined below, followed by an attenuator in cascade. It is characterized by the following equation:

$$y(t) = \sum_{v=0}^R k_{fv} \theta_v[x(t)] \quad (4.23)$$

where k_{fv} are constants and $\theta_v[x]$ are a set of orthonormal polynomials with respect to $p(x)$ the first probability density function of $x(t)$ as weighting function over the whole range of x .

$$\text{i.e. } \int \theta_n(x) \theta_m(x) p(x) dx = \begin{cases} 1 & n=m \\ 0 & n \neq m \end{cases} \quad (3.1'')$$

Since $\overline{\theta_v[x(t)]} = 0$, $v \neq 0$ and $\theta_0[x] = 1$ (see chapter 3),

we have that $k_{f_0} = \overline{y(t)}$

A multipath zero memory filter can be made to approximate to any zero memory filter with any desired accuracy (by making R large enough), provided the expansion (equation 4.23) converges.

4.4.5 Multipath memory class

Fig.6 shows a multipath memory filter of degree R , each path of which consists of a polynomial function generator $\theta_v[x]$ defined above, followed by a linear memory filter in cascade. It is characterized by the equation:

$$y(t) = \sum_{v=1}^R \int_0^{\infty} \theta_v[x(t-\tau)] w_v(\tau) d\tau \quad (4.24)$$

where $w_v(\tau)$ are delta function responses and $\overline{y(t)} = \int_0^{\infty} w_0(\tau) d\tau$

A multipath memory filter can be made to approximate to a filter of class η , with any desired degree of accuracy (by making R large enough) provided the Kernel function of equation (4.16) has a convergent expansion of the form indicated below:

$$F[x(t-\tau), \tau] = \sum_{v=1}^{\infty} \theta_v[x(t-\tau)] w_v(\tau) \quad (4.25)$$

5.0 THE OPTIMIZATION OF SOME FILTERS OF CLASS η_1 .

In this chapter some filters belonging to the various subclasses of class γ_1 are optimized. The optimum filter is obtained by minimizing the mean-squared error defined by equation 1.1.

$$\overline{e^2} = \overline{[y-z]^2} = \overline{y^2} + \overline{z^2} - 2\overline{yz} \quad (1.1)$$

where the bar indicates an ensemble or time average; in this and subsequent chapters, all signals are assumed to be ergodic, whence ensemble and time averages are identical.

Let $y(t) = f[x(t)]$ where, for the purposes of this chapter only, f is a general symbol which represents a filter operation of a specified class (not necessarily zero memory), on the whole, or part of the past and present input and having finite mean square.

The variation in $\overline{e^2}$ produced by a variation in $f[x]$ or $\epsilon h[x]$, is given by:

$$\Delta \overline{e^2} = 2\epsilon \overline{h[x] \{f[x] - z\}} + \epsilon^2 \overline{h[x]^2} \quad (5.1)$$

where ϵ is a small constant and $h[x]$ is an operation of the same class as $f[x]$.

The necessary condition for $f[x]$ to be optimum is given by:

$$\frac{\partial \Delta \overline{e^2}}{\partial \epsilon} \bigg|_{\epsilon=0} = -2 \overline{h[x_1] \{f[x_2] - z\}} = 0$$

$$\therefore \overline{h[x_1] \{f[x_2] - z\}} = 0 \quad (5.2)$$

Sufficiency of this condition is easily proved by showing that $\frac{\partial \Delta \bar{e}^2}{\partial \epsilon}$ is always positive, irrespective of

From equation 5.2 the following two conditions are derived:

- i. For a chosen variation $\epsilon_h[x] = \epsilon f[x]$

$$\overline{f^2(x)} = \overline{z f(x)} \quad (5.3)$$

i.e. $\overline{y^2(t)} = \overline{z(t) y(t)}$

- ii. For a chosen variation $\epsilon_h[x] = \epsilon x$

$$\overline{f[x]} = \overline{z} \quad (5.4)$$

i.e.

$$\overline{y(t)} = \overline{z(t)}$$

The above two conditions 5.3 and 5.4 are necessary for $f[x]$ to represent an optimum filter of any class.

Substitution of equations 5.3 and 5.4 in equation 1.1 yields:

$$\overline{e^2} = \sigma_z^2 - \sigma_y^2 \quad (5.5)$$

where σ_z is the variance of the desired output
and σ_y is the variance of the optimum system output.

This equation for mean-squared error is true for an optimum filter belonging to any class.

Furthermore, suppose there should exist more than one optimum solution for $f[x]$; by taking any two solutions $f_1[x]$ and $f_2[x]$ we may write using equation 5.2

$$\overline{f_1[x] \{f_2[x] - z\}} = 0$$

Similarly

$$\overline{f_2[x] \{f_1[x] - z\}} = 0$$

whence

$$\overline{f_1[x] z} = \overline{f_2[x] z} \quad (5.6)$$

i.e.

$$\overline{y_1(t) z(t)} = \overline{y_2(t) z(t)}$$

also from equation 5.4

$$\overline{y_1(t)} = \overline{y_2(t)} = \overline{z(t)} \quad (5.7)$$

From equations 1.1, 5.3, ... and 5.6 we see that two solutions of the same class yield the same mean squared error; thus although any one solution may not be unique, it is no worse than any other solution.

5.1 Filters of class η_1

The input-output relationship for filters of class η_1

is of the form:

$$y(t) = \int_0^{\infty} F[x(t-\tau), \tau] d\tau \quad (4.16)$$

To obtain the optimum filter of this class we must solve equations^{5.2} which can be written in the form:

$$\int h[x_1] \left[\int f[x_2] p_2(x_1, x_2; \tau) dx_2 - \int z p_2(x_1, z; \tau) dz \right] dx_1 = 0 \quad (5.8)$$

where

$$f[x] = \int_0^{\infty} F[x, \tau] d\tau$$

and

$$h[x] = \int_0^{\infty} H[x, \tau] d\tau$$

By substitution we obtain:

$$\int \int_0^{\infty} H[x_1, \tau_1] \left[\int \int_0^{\infty} F[x_2, \tau_2] p_2(x_1, x_2; \tau_1 - \tau_2) dx_2 d\tau_2 - \int z p_2(x_1, z; \tau_1) dz \right] dx_1 d\tau_1 = 0 \quad (5.9)$$

Using the fundamental lemma of variational calculus, we obtain from equation 5.9 an integral equation which gives

the implicit solution for the Kernel function of the optimum non-linear filter.

$$\begin{aligned} \int_0^\infty \int F[x_2, \tau_2] p_2(x_1, x_2; \tau_1 - \tau_2) dx_2 d\tau_2 \\ = \int z p_2(x, z; \tau_1) dz \end{aligned} \quad (5.10)$$

$\tau_1 \geq 0$

Thus the information needed in order to optimize a filter of class γ_1 is the second probability density of the input $p_2(x_1, x_2; \tau)$ and the joint second probability density $p_2(x, z; \tau)$ between the input and desired output.

5.2 Zero memory non-linear filters.

The input-output relationship for a zero memory filter is of the form:

$$y(t) = f[x(t)] \quad (5.11)$$

To obtain the optimum filter function of this class we must solve equation 5.8 as it stands where $f[x]$ and $h[x]$ are instantaneous (zero memory) functions of x

Using the fundamental lemma of variational calculus equation 5.8 becomes:

$$\int f(x_2) p_2(x_1, x_2; 0) dx_2 - \int z p_2(x, z; 0) dz = 0 \quad (5.12)$$

But

$$p_2(x_1, x_2; 0) = p(x_1) \delta(x_2 - x_1) \quad (5.13)$$

So that the zero memory filter function is given by the explicit equation:

$$f(x) = \frac{\int z p_2(x, z) dz}{p(x)} \quad (5.14)$$

where $p_2(x, z) \delta x \delta z$ is the probability that x lies in the range x to $x + \delta x$ and z lies in the range z to $z + \delta z$ simultaneously, as both δx and δz tend to zero.

Thus the optimum filter function $f(x)$ is equal to the conditional mean value of z , given x and the information needed in the optimization of a zero memory filter is the whole information content of the first probability density function of the input and the second joint density function of the input and desired output with $\tau = 0$.

5.3 Multipath zero memory filters.

We have defined a class of non-linear filters in chapter 4, known as multipath zero memory filters of degree R by the following relationship between filter input and output.

$$y(t) = f[x] = \sum_{m=0}^R k_{fm} \theta_m[x] \quad (5.15)$$

where k_{fm} are constants and the set of functions $\theta_m[x]$ are orthonormal polynomials with respect to the first probability density of the input $p(x)$ over the whole range of x . (See chapter 3).

The most general form for $h[x]$ is:

$$h[x] = \sum_{n=0}^R k_{hn} \theta_n(x)$$

By substitution in equation 5.2 we obtain:

$$\sum_{n=0}^R k_{hn} \left\{ \sum_{m=0}^R k_{fm} \overline{\theta_n[x] \theta_m[x]} - \overline{\theta_n[x] z} \right\} = 0$$

Since $\int \theta_n[x] \theta_m[x] p(x) dx = \overline{\theta_n[x] \theta_m[x]} = \begin{cases} 0 & m \neq n \\ 1 & m = n \end{cases}$

(See chapter 3)

we obtain $\sum_{n=0}^R k_{hn} k_{fn} = \sum_{n=0}^R k_{hn} \overline{\theta_n[x] z}$ (5.16)

Since the coefficients k_{hn} can be given any arbitrary set of values, for equation 5.16 to be true, only one solution for the constants k_{fn} can exist.:

$$k_{fn} = \overline{\theta_n[x] z} \quad (5.17)$$

also since $\theta_0[x] = 1$ (See chapter 3)

$$k_{f0} = \overline{f[x]} = \overline{y(t)} = \overline{z(t)} \quad (5.18)$$

Thus the optimum multipath zero memory filter of degree R is given by:

$$f[x(t)] = \sum_{n=1}^R \sigma_z D_n \theta_n[x(t)] + \overline{z(t)} \quad (5.19)$$

$$\text{where } D_n = \frac{\overline{\theta_n[x] (z - \overline{z})}}{\sigma_z} \quad \left[\begin{array}{l} \text{from} \\ \text{equations} \\ 3.4 \text{ and} \\ 3.14 \end{array} \right]$$

$$= \frac{\overline{\theta_n[z] z}}{\sigma_z} \quad n \neq 0$$

From equation 5.19 we obtain

$$\sigma_y^2 = \overline{\{f[x(t)] - \overline{z(t)}\}^2} = \sum_{n=1}^R \sigma_z^2 D_n^2 \quad (5.20)$$

$$\left\{ \begin{array}{l} \text{Since } \overline{\theta_n(x)} = 0 \quad n \neq 0 \\ \text{and} \\ \theta_n[x] \theta_m[x] = 1 \quad n = m \\ \quad \quad \quad = 0 \quad n \neq m \end{array} \right\}$$

Substitution in equation 5.5 gives:

$$\overline{e^2} = \sigma_z^2 \left(1 - \sum_{n=1}^R D_n^2 \right) \quad (5.21)$$

The coefficients D_n and the constants σ_z and $\overline{z}$ can be determined by analogue measurements if long enough samples of the filter input and desired output are available or they may be determined analytically from a knowledge of the statistics of the input signal and desired output. The following recurrence equation which gives $\sigma_z D_n$ in terms

of the first moments $\overline{x^*}$ and the averaged products between z and x^* is obtained from equation (3.6)

$$\delta_z D_{n1} = \overline{\theta_n(x) z} = \frac{\overline{x^n z} - \sum_{r=0}^{n-1} \overline{x^n \theta_r(x)} D_{r1} \delta_z}{\left[\overline{x^{2n}} - \sum_{r=0}^{n-1} \overline{x^n \theta_r[x]^2} \right]^{1/2}} \quad (5.22)$$

5.4 Multipath memory filters.

We have defined a class of multipath memory filters of degree R in chapter 4, by the following relationship between input and output

$$y(t) = f[x] = \sum_{m=0}^R \int_0^{\infty} w_{f_m}(\tau) \theta_m[x(t-\tau)] d\tau \quad (5.23)$$

The most general form for $h[x]$ is:

$$h[x] = \sum_{n=0}^R \int_0^{\infty} w_{h_n}(\tau) \theta_n[x(t-\tau)] d\tau$$

Substitution in equation 5.2 gives:

$$\sum_{n=0}^R \int_0^{\infty} w_{h_n}(\tau_1) \left[\sum_{m=0}^R \int_0^{\infty} w_{f_m}(\tau_2) C_{nm}(\tau_1, \tau_2) d\tau_2 - \overline{\theta_n[x(t-\tau_1) z(t)]} \right] d\tau_1 = 0 \quad (5.24)$$

where $C_{nm}(\tau) = \overline{\theta_n[x(t)] \theta_m[x(t+\tau)]}$

Using the fundamental lemma of variational calculus and noting that all $w_{h_n}(\tau)$ are arbitrary but physically

realizable weighting functions of linear networks (i.e.

$w_{hn}(\tau) = 0$ for $\tau < 0$) we obtain:

$$\sum_{m=0}^R \int_0^{\infty} w_m(\tau_2) C_{nm}(\tau_1 - \tau_2) d\tau_2 = \overline{\theta_n [x(t-\tau_1) z(t)]} \quad (5.25)$$

$\tau_1 \geq 0$

since

$$n = 0, 1, 2, \dots, R$$

$$C_{0m}(\tau) \equiv 0 \quad m \neq 0$$

$$= 1 \quad m = 0$$

and

$$\theta_0 [x(t)] = 1$$

we obtain

$$\int_0^{\infty} w_0(\tau) d\tau = w_0 = \overline{z(t)} \quad (5.26)$$

and

$$\sum_{m=1}^R \int_0^{\infty} w_m(\tau_2) C_{nm}(\tau_1 - \tau_2) d\tau_2 = \sigma_z D_{n1}(\tau_1) \quad (5.27)$$

$\tau_1 \geq 0$

$n = 1, 2, 3, \dots, R$

where

$$D_{n1}(\tau) = \frac{\overline{\theta_n [x(t-\tau)] [z(t) - \overline{z(t)}]}}{\sigma_z}$$

$$= \frac{\overline{\theta_n [x(t-\tau)] z(t)}}{\sigma_z} \quad n \neq 0$$

From equations ^{5.23} 5.26 and 5.27 it follows that: $\sigma_y^2 = \overline{y^2} - \overline{y}^2$

$$\begin{aligned}
 &= \sum_{n=1}^R \sum_{m=1}^R \int_0^\infty \int_0^\infty w_n(\tau_1) w_m(\tau_2) C_{nm}(\tau_1 - \tau_2) d\tau_1 d\tau_2 \quad (5.28) \\
 &= \sum_{n=1}^R \int_0^\infty D_{nn}(\tau) \sigma_z w_n(\tau) d\tau
 \end{aligned}$$

By substitution in equation 5.5 we obtain:

$$\overline{e^2} = \sigma_z^2 \left(1 - \frac{\sigma_y^2}{\sigma_z^2} \right) = \sigma_z^2 \left(1 - \sum_{n=1}^R \int_0^\infty \frac{D_{nn}(\tau)}{\sigma_z} w_n(\tau) d\tau \right) \quad (5.29)$$

The simultaneous integral equations 5.27 can be solved by the method of undetermined coefficients as described elsewhere ¹² and in appendix 13.3.

It seems that often there is a unique set of solutions for the optimum linear weighting functions $w_n(\tau)$ although it has already been shown that if more than two sets of solutions exist they yield the same mean-squared error.

The multipath memory filter of degree R can only approximate to a more general filter of class η_1 , because for physical realizability the number of paths R must be finite. For a large value of R more information is required to optimize the filter until in the limit as

$R \rightarrow \infty$ the whole information content of $p_2(x_1, x_2; \tau)$

and a large part of the information in $\rho_2(x, z; \tau)$ is required, i.e. all the coefficients $C_{nm}(\tau)$ in the infinite expansion of $\rho_2(x_1, x_2; \tau)$ and $D_{n1}(\tau)$ in the infinite expansion of $\rho_2(x, z; \tau)$

The coefficients $C_{nm}(\tau)$ and $D_{n1}(\tau)$ can be determined by analogue measurements given long enough samples of the filter input and desired output, or they may be determined analytically from a knowledge of the statistics of the input and desired output as described in chapter 3.

5.5 Linear attenuators.

The optimum attenuator defined by equation (4.20) can be found by minimizing the mean square error using the differential calculus or from equation (5.19), since the class of attenuators is a subclass of the class of zero memory filters.

Therefore if

$$y(t) = k [x(t) - \overline{x(t)}] + K \quad (4.20)$$

and (5.19) putting $R = 1$,
using equations (3.4) the optimum condition is given by:

$$k = \frac{\overline{(x(t) - \overline{x(t)})(z(t) - \overline{z(t)})}}{\sigma_x^2} = \frac{\phi_{xz}(0)}{\sigma_x^2} \quad (5.30)$$

and

$$K = \overline{z(t)}$$

whilst the mean-square error of the optimum attenuator is found by putting $R = 1$ in

using equation (5.21)

$$\overline{e_a^2} = \sigma_z^2 (1 - D_n^2) = \sigma_z^2 - \frac{\sigma_z}{\sigma_x} \phi_{xz}(0) \quad (5.31)$$

5.6 Linear memory filters.

The optimum linear filter defined by equation (4.21) can be found by using equation (5.2) or from equation (5.27), since the class of linear filters is a subclass of the class of multi-path memory filters.

For the linear class:

$$y(t) = \int_0^\infty [x(t-\tau) - \bar{x}] w(\tau) d\tau + K \quad (4.21)$$

Using equation (5.27) it can be shown that the optimum $w(\tau)$ is given by the solution of the integral equation below*:

$$\int_0^\infty w(\tau_2) \phi_{xx}(\tau_1 - \tau_2) d\tau_2 = \phi_{xz}(\tau_1) \quad (5.32)$$

[and $K = \overline{z(t)}$]

$\tau_1 \geq 0$

This equation is known as the Wiener Hopf equation and its solution is given in appendix 13.3.

The mean square error of the optimum filter is found using equation (5.29)

$$\overline{e^2} = \sigma_z^2 - \int_0^\infty \phi_{xz}(\tau) w(\tau) d\tau \quad (5.33)$$

where $\phi_{xz}(\tau) = \sigma_z \sigma_x D_n(\tau)$

* Owing to a slight confusion of notation $w(\tau)$ of equation (4.21) corresponds to $\frac{w_1(\tau)}{\sigma_x}$ of equations (5.23) and (5.27).

5.7 Applications.

A number of applications in the control and communications field are immediately apparent; the following is a list of some applications which are considered in this thesis:

1. Noise filtering
2. Prediction
3. System analysis
4. System compensation.

Optimization procedures for some of the filters of class η_1 are described for each of these applications and some examples are worked out. In all the examples of multipath filters, no more than five paths are used. Where possible, analytical examples are chosen. Experimental results are used only in cases which appeared to be intractable analytically.

Under certain circumstances, it may be possible to find a simpler and better non-linear filter belonging to some class other than η_1 , in which event it would obviously not be desirable to use the optimization procedures of this thesis, but although the proposed methods do not always yield the simplest solutions, they can often be used to produce a physically realizable filter, which is better than a linear filter under conditions when other methods fail due to complexity or intractability of the mathematics.

6.0 SOME SPECIAL CLASSES OF SECOND PROBABILITY DENSITY FUNCTIONS.

Classes of second probability density functions can be defined in terms of the expansion-coefficient matrix-pattern and in other ways. The classes defined below are of interest because they occur in the non-linear filter examples which are worked out in chapters 7,8,9 and 10.

6.1 Class Λ .

There exists a class of second probability density functions¹³ Λ of physically realizable processes for which it is possible to expand the density function in the form of a single series expansion consisting of diagonal terms of the coefficient matrix only. This is a special case of the expansion 3.8.

viz.

$$\rho_2(x, z; t_1, t_2) = \sum_{n=0}^{\infty} D_{nn}(t_1, t_2) \theta_n[x, t_1] \gamma_n[z, t_2] \rho(x, t_1) \times q(z, t_2) \quad (6.1)$$

6.1.1 The Gaussian Process.

An example of class Λ is the Gaussian second probability density with zero mean value $\bar{x} = 0$, for which we may write:

$$\rho_2(x_1, x_2; \tau) = \frac{1}{2\pi\sqrt{1-\rho^2(\tau)}} \exp \left[- \frac{x_1^2 + x_2^2 - 2\rho(\tau)x_1x_2}{2\{1-\rho^2(\tau)\}} \right] \quad (6.2)$$

where $\tau = t_2 - t_1$, since the process is stationary and independent of time origin.

It is shown by Barrett and Coales¹⁴ for this case that the series expansion becomes:

$$p_2(x_1, x_2; \tau) = \frac{e^{-\frac{(x_1^2 + x_2^2)}{2}}}{2\pi} \sum_{n=0}^{\infty} H_n(x_1) H_n(x_2) \frac{\rho^n(\tau)}{n!} \quad (6.3)$$

where $H_n(x)$ is the Hermite polynomial type 2, using the notation of Bateman¹⁵ and is defined by:

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} \quad (6.4)$$

so that

$$\begin{aligned} \int H_n(x) H_m(x) e^{-x^2/2} dx &= n! \sqrt{2\pi} \quad n=m \\ &= 0 \quad n \neq m \end{aligned} \quad (6.5)$$

Therefore

$$\begin{aligned} \phi_n(x) &= \frac{H_n(x)}{\sqrt{n!}} \\ \rho(x) &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \end{aligned} \quad (6.6)$$

and

$$D_{nn}(\tau) = \rho^n(\tau)$$

6.1.2 An asymmetrical probability density of class Λ .

Another example of a second probability density of class Λ is used as the desired output for problem 2 in

chapter 7, and is obtained by passing Gaussian noise through a square law device. For convenience the square law device is chosen so that its output has zero mean and unit variance, viz.

$$z(t) = \frac{x^2 - 1}{\sqrt{2}} = \frac{He_2[x(t)]}{\sqrt{2}} \quad (6.7)$$

where $z(t)$ is its output
and $x(t)$ its input.

It is shown below that the second probability density of the output $p_2(z_1, z_2; \tau)$ is of class Λ . It is always possible to find a polynomial θ_m of order m such that

$$y = \theta_m \left[\frac{He_2(x)}{\sqrt{2}} \right] = \frac{He_{2m}(x)}{\sqrt{(2m)!}} \quad (6.8)$$

Although no proof of this statement is given, it can be easily verified for all finite-values of m .

Let θ_n and θ_m be two such polynomials, we write

$$\phi_{nm}(\tau) = \iint \theta_n(z_1) \theta_m(z_2) p_2(z_1, z_2; \tau) dz_1 dz_2, \quad (6.9)$$

where ϕ_{nm} is the cross-correlation between the outputs y_1 and y_2 of the polynomial function generators θ_n and θ_m , then

$$\begin{aligned}
\phi_{nm}(\tau) &= \iint \theta_n \left[\frac{He_2(x_1)}{\sqrt{2}} \right] \theta_m \left[\frac{He_2(x_2)}{\sqrt{2}} \right] p_2(x_1, x_2; \tau) dx_1 dx_2 \\
&= \iint \frac{He_{2m}(x_1)}{\sqrt{(2m)!}} \frac{He_{2n}(x_2)}{\sqrt{(2n)!}} p_2(x_1, x_2; \tau) dx_1 dx_2 \quad (6.10) \\
&= \frac{1}{2\pi} \sum_{r=0}^{\infty} \frac{\rho^r}{r!} \int \frac{He_{2m}(x_1)}{\sqrt{(2m)!}} He_r(x_1) e^{-x_1^2/2} dx_1 \int \frac{He_{2n}(x_2)}{\sqrt{(2n)!}} He_r(x_2) e^{-x_2^2/2} dx_2
\end{aligned}$$

(by substituting equation 6.3)

whence

$$\begin{aligned}
\phi_{nm}(\tau) &= 0 \quad n \neq m \\
&= \rho_{x, x_1}^2(\tau) \quad n = m \quad (6.11)
\end{aligned}$$

(by using equation 6.5).

But ϕ_{nm} can be identified as the expansion coefficients, therefore $\rho_2(z_1, z_2; \tau)$ is of class Λ .

We note also that:

$$\phi_{11}(\tau) = \rho_{z, z_1}(\tau) = \rho_{x, x_1}^2(\tau)$$

Thus if

$$\rho_{x, x_1}(\tau) = e^{-\rho|\tau|} \quad \text{then} \quad \rho_{z, z_1}(\tau) = e^{-2\rho|\tau|} \quad (6.12)$$

This result is used in chapter 7.

The first probability density of Z can be obtained as follows:-

$$\rho(x) = \frac{1}{2\pi} e^{-x^2/2}$$

for zero mean and unit variance

$$\text{But } x = \sqrt{\sqrt{2}z + 1}$$

$$\text{and } 2x dx = 2 dz$$

$$\begin{aligned} \therefore p(z) &= \frac{1}{2\sqrt{\pi}} e^{-\frac{\sqrt{z+1}}{2}} \times \frac{1}{\sqrt{\sqrt{2}z+1}} \quad z > -\frac{1}{\sqrt{2}} \\ &= 0 \quad z \leq -\frac{1}{\sqrt{2}} \end{aligned} \quad (6.13)$$

The moments of z can be found from the equation

$$\overline{z^n} = \int z^n p(z) dz \quad (6.14)$$

or from the expression

$$\overline{z^n} = \overline{\left(\frac{x^2-1}{\sqrt{2}}\right)^n} \quad (6.15)$$

6.2 Separable random processes and class Δ

16

Nuttall¹⁶ investigated a class of random processes which he called separable processes.

Writing

$$g(x; t_1, t_2) = \int [s - \langle s(t_2) \rangle] p_2(x, s; t_1, t_2) ds \quad (6.16)$$

Then the random process X is said to be separable with respect to the random process^{*} S if $g(x; t_1, t_2)$ can be written in the form:

$$g(x; t_1, t_2) = g_1(x, t_1) g_2(t_1, t_2) \quad (6.17)$$

6.2.1 Some properties of separable processes.

Consider fig. 7, a signal $x(t)$ is connected as the input to a zero memory nonlinearity, the output of which is given by:

$$z = f[x(t)]$$

The output cross-correlation is given by:

$$\begin{aligned} \phi_{zs}(t_1, t_2) &= \iint f[x] [s - \langle s(t_2) \rangle] p_2(x, s; t_1, t_2) dx ds \\ &= \int f[x] g(x; t_1, t_2) dx \end{aligned}$$

If X is separable with respect to S then:

$$\phi_{zs}(t_1, t_2) = g_2(t_1, t_2) \int f[x] g_1(x, t_1) dx$$

---

* Nuttall uses the opposite convention. In this thesis we shall refer to x as the first variable and S as the second variable.

Similarly

$$\phi_{xs}(t_1, t_2) = g_2(t_1, t_2) \int x g_1(x, t_1) dx$$

$$\therefore \phi_{zs}(t_1, t_2) = a(t_1) \phi_{xs}(t_1, t_2) \quad \left. \vphantom{\phi_{zs}(t_1, t_2)} \right\} \quad (6.18)$$

where

$$a(t_1) = \frac{\int f[x] g_1(x, t_1) dx}{\int x g_1(x, t_1) dx}$$

Thus separability of X with respect to S is a sufficient condition for the cross-correlation between X and S to be invariant when X undergoes non-linear transformation, provided $\phi_{zs}(t_1, t_2)$ is neither infinite nor zero.

This property of the correlation coefficients is known as the invariance property and we say that X is invariant with respect to S . Invariance or separability of X with respect to S does not imply invariance or separability of S with respect to X .

Nuttall has shown that separability is sufficient as

well as necessary for the invariance property to hold provided the output cross-correlation exists.

If a single process is separable with respect to itself, then writing:

$$g(x_1; t_1, t_2) = \int [x_2 - \langle x(t_2) \rangle] p_2(x_1, x_2; t_1, t_2) dx_2 \quad (6.19)$$

and

$$g(x_1; t_1, t_2) = g_1(x_1; t_1) g_2(t_1, t_2) \quad (6.20)$$

also

$$\phi_{xx}(t_1, t_2) = g_2(t_1, t_2) \int x g_1(x, t_1) dx \quad (6.21)$$

when $t_1 = t_2$ $p_2(x_1, x_2; t_1, t_2) = p(x_1; t_1) \delta(x_2 - x_1)$

and

$$g(x_1; t_1) = g_1(x_1, t_1) g_2(t_1, t_1)$$

For convenience we let $g_2(t_1, t_1) = 1$

then $g(x_1; t_1) = g_1(x_1; t_1) = \int [x_2 - \langle x(t_1) \rangle] p(x_1; t_1) \delta(x_2 - x_1) dx_2$

$$= [x_1 - \langle x(t_1) \rangle] p(x_1; t_1) \quad (6.22)$$

Substituting in equation (6.21) gives:

$$g_2(t_1, t_2) = \frac{\phi_{xx}(t_1, t_2)}{\sigma_x^2(t_1)} \quad (6.23)$$

Substituting equations (6.23) and (6.22) in equation (6.20) gives:

$$g(x_i; t_1, t_2) = [x_i - \langle x(t_1) \rangle] p(x_i; t_1) \frac{\phi_{xx}(t_1, t_2)}{\sigma_x^2(t_1)} \quad (6.24)$$

Also substituting equation (6.22) in equations (6.18) gives:

$$\phi_{zx}(t_1, t_2) = a(t_1) \phi_{xx}(t_1, t_2)$$

$$\begin{aligned} \text{where } a(t_1) &= \frac{\int f(x) [x - \langle x(t_1) \rangle] p(x; t_1) dx}{\int x [x - \langle x(t_1) \rangle] p(x; t_1) dx} \quad (6.25) \\ &= \frac{\phi_{zx}(t_1, t_1)}{\phi_{xx}(t_1, t_1)} \end{aligned}$$

6.2.2 A property of the expansion coefficients.

Some interesting results concerning the expansion coefficients of equation (3.8) are obtained by substitution of equation (3.8) in equation (6.16)

$$g(x; t_1, t_2) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} D_{nm}(t_1, t_2) \theta_n(x, t_1) p(x, t_1) \\ \times \int [s - \langle s(t_2) \rangle] \gamma_m(s, t_2) q(s, t_2) ds$$

By using the orthogonal property of $\gamma_m(s, t)$ and that:

$$\gamma_1(s, t_2) = \frac{s - \langle s(t_2) \rangle}{\sigma_s(t_2)}$$

we obtain:

$$g(x; t_1, t_2) = \sum_{n=0}^{\infty} D_{n1}(t_1, t_2) \theta_n(x, t_1) p(x, t_1) \sigma_s(t_2) \quad (6.26)$$

Since $D_{n0} = 0$, if the expansion 3.8 exists and $(n \neq 0)$

$$D_{n1}(t_1, t_2) = b_n(t_1) D_{n1}(t_1, t_2) = b_n(t_1) \rho_{xs}(t_1, t_2) \quad (6.27)$$

then the process is separable; conversely if expansion 3.8 exists and the process is separable then condition 6.27 follows. This condition indicates that the coefficients in the first column of the expansion coefficient matrix are all proportional to the normalized cross correlation coefficient.

The class of processes for which an expansion of the form of equation (3.8) is possible and which satisfies condition 6.27 was first introduced by Brown¹⁷, who showed that condition 6.27 is both necessary and sufficient for the invariance property. Class Λ of Barrett and Lampard¹³

is a subclass of Brown's class and therefore also exhibits the invariance property, which Busgang¹⁸ originally demonstrated as a property of Gaussian processes.

For a single separable process, by equating the right hand side of equations (6.24) and (6.26) we obtain:

$$\begin{aligned} [x_1 - \langle x(t_1) \rangle] p(x_1; t_1) p_{xx}(t_1, t_2) \\ = \sum_{n=0}^{\infty} C_n(t_1, t_2) \theta_n(x_1, t_1) p(x_1; t_1) \sigma_x(t_2) \end{aligned}$$

(using the symbol C for coefficients ^{the} in expansion of a second probability density of a single process and symbol D for joint probability densities). and using equation 3.4

Thence equating coefficients of θ_n we obtain

$$C_{11}(t_1, t_2) = p_{xx}(t_1, t_2) \quad \left. \begin{array}{l} \text{which is a familiar} \\ \text{result} \end{array} \right\} \quad (6.28)$$

$$\text{and } C_n(t_1, t_2) = 0 \quad n \geq 2$$

This result is true of single separable processes, conversely it can be shown if (6.28) holds and equation (3.8) exists for $p_2(x_1, x_2; t_1, t_2)$ then the process is separable.

Separable processes are used as inputs to both the noise filters which are optimized in chapter 7. These inputs are obtained in each case by adding signal and noise, which are both separable and with identical correlation coefficients. It can be shown by a general proof that the sum of separable processes with ^{the} same auto-correlation is separable.

These results are reproduced here because they are needed later in this thesis but for a more complete treatment of the separable class, reference should be made to Nuttall's work¹⁶.

6.2.3 Linear transformation.

A property not mentioned by Nuttall concerns linear transformation of the second variable.

Let $S(t)$ be the input to a linear filter, which gives a response $K(t)$ independent of its input and a response $\psi(\lambda, t) = K(t)$ to a delta function input (fig.8).

$$\therefore z(t) = \int_{-\infty}^t \psi(t, \lambda) [S(\lambda) - \langle S(\lambda) \rangle] d\lambda + K(t) \quad (6.29)$$

where $K(t) = \langle z(t) \rangle$

and $z(t)$ is the linear filter output.

$$\begin{aligned}\phi_{xz}(t_1, t_2) &= \iint x [z - \langle z \rangle] p_2(x, z; t_1, t_2) dx dz \\ &= E \left[x(t_1) \left\{ z(t_2) - \langle z(t_2) \rangle \right\} \right] \quad (6.30)\end{aligned}$$

$$\begin{aligned}\therefore \phi_{xz}(t_1, t_2) &= E \left\{ x(t_1) \int_{-\infty}^{t_2} w(t_2, \lambda) [s(\lambda) - \langle s(\lambda) \rangle] d\lambda \right\} \\ &= \int_{-\infty}^{t_2} w(t_2, \lambda) E \left\{ x(t_1) [s(\lambda) - \langle s(\lambda) \rangle] \right\} d\lambda \quad (6.31)\end{aligned}$$

provided $w(t)$ is continuous

and $E [x(s - \langle s \rangle)]$ exists.

Equating the right hand sides of equation (6.30) and (6.31) and using equation ^{6.16} (6.17) gives:

$$\begin{aligned}\iint x [z - \langle z \rangle] p_2(x, z; t_1, t_2) dx dz \\ = \int_{-\infty}^{t_2} w(t_2, \lambda) g_2(t_1, \lambda) d\lambda \int x g_1(x, t_1) dx \quad (6.32)\end{aligned}$$

Since the right hand side separates, it is concluded that the process X is separable with respect to Z if it is separable with respect to S .

writing

$$h(x, ; t_1, t_2) = \int [z - \langle z \rangle] p_z(x, z; t_1, t_2) dz$$

Then

$$h(x, ; t_1, t_2) = g_1(x, t_1) h_2(t_1, t_2) \quad (6.33)$$

where

$$h_2(t_1, t_2) = \int_{-\infty}^{t_2} \omega(t_2, \lambda) g_2(t_1, \lambda) d\lambda$$

If x and z belong to the same random processes X then substituting equations (6.22) and (6.23) in (6.32) gives:

$$\phi_{xz}(t_1, t_2) = \int_{-\infty}^{t_2} \omega(t_2, \lambda) \phi_{xx}(t_1, \lambda) d\lambda \quad (6.34)$$

which is a well-known result.

It is equally apparent that:

$$\phi_{\theta_n(x)z}(t_1, t_2) = \int_{-\infty}^{t_2} \omega(t_2, \lambda) \phi_{\theta_n(x)x}(t_1, \lambda) d\lambda$$

where $\theta_n(x, t_1)$ is an orthonormal polynomial function defined in section 3.1 and $\phi_{\theta_n(x)z}(t_1, t_2)$ is the cross-correlation coefficient between the output

of the polynomial function generator $\theta_n(x)$ and the output of the linear filter $z(t)$.

$$\text{Since } D_{n1}^{(xz)}(t_1, t_2) = \frac{\theta_n(x) [z(t) - \langle z \rangle]}{\sigma_z}$$

$$\text{and likewise } = \sigma_z \phi_{\theta_n(x)z}(t_1, t_2)$$

$$C_{n1}^{(xx)}(t_1, t_2) = \frac{\theta_n(x) [x(t) - \langle x \rangle]}{\sigma_x}$$

$$= \sigma_x \phi_{\theta_n(x)x}(t_1, t_2)$$

$$\text{Then } D_{n1}^{(xz)}(t_1, t_2) = \frac{\sigma_x}{\sigma_z} \int_{-\infty}^{t_2} w(t_2, \lambda) C_{n1}^{(xx)}(t_1, \lambda) d\lambda \quad (6.35)$$

But if the random process X is separable, then

$$C_{n1}^{(xx)}(t_1, t_2) \equiv 0 \quad \text{and} \quad D_{n1}^{(xz)}(t_1, t_2) \equiv 0 \quad n \geq 2$$

(by using equation 6.28 for $n \geq 2$ in equation 6.35)

Thus the separability of x with respect to z is of a special kind, since the expansion coefficients $D_{n1}^{(xz)}$ are not only proportional to each other, but they are all zero for $n \geq 2$ (c.f. equation 6.28). This leads to the following definition of a class of second probability density functions based on the pattern of expansion coefficients.

6.2.4 Class Δ .

We can define a class Δ of second probability density functions $p_2(x, z; t_1, t_2)$ for which an expansion of the form of equation (3.8) exists and the expansion coefficients have the following pattern:

$$\left. \begin{aligned} D_{11}(t_1, t_2) &= p_{xz}(t_1, t_2) \\ D_{n1}(t_1, t_2) &= 0 \quad n \geq 2 \end{aligned} \right\} \quad (6.36)$$

This class, which is of interest in non-linear filtering (section 6.4), is less general than the separable class but x is separable and invariant with respect to z . Class Δ of Barrett and Lampard is a subclass of class Δ .

If x and z are members of the same ensemble, then the equations (6.36) are identical to equation (6.28), therefore separable single processes for which the expansion 3.8 exists, are members of class Δ . One example of a joint probability density of this class has already been given, a second example is as follows:

Let $z(t_2) = x_2(t_1 + \tau)$ and $\rho_2(x_1, x_2; t_1, t_2)$ be the second probability density of a single separable process ^{which} for expansion 3.8 exists, then $\rho_2(x_1, z; t_1, t_2)$ is of class Δ ; this case is of interest in non-linear prediction (section 6.4).

6.3 Semi-separable random processes.

Following the arguments of the previous sub-section, a class of semi-separable random processes can be defined:

Let

$$g(x; t_1, t_2) = \int [s - \langle s(t_2) \rangle] \rho_2(x, s; t_1, t_2) ds \quad (6.16)$$

The random process X is said to be semi-separable with respect to the random process S if $g(x; t_1, t_2)$ can be written in the form:

$$g(x; t_1, t_2) = g_1(x, t_1) g_2(t_1, t_2) \quad (6.37)$$

for $t_1 \geq t_2$ only.

It is apparent that semi-separability can be defined either for $t_1 \geq t_2$ or $t_2 \geq t_1$ and, in fact, both cases can exist. Where necessary, we shall use the terms "semi-separable negative" and "semi-separable positive", to distinguish these two cases.

6.3.1 Some properties of semi-separable processes.³⁷

All the results proved previously for separable processes are true of semi-separable processes over the restricted range $t_1 \geq t_2$ or $t_2 \geq t_1$ as the case may be; thus under non-linear transformation (fig.7) if X is semi-separable negative with respect to S :

$$\phi_{zs}(t_1, t_2) = a(t_1) \phi_{xs}(t_1, t_2)$$

$$\text{for } t_1 \geq t_2 \text{ only}$$

$$\text{where } z(t) = f[x(t)]$$

$$\text{and } a(t_1) = \frac{\int f[x] g_1(x, t_1) dx}{\int x g_1(x, t_1) dx}$$

(6.38)

For a single process

$$g(x_1; t_1, t_2) = [x_1 - \langle x(t_1) \rangle] p(x_1; t_1) \frac{\phi_{xx}(t_1, t_2)}{\sigma_x^2(t_1)} \quad (6.39)$$

$$\text{for } t_1 \geq t_2 \text{ only}$$

then equations (6.38) become:

$$\phi_{zx}(t_1, t_2) = \frac{\phi_{zx}(t_1, t_1)}{\phi_{xx}(t_1, t_1)} \phi_{xx}(t_1, t_2) \quad (6.40)$$

for $t_1 \geq t_2$ only

also the expansion coefficients satisfy the condition

$$D_{n1}(t_1, t_2) = b_n(t_1) D_{11}(t_1, t_2) = b_n(t_1) \rho_{xs}(t_1, t_2) \quad (6.41)$$

for $t_1 \geq t_2$ only

and for a single process

$$\left. \begin{aligned} C_{11}(t_1, t_2) &= \rho_{xx}(t_1, t_2) \\ C_{n1}(t_1, t_2) &= 0 \quad n \geq 2 \end{aligned} \right\} \text{ for } t_1 \geq t_2 \text{ only} \quad (6.42)$$

Since $C_{n1}(t_1, t_2) = C_{1n}(t_2, t_1)$ it follows that for a single process:

$$C_{1n}(t_2, t_1) = 0 \quad n \geq 2 \quad t_1 \geq t_2 \text{ only}$$

Thus for a single semi-separable process, if the first variable is semi-separable negative with respect to the second, then the second is semi-separable positive with respect to the first.

This statement follows directly from the symmetry of the second probability density in the pairs of variables x, t_1 and x, t_2 .

It is also interesting to note that a semi-separable negative process can be turned into a semi-separable positive process by taking the 'time image' of it, e.g. by recording it and replaying it backwards. A physical example of a semi-separable process, given below, is used in the prediction example of chapter 8, whilst its 'image' is used in the example on approximate linear filters, section 9.1.3.

Like the separable property, the semi-separable property is preserved under linear transformation of the second variable. The proof for semi-separable processes is the same as that given in section 6.2.3 for separable processes, except that we have the restriction $t_1 \leq t_2$ or $t_2 \leq t_1$ on equations (6.32) and (6.33). By exactly analogous arguments we are lead to the knowledge of the existence of Δ_1 & Δ_2 of second probability density function, ^{which are} of interest in non-linear filtering and defined as follows:

6.3.2 Classes Δ_1 and Δ_2 .

A class Δ_1 of second probability density functions

$p_2(x, s; t_1, t_2)$ can be defined, for which an expansion of the form of equation (3.8) is possible and the expansion coefficients of which satisfy the following equations:

$$\left. \begin{aligned} D_{11}(t_1, t_2) &= \rho_{xz}(t_1, t_2) \\ D_{nn}(t_1, t_2) &= 0 \quad n \geq 2 \end{aligned} \right\} \text{ for } t_1 \geq t_2 \text{ only} \quad (6.43)$$

Also a class Δ_1 can be defined as for Δ_2 except the conditions 6.43 apply over the range $t_2 \geq t_1$ only.

6.33 An example of a semi-separable process.

Wonham¹⁹ in a recent paper obtains the transition probability functions for the output of a single stage low pass RC network when the input is excited by a random telegraph signal of the type described by Rice²⁰. In an interval of continuity, the random telegraph signal has a value $S(t) = \pm 1$; n changes of sign occur during an interval of length T with probability given by the Poisson distribution

$$p(K; T) = \frac{(\mu T)^n}{n!} e^{-\mu T} \quad (6.44)$$

The low pass filter function response is given by:

$$\frac{1}{RC} e^{-t/RC}$$

first and

The second probability densities of the filter output are easily derived from Wonham's transition probability function and for the case $\beta = \mu RC = 1$; $\mu = RC = 1$

$$\left. \begin{aligned} p_2(x_1, x_2; \tau) &= \frac{1}{4} + \frac{\psi(1+x_1)}{4} \delta[x_2 - (\psi x_1 + 1 - \psi)] \\ &\quad + \frac{\psi(1-x_1)}{4} \delta[x_2 - (\psi x_1 - 1 + \psi)] \end{aligned} \right\} \quad (6.45)$$

$$p_1(x) = \frac{1}{2}$$

where $\psi = e^{-\tau}$ $\tau \geq 0$

and the range of x_2 is $\psi x_1 - 1 + \psi$ to $\psi x_1 + 1 - \psi$

whilst for x_1 , the range is ± 1 .

Separability for $\tau \geq 0$ is easily verified by using equations

(6.16)

$$g(x_1; \tau) = \int_{\psi x_1 - 1 + \psi}^{\psi x_1 + 1 - \psi} x_2 \left[\frac{1}{4} + \frac{\psi(1+x_1)}{4} \delta\{x_2 - (\psi x_1 + 1 - \psi)\} + \frac{\psi(1-x_1)}{4} \delta\{x_2 - (\psi x_1 - 1 + \psi)\} \right] dx_2$$

$$\tau \geq 0$$

This integral, which is easy to evaluate becomes:

$$g(x_1; \tau) = \frac{x_1(2-\psi)\psi}{2} \quad \tau \geq 0 \quad (6.46)$$

This expression is separable in the form:

$$g_2(\tau) = f_{xx}(\tau) = (2-\psi)\psi$$

$$g_1(x_1) = \frac{x_1}{2} = x_1 p(x_1) \quad \tau \geq 0 \quad (6.47)$$

It is not so easy to establish that $p_2(x_1, x_2; \tau)$ is not separable for $\tau < 0$ since the bounds of x_2 depend upon x_1 and $p_2(x_1, x_2; \tau)$ is a discontinuous function. In chapter 8 the coefficients of the ^{two}dimensional expansion of $p_2(x_1, x_2; \tau)$ (equation 3.8) are worked out and shown to satisfy equations 6.42 but not equations 6.28. Equations 6.42 are both necessary and sufficient for semi-separability provided the expansion 3.8 exists. A more direct way of establishing semi-separability is by a characteristic function method. For this purpose a modified characteristic function is defined (see Nuttall reference 16).

$$f(u, t) = \int e^{ju(x - \langle x \rangle)} p(x, t) dx \quad \left. \vphantom{\int} \right\} (6.48)$$

Similarly:

$$f(u, u_2; t, t_2) = \iint e^{ju_1(x_1 - \langle x_1 \rangle)} e^{ju_2(x_2 - \langle x_2 \rangle)} p(x_1, x_2; t, t_2) dx_1 dx_2$$

(These definitions differ from those of equations 2.26 and 2.30 in the subtraction of the means $\langle x_1 \rangle$ and $\langle x_2 \rangle$)

Let

$$\begin{aligned} G(u_1; t, t_2) &= \frac{\partial f(u, u_2; t, t_2)}{\partial u_2} \bigg|_{u_2=0} \\ &= j \iint (x_2 - \langle x_2 \rangle) e^{ju_1(x_1 - \langle x_1 \rangle)} p(x_1, x_2; t, t_2) dx_1 dx_2 \end{aligned}$$

$$\therefore G(u,; t_1, t_2) = j \int e^{ju_1(x_1 - \langle x_1 \rangle)} g(x_1; t_1, t_2) dx_1 \quad (6.49)$$

(using equation 6.16)

Thus $G(u,; t_1, t_2)$ is the Fourier transform of $g(x_1; t_1, t_2)$ and if g is separable, in either half range $\tau > 0$ or $\tau < 0$, so also is G . Conversely, from the inverse transform it can be seen that if G separates so also does g .

Using equation 6.24 for a single separable process and substituting in equation 6.49 gives:

$$G(u,; t_1, t_2) = j \frac{\phi_{xx}(t_1, t_2)}{\sigma_x^2(t_1)} \int e^{ju_1(x_1 - \langle x_1 \rangle)} (x_1 - \langle x_1 \rangle) p(x_1, t_1) dx_1$$

But

$$\frac{df(u_1, t_1)}{du_1} = j \int (x_1 - \langle x_1 \rangle) e^{ju_1(x_1 - \langle x_1 \rangle)} p(x_1, t_1) dx_1$$

(from equation 6.48)

$$\therefore G(u,; t_1, t_2) = \frac{df(u_1, t_1)}{du_1} \frac{\phi_{xx}(t_1, t_2)}{\sigma_x^2(t_1)} \quad (6.50)$$

For a stationary process

$$G(u_1, t_1) = \frac{d f(u_1)}{d u_1} \rho(\tau) \quad (6.51)$$

$$\text{where } \rho(\tau) = \frac{\phi_{xx}(\tau)}{\sigma_x^2}$$

For the present example, using equations 6.48 and 6.45,

$$\begin{aligned} f(u) &= \int_{-1}^{+1} e^{jux} \frac{1}{2} dx \\ &= \frac{1}{2j u} (e^{ju} - e^{-ju}) = \frac{\sin u}{u} \end{aligned} \quad (6.52)$$

Similarly it can be shown, using equations 6.48 and 6.45

that:

$$\begin{aligned} f(u_1, u_2; \tau) &= \frac{\sin(u_1 + \psi u_2)}{u_1 + \psi u_2} \left[\frac{1}{u_2} \sin\{(1-\psi)u_2\} + \psi \cos\{u_2(1-\psi)\} \right] \\ &+ \psi \sin\{(1-\psi)u_2\} \left[\frac{\cos(u_1 + \psi u_2)}{u_1 + \psi u_2} - \frac{\sin(u_1 + \psi u_2)}{(u_1 + \psi u_2)^2} \right] \end{aligned} \quad (6.53)$$

$$\text{where } \psi = e^{-\tau} \quad \tau \geq 0$$

As a check we put $u_2 = 0$ whence:

$$f(u_1, 0; \tau) = \frac{\sin u_1}{u_1} [1 - \psi + \psi] + 0 = f(u_1)$$

is
which is the result to be expected.

When $u_1 = 0$ we get:

$$\begin{aligned}
 f(0, u_2; \tau) &= \frac{\sin \psi u_2}{\psi u_2} \left[\frac{1}{u_2} \sin \{(1-\psi)u_2\} + \psi \cos \{u_2(1-\psi)\} \right] \\
 &+ \psi \sin \{(1-\psi)u_2\} \left[\frac{\cos \psi u_2}{\psi u_2} - \frac{\sin \psi u_2}{\psi^2 u_2^2} \right] \\
 &= \frac{1}{u_2} \sin [\psi u_2 + u_2 - \psi u_2] = \frac{\sin u_2}{u_2} = f(u_2)
 \end{aligned}$$

Also

$$\begin{aligned}
 \frac{\partial f(u_1, u_2; \tau)}{\partial u_2} &= \frac{\psi(u_1 + \psi u_2) \cos(u_1 + \psi u_2) - \psi \sin(u_1 + \psi u_2)}{(u_1 + \psi u_2)^2} \\
 &\times \left[\frac{1}{u_2} \sin \{(1-\psi)u_2\} + \psi \cos \{u_2(1-\psi)\} \right] \\
 &+ \frac{\sin(u_1 + \psi u_2)}{u_1 + \psi u_2} \left[\frac{u_2(1-\psi) \cos \{u_2(1-\psi)\} - \sin \{(1-\psi)u_2\}}{u_2^2} \right. \\
 &\quad \left. - \psi(1-\psi) \sin \{u_2(1-\psi)\} \right] \\
 &+ \psi(1-\psi) \cos \{(1-\psi)u_2\} \left[\frac{\cos(u_1 + \psi u_2)}{u_1 + \psi u_2} - \frac{\sin(u_1 + \psi u_2)}{(u_1 + \psi u_2)^2} \right] \\
 &+ \psi \sin \{(1-\psi)u_2\} \left[\frac{-\psi(u_1 + \psi u_2) \sin(u_1 + \psi u_2) - \psi \cos(u_1 + \psi u_2)}{(u_1 + \psi u_2)^2} \right. \\
 &\quad \left. - \frac{-\psi(u_1 + \psi u_2) \cos(u_1 + \psi u_2) + 2\psi \sin(u_1 + \psi u_2)}{(u_1 + \psi u_2)^3} \right]
 \end{aligned}$$

(6.54)

Putting $u_2 = 0$ gives,

$$\begin{aligned} & \psi \left[\frac{\cos u_1}{u_1} - \frac{\sin u_1}{u_1^2} \right] (1 - \psi + \psi) + \frac{\sin u_1}{u_1} [0 - 0] \\ & + \psi(1 - \psi) \left[\frac{\cos u_1}{u_1} - \frac{\sin u_1}{u_1^2} \right] + 0 \\ \therefore \frac{\partial f(u_1, u_2; \tau)}{\partial u_2} \bigg|_{u_2=0} &= \psi(2 - \psi) \left[\frac{\cos u_1}{u_1} - \frac{\sin u_1}{u_1^2} \right] \quad (6.55) \\ & \text{where } \psi = e^{-\tau} \quad \tau \geq 0 \end{aligned}$$

Comparison of this result with equations 6.51 and 6.52 shows that χ_1 is separable with respect to x_2 for $\tau \geq 0$. This result was proved more simply on page 97, but the characteristic function method can be used to show that the process is not separable for $\tau < 0$. Since equations 6.45 and 6.53 are for $\tau \geq 0$ only, we use the fact that if χ_1 is semi-separable positive with respect to x_2 , then χ_2 is semi-separable negative with respect to x_1 . Therefore we shall test for separability of $G(u_2; \tau)$.

$$\begin{aligned} \frac{\partial f(u_1, u_2; \tau)}{\partial u_1} &= \frac{(u_1 + \psi u_2) \cos(u_1 + \psi u_2) - \sin(u_1 + \psi u_2)}{(u_1 + \psi u_2)^2} \\ &\times \left[\frac{1}{u_2} \sin \{ (1 - \psi) u_2 \} + \psi \cos \{ u_2 (1 - \psi) \} \right] \\ &+ \psi \sin \{ (1 - \psi) u_2 \} \left[\frac{-(u_1 + \psi u_2) \sin(u_1 + \psi u_2) - \cos(u_1 + \psi u_2)}{(u_1 + \psi u_2)^2} \right. \\ &\quad \left. - \frac{(u_1 + \psi u_2) \cos(u_1 + \psi u_2) + 2 \sin(u_1 + \psi u_2)}{(u_1 + \psi u_2)^3} \right] \end{aligned}$$

Putting $u_1 = 0$ gives:

$$\begin{aligned}
 G(u_2; \tau) &= \left. \frac{\partial f(u_1, u_2; \tau)}{\partial u_1} \right|_{u_1=0} \\
 &= \left[\frac{\cos \psi u_2}{\psi u_2} - \frac{\sin \psi u_2}{\psi^2 u_2^2} \right] \left(\psi \cos \{u_2(1-\psi)\} \right) \\
 &\quad + \psi \sin \{(1-\psi)u_2\} \left[-\frac{\sin \psi u_2}{\psi u_2} - \frac{\cos \psi u_2}{\psi^2 u_2^2} + \frac{\sin \psi u_2}{\psi^3 u_2^3} \right]
 \end{aligned} \tag{6.56}$$

where $\psi = e^{-\tau}$ $\tau \geq 0$

This expression is separable only at $\tau=0$.

6.4 A theorem on optimum multipath filters.

Although for many cases of input and desired output a non-linear filter may be capable of giving considerably smaller mean-squared error than an optimum linear filter, there are situations in which the performance of the optimum non-linear filter is no better than an optimum linear filter.

For instance, the following theorem may be proved:

If $p_2(x_1, x_2; \tau)$ belongs to class Δ and $p_2(x_1, z; \tau)$ belongs to class Δ_1 , then an optimum multipath filter of any degree is linear and contains only one path.

The proof is trivial since for classes Δ and Δ_1 , the

expansion for the probability density functions $p_2(x_1, x_2; \tau)$ and $p_2(x_1, z; \tau)$ is valid and the coefficients satisfy the following conditions

$$C_{nm}(\tau) \equiv 0 \quad \begin{matrix} n=1 & m \neq 1 \\ m=1 & n \neq 1 \end{matrix} \quad \text{for} \quad p_2(x_1, x_2; \tau) \quad (6.57)$$

$$D_n(\tau) \equiv 0 \quad n \neq 1 \quad \tau \geq 0 \quad \text{for} \quad p_2(x_1, z; \tau) \quad (6.58)$$

Therefore the simultaneous integral equations 5.27 can be satisfied by putting

$$W_m(\tau) \equiv 0 \quad m = 2, 3, \dots, R$$

and

$$\int_0^\infty C_{11}(\tau_1 - \tau_2) W_1(\tau_2) d\tau_2 = \sigma_z D_{11}(\tau_1) \quad \tau_1 \geq 0$$

Therefore $W_1(\tau)$ is given by the solution of the Wiener Hopf equation and the optimum multipath filter is linear

since

$$\theta_1[x] = \frac{x - \bar{x}}{\sigma_x} \quad (\text{see chapter 3}).$$

Two corollaries of this theorem are:

1. The optimum multipath predictor is linear if its input is of class Δ

This corollary is easy to prove as follows:

$$D_{nm}(\tau) = C_{nm}(\tau + \tau) \quad \text{for prediction problems.}$$

But for class Δ , $C_{n1}(\tau) \equiv C_{1m}(\tau) \equiv 0$ $\begin{matrix} n \neq 1 \\ m \neq 1 \end{matrix}$. Therefore both conditions 6.57 and 6.58 are satisfied and the predictor is linear.

11. If the expansion 4.25 for the Kernel function of the optimum filter of class η_1 is exact, then the optimum filter of class η_1 is linear if both $\rho_2(x_1, x_2; \tau)$ and $\rho_2(x, z; \tau)$ are of classes Δ and Δ , respectively.

7.0 NON-LINEAR NOISE FILTERING.

In noise filtering problems the input to the filter consists of some function of the desired output and an unwanted component called noise.

$$x(t) = k [z(t), v(t)] \quad (7.1)$$

where $x(t)$ is the filter input

$z(t)$ is the desired output of the filter

$v(t)$ is the noise component of the input.

In practice, noise may be introduced during transmission of the desired signal, by line, radio or other means, and may be due to random causes, such as atmospherics, thermal or shot effects, or cross talk from a neighbouring channel, or it may be caused by distortion.

Two special cases only are considered in this thesis;

- i. when the noise and signal are linearly additive
- ii. when they are multiplicative.

Undoubtedly the optimization procedure outlined below could be extended to cover cases when the noise and signal are combined partly in a multiplicative fashion and partly additively, but the above two cases are sufficient for most practical needs.

7.1 Additive signal and noise.

For additive signal and noise we may write:

$$x(t) = a z(t) + b v(t) \quad (7.2)$$

where a and b are constants.

The various moments required for the determination of the orthonormal polynomials and the expansion coefficients

$C_{nm}(\tau)$ and $D_{nm}(\tau)$, which are required for the optimization of the non-linear multipath filter are given by:

$$\overline{x^n} = \sum_{r=0}^n \binom{n}{r} a^r b^{n-r} \overline{z^r v^{n-r}} \quad (7.3)$$

where $\binom{n}{r}$ is the r th binomial coefficient.

Similarly

$$\overline{x_1^n x_2^m} = \sum_{r=0}^n \sum_{s=0}^m \binom{n}{r} \binom{m}{s} a^{r+s} b^{n+m-r-s} \overline{z_1^r z_2^s v_1^{n-r} v_2^{m-s}} \quad (7.4)$$

where the suffixes indicate different instants in time, although since we are considering ^{only} stationary signals $\overline{x_1^n x_2^m}$ is a function of the time interval τ only.

If $z(t)$ and $v(t)$ are statistically independent:

$$\overline{x^n} = \sum_{r=0}^n \binom{n}{r} a^r b^{n-r} \overline{z^r} \overline{v^{n-r}} \quad (7.5)$$

and

$$\overline{x_1^n x_2^m} = \sum_{r=0}^n \sum_{s=0}^m \binom{n}{r} \binom{m}{s} a^{r+s} b^{n+m-r-s} \overline{z_1^r z_2^s} \overline{v_1^{n-r} v_2^{m-s}} \quad (7.6)$$

From the set of moments $\overline{x^n}$ up to $n = 2N$, we can generate the first N polynomials, $\theta_n[x]$ using equation (3.6), also the coefficients $C_{nm}(\tau)$ can be obtained from the functions $\overline{x_1^n x_2^m}(\tau)$ by repeated use of equation (3.6).

Furthermore:

$$\overline{x_1^n z_2} = \sum_{r=0}^n \binom{n}{r} a^r b^{n-r} \overline{z_1^r z_2} \overline{v^{n-r}} \quad (7.7)$$

If $z(t)$ and $v(t)$ are statistically independent, then:

$$\overline{x_1^n z_2} = \sum_{r=0}^n \binom{n}{r} a^r b^{n-r} \overline{z_1^r z_2} \overline{v^{n-r}} \quad (7.8)$$

from which the coefficients $\phi_{zD_n}(\tau)$ can be found using equation (3.6).

7.2 Multiplicative signal and noise.

For multiplicative signal and noise, similar equations exist:

$$\overline{x^n} = \overline{z^n v^n} \quad (7.9)$$

or

$$\overline{x^n} = \overline{z^n} \overline{v^n} \quad (7.10)$$

(if z and v are statistically independent)

also

$$\overline{x_1^n x_2^m} = \overline{z_1^n v_1^n z_2^m v_2^m} \quad (7.11)$$

$$\therefore \overline{x_1^n x_2^m} = \overline{z_1^n z_2^m} \overline{v_1^n v_2^m} \quad (7.12)$$

(if z and v are statistically independent

and

$$\overline{x_1^n z_2} = \overline{z_1^n z_2} \overline{v_1^n} \quad (7.13)$$

$$\therefore \overline{x_1^n z_2} = \overline{z_1^n z_2} \overline{v_1^n} \quad (7.14)$$

(if z and v are statistically independent).

7.3 Problem 1. The optimization of some non-linear filters for a random telegraph signal and Gaussian noise with identical spectra.

To demonstrate that a non-linear filter can produce a substantially lower mean-squared error than the best linear filter, an example has been chosen in which the signal and noise are linearly additive and have the same correlation coefficient and hence the same spectral density.

The signal component of the input is a random square wave with variance K^2 , zero mean and has a first probability density function:

$$p_z(z) = \frac{1}{2} [\delta(z-K) + \delta(z+K)]$$

It can be shown that for this signal:

$$\overline{z_1^n z_2^m} = \frac{K^{n+m}}{4} \left\{ [1+(-1)^m][1+(-1)^n] + \rho [1-(-1)^m][1-(-1)^n] \right\} \quad (7.15)$$

where ρ , the normalized auto-correlation coefficient, is exponential

$$\overline{z_1 z_2} = \rho = e^{-\rho|\tau|} \quad (7.16)$$

The noise is Gaussian with variance σ^2 , zero mean and a first probability density function:

$$p_v(v) = \frac{1}{\sqrt{2\pi} \sigma} e^{-v^2/2\sigma^2} \quad (7.17)$$

The joint moments $\overline{v_1^n v_2^m}$ for the Gaussian noise are conveniently determined from the joint characteristic function, given below, using equation (3.13)

$$\psi(u_1, u_2; \tau) = e^{-\frac{1}{2}(u_1^2 + 2\rho u_1 u_2 + u_2^2)\sigma^2} \quad (7.18)$$

where ρ the normalized auto-correlation coefficient is exponential

$$\overline{v_1 v_2} = \rho = e^{-\rho|\tau|} \quad (7.19)$$

The filter input is given by:

$$x(t) = z(t) + v(t) \quad (7.2)$$

The signal and noise are independent.

7.3.1 A zero memory filter.

Since the noise and signal are independent and added linearly

$$\begin{aligned}
 p_z(x, z) &= p_v(x-z) p_z(z) \\
 &= \frac{1}{2\sqrt{2\pi}\sigma} \left[\delta(z-k) + \delta(z+k) \right] e^{-\frac{(x-z)^2}{2\sigma^2}}
 \end{aligned}
 \tag{7.20}$$

The optimum zero memory filter function is given by direct substitution in equation (5.14)

$$f[x] = \frac{\int_{-K}^K \frac{1}{2\sqrt{2\pi}\sigma} \left[\delta(z-k) + \delta(z+k) \right] e^{-\frac{(x-z)^2}{2\sigma^2}} dz}{\int_{-K}^K \frac{1}{2\sqrt{2\pi}\sigma} \left[\delta(z-k) + \delta(z+k) \right] e^{-\frac{(x-z)^2}{2\sigma^2}} dz}$$

which can be reduced to:

$$y = f[x] = K \tanh \frac{Kx}{\sigma^2}
 \tag{7.21}$$

As $\frac{K}{\sigma}$ increases, the error decreases. For small values of $\frac{K}{\sigma}$ there is a high probability that the value of x will

be negative when z is positive, such a condition gives rise to large error (see fig.9).

7.3.2 Multipath zero memory filters.

The following moments of the Gaussian process and random square-wave are easily derived using equations 3.13, 7.18 and 7.15. signal and noise.

$$\overline{V^n} = \sigma^n \times 1 \times 3 \times 5 \times 7 \dots (n-1) \text{ for } n \text{ even} \quad (7.22)$$

$$= 0 \text{ for } n \text{ odd}$$

$$\begin{aligned} \overline{z^n} &= K^n \text{ for } n \text{ even} \\ &= 0 \text{ for } n \text{ odd} \end{aligned} \quad (7.23)$$

The orthonormal polynomials are then found, using equations (7.4) and (3.6), for three different cases of K/σ , also the values of $\sigma_z D_n(0)$ are computed by direct evaluation of $\overline{\theta_n(x)z}$ using only the data given in equations (7.22) and (7.23) above. The polynomials and values of $\sigma_z D_n(0)$ are tabulated below:

The multipath filter functions, as listed below, are obtained from the expression:

$$y = \sum_{n=1}^R \sigma_z D_n \theta_n [x] \quad (7.24)$$

Since $\overline{z} = 0$ in equation (5.19)

* In all the computations at least six decimal places were used and sometimes more, if necessary, although only three or four decimal figures are recorded in the tables.

$K/\sigma_z = 1$	$K^2 + \sigma_z^2 = 1$		
$\theta_1(x)$	x	$\sigma_z^{D_{11}}(0)$	0.5000
$\theta_2(x)$	$-0.8165 + 0.8165x^2$	$\sigma_z^{D_{21}}(0)$	0
$\theta_3(x)$	$-1.3868x + 0.5547x^3$	$\sigma_z^{D_{31}}(0)$	-0.1387
$\theta_4(x)$	$0.7290 - 1.5702x^2 + 0.3365x^4$	$\sigma_z^{D_{41}}(0)$	0
$\theta_5(x)$	$1.6564x - 1.3650x^3 + 0.1848x^5$	$\sigma_z^{D_{51}}(0)$	-0.0640

$K/\sigma_z = 2$	$K^2 + \sigma_z^2 = 1$		
$\theta_1(x)$	x	$\sigma_z^{D_{11}}(0)$	0.8000
$\theta_2(x)$	$-1.1785 + 1.1785x^2$	$\sigma_z^{D_{21}}(0)$	0
$\theta_3(x)$	$-1.6918x + 0.9836x^3$	$\sigma_z^{D_{31}}(0)$	-0.2518
$\theta_4(x)$	$1.2215 - 2.6852x^2 + 0.8509x^4$	$\sigma_z^{D_{41}}(0)$	0
$\theta_5(x)$	$2.3127 - 2.7823x^3 + 0.6194x^5$	$\sigma_z^{D_{51}}(0)$	0.1414

$K/\sigma_z = 4$	$K^2 + \sigma_z^2 = 1$		
$\theta_1(x)$	x	$\sigma_{zD_{11}}(0)$	0.9412
$\theta_2(x)$	$-2.0926 + 2.0926x^2$	$\sigma_{zD_{21}}(0)$	0
$\theta_3(x)$	$-2.4378x + 1.9764x^3$	$\sigma_{zD_{31}}(0)$	-0.2060
$\theta_4(x)$	$3.3232 - 6.9647x^2 + 2.9645x^4$	$\sigma_{zD_{41}}(0)$	0
$\theta_5(x)$	$4.4293x - 7.4235x^3 + 2.65708x^5$	$\sigma_{zD_{51}}(0)$	0.0895

K/σ_z	$K^2 + \sigma_z^2 = 1$
1	$y = 0.7983x - 0.1643x^3 + 0.01183x^5$
2	$y = 1.5530x - 0.6411x^3 + 0.08759x^5$
4	$y = 1.8377x - 1.0716x^3 + 0.2378x^5$

The^{above} polynomial functions are plotted in fig.9, for each value of K/σ_z , to the same axes as the filter function

$$y = K \tanh \frac{Kx}{\sigma_{\text{auto}}^2}$$

For this example, since the_{the} correlation coefficients of the noise and/signal are identical, the optimum linear filter will be a simple attenuator with a gain constant $\mathcal{D}_{11}$. The mean-squared error of the^{optimum} attenuator is given by:

$$\overline{e_a^2} = \sigma_z (1 - D_{11}^2) \quad (5.31)$$

The ratio of the mean-squared error to $\overline{e_a^2}$ is computed using equations (5.21) and (5.31) for each of the optimum filters.

These ratios, referred to subsequently as the performance index, are tabulated below and plotted in fig.10, for different numbers of paths.

K/σ_z	$\overline{e^2}$	$\overline{e^2} / \overline{e_a^2}$
1	0.5000	1.0000
	0.4615	0.9231
	0.4534	0.9067
2	0.2000	1.0000
	0.1207	0.6037
	0.0958	0.4788
4	0.0588	1.0000
	0.0137	0.2337
	0.0052	0.0890

Results, are plotted for different numbers of paths.

The following conclusions arise out of the results:-

1. As expected, the optimum zero-memory non-linear filter gives in some cases a substantially better performance than the best linear filter.

2. Also as expected, the zero memory filter performance improves for larger values of K/σ .

3. The mean-squared error converges rapidly with number of paths, indicating that only few paths may be needed, thus simplifying the optimization and the filter synthesis.

4. The multipath zero memory filter functions give a reasonably good approximation to the corresponding tanh curve over the working range.

In a sense, this example is trivial, since the filter function $f(x) = K \tanh \frac{Kx}{\sigma}$ can be found so easily, thus removing the need to use multipath filters. However, this example does enable a comparison of the two methods to be obtained, ^(fig.9) and in other cases it is not always possible to evaluate the integral of equation (5.14) except by numerical methods, for instance, as in the next problem.

7.3.3 A multipath memory filter.

For the case when $K/\sigma = 2$ the optimum multipath memory filter is designed as follows:-

The joint moments $\overline{x_1^n x_2^m}$ and $\overline{x_1^n z}$ are computed using equations (7.4), (7.8) and the example data in equations (7.15) and (7.18) and the fact that the signal and noise are statistically independent.

The expansion coefficients $C_{nm}(\tau)$ are then computed from the joint moments $\overline{x_1^n x_2^m}$ and $\sigma_z^2 D_{n1}(\tau)$ from the joint moments $\overline{x_1^n z}$ as described in section 3.4.

The results are given below:

$$\begin{aligned}
 C_{11}(\tau) &= \rho; & C_{22}(\tau) &= \rho^2 \\
 C_{33}(\tau) &= 0.3963\rho + 0.6037\rho^3 \\
 C_{35}(\tau) &= -0.2225\rho + 0.2225\rho^3 \\
 C_{44}(\tau) &= 0.5273\rho^2 + 0.4727\rho^4 \\
 C_{55}(\tau) &= 0.1250\rho + 0.5656\rho^3 + 0.3094\rho^5 \\
 \text{all other } C_{nm}(\tau) &= 0 \text{ for } 1 \leq \frac{n}{m} \leq 5 \\
 D_{11}(\tau) &= 0.8000\rho \\
 D_{31}(\tau) &= -0.2518\rho; & D_{21}(\tau) &= 0 \\
 D_{51}(\tau) &= 0.1414\rho; & D_{41}(\tau) &= 0 \\
 \text{where } \rho &= e^{-\beta|\tau|}
 \end{aligned} \tag{7.25}$$

It would appear that the probability density function $p_2(x_1, x_2; \tau)$ is of class Δ and it appears that it belongs to an even more specialized class, since $C_{n2}(\tau) = C_{2n}(\tau) = 0$

$n = 1, 2, \dots, 5.$

It can be shown that $p_2(x_1, x_2; \tau)$ is separable if $p_2(z_1, z_2; \tau)$ and $p_2(v_1, v_2; \tau)$ are separable and possess the same correlation coefficient (see reference 16).

Since $\beta_2(x, x_2; \tau)$ is of class Δ , the first equation of the set of equations 5.27 is independent of the remainder; their solution (see appendix 13.3) is thereby simplified. The optimum weighting function for 3 paths ($R=3$) are found to be:

$$w_1(\tau) = 0.8000 \delta(\tau)$$

$$w_3(\tau) =$$

(7.26 a)

whilst for 5 paths ($R=5$)

$$w_1(\tau) = 0.8000 \delta(\tau)$$

$$w_3(\tau) = -0.1835 \delta(\tau) - 0.1117 \beta e^{-4.1414 \beta \tau} - 0.2288 \beta e^{-1.8165 \beta \tau} \quad (7.26b)$$

$$w_5(\tau) = 0.0960 \delta(\tau) - 0.0405 \beta e^{-4.1414 \beta \tau} + 0.1555 \beta e^{-1.8165 \beta \tau}$$

R	$\overline{e^2}$	$\frac{\overline{e^2}}{e_a^2}$
1	0.2000	1.0000
3	0.1160	0.5799
5	0.0879	0.4393

The optimum filter is shown in fig.11. The performance index for this memory filter is worked out using equations (5.29) and (5.31) and plotted in fig.10. It is concluded that only slight improvement over the zero memory case is achieved.

7.4 Problem 2. The optimization of some non-linear^{noise} filters for a continuous input signal and Gaussian noise with identical spectra.

In this second example the signal and noise again have the same^{auto-} correlation coefficient and they are additive. The difference in this problem lies in the fact that both signal and noise are continuous and the signal has an asymmetrical probability density.

The signal component of the input is generated by passing Gaussian noise through a square law device

$$z(t) = \frac{s^2(t) - 1}{\sqrt{2}} = \frac{He_2[s(t)]}{\sqrt{2}} \quad (6.7)$$

where $s(t)$ is Gaussian noise with zero mean and unit variance but uncorrelated with $v(t)$.

The first probability is given by equation (6.13) and the moments, obtained using equation (6.15) are:

$$\begin{aligned} \overline{z} &= 0 & \overline{z^2} &= 1 & \overline{z^3} &= 2.8284 & \overline{z^4} &= 15 \\ \overline{z^5} &= 96.1665 & \overline{z^6} &= 755 & \overline{z^7} &= 6983.3866 & & (7.27) \\ \overline{z^8} &= 74417 & \overline{z^9} &= 897799.3379 & \overline{z^{10}} &= 12096873 \end{aligned}$$

The moments of the Gaussian noise with unit variance and zero mean are as given in equation (7.22).

The filter input is given by:

$$x(t) = \frac{z(t) + v(t)}{\sqrt{2}} \quad (7.28)$$

The signal and noise are statistically independent.

7.4.1 A multipath zero memory filter.

Following the same procedure as before, the orthonormal polynomials $\theta_n[x]$ and the coefficients $D_{n1}(0)$ are found to be

$\theta_1(x)$	x	D_{11}	0.7071
$\theta_2(x)$	$\frac{1}{2}(x^2 - x - 1)$	D_{21}	0.3536
$\theta_3(x)$	$\frac{1}{6.0622}(x^3 - 3.75x^2 - 2.25x + 2.75)$	D_{31}	-0.0875
$\theta_4(x)$	$\frac{1}{24.6469}(x^4 - 8.2041x^3 + 5.2653x^2 + 21.9592x - 3.0612)$	D_{41}	-0.0386
$\theta_5(x)$	$\frac{1}{125.0755}(x^5 - 11.4628x^4 + 40.0826x^3 + 50.9917x^2 - 103.3058x - 26.2975)$	D_{51}	0.0617

$$\sigma_z = 1 \quad (7.29)$$

The final polynomial becomes

$$y = \sum_{n=1}^5 D_n \theta_n(x) = 0.0005x^5 - 0.0087x^4 + 0.0182x^3 + 0.2478x^2 + 0.4774x - 0.2246 \quad (7.30)$$

The curve for this filter function is plotted in fig.12.

The performance index is worked out using equations (5.21) and (5.31) for different numbers of paths and tabulated on page 121. It is plotted against numbers of paths in fig.13.

The performance index again seems to converge rapidly to an asymptotic value as the number of paths is increased.

7.4.2 A multipath memory filter.

The joint Gaussian moments $\overline{V_1^n V_2^m}$ which were calculated for the previous set of examples are used in this example, since the noise is Gaussian and the correlation coefficient is again exponential.

$$\rho = e^{-\beta|\tau|}$$

The joint moments $\overline{z_1^n z_2^m}$ are calculated using equation (6.7) and the joint Gaussian moments for a correlation coefficient of $e^{-\frac{\beta|\tau|}{2}}$. Thus if

$$\overline{S_1 S_2} = e^{-\frac{\beta|\tau|}{2}} \quad \text{by equation (6.12) we have}$$

$$\overline{z_1 z_2} = e^{-\beta|\tau|}$$

The joint moments $\overline{x_1^n x_2^m}$ and $\overline{x_1^n z}$ are found exactly as before, from which the coefficients $C_{nm}(\tau)$ and $D_n(\tau)$ are found to be:

$$c_{11}(\tau) = \rho(\tau); \quad c_{1n}(\tau) = c_{n1}(\tau) = 0$$

$$c_{22}(\tau) = 0.25 [\rho(\tau) + 3\rho^2(\tau)]$$

$$c_{23}(\tau) = c_{32}(\tau) = \frac{0.75}{12.1244} [-\rho(\tau) + \rho^2(\tau)]$$

$$c_{24}(\tau) = c_{42}(\tau) = \frac{-1.3460\rho(\tau) + 1.3460\rho^2(\tau)}{49.25}$$

$$c_{33}(\tau) = \frac{0.5625\rho(\tau) + 15.1875\rho^2(\tau) + 21\rho^3(\tau)}{36.75} \quad (7.31)$$

$$c_{34}(\tau) = c_{43}(\tau) = \frac{1.0088\rho(\tau) - 26.7522\rho^2(\tau) + 25.6761\rho^3(\tau)}{149.2817}$$

$$c_{44}(\tau) = \frac{1.8097\rho(\tau) + 71.1243\rho^2(\tau) + 252.5364\rho^3(\tau) + 2820875\rho^4(\tau)}{606.3906}$$

$$D_{11}(\tau) = 0.7071\rho(\tau)$$

$$D_{21}(\tau) = 0.3536\rho(\tau)$$

$$D_{31}(\tau) = -0.0875\rho(\tau)$$

$$D_{41}(\tau) = -0.0386\rho(\tau)$$

$$D_{51}(\tau) = 0.0617\rho(\tau)$$

(7.32)

where $\rho(\tau) = e^{-\beta|\tau|}$

It can be proved that $\rho_2(x_1, x_2; \tau)$ is of class Δ , this is because both $\rho_2(z_1, z_2; \tau)$ and $\rho_2(v_1, v_2; \tau)$ are of class Δ and have the same correlation coefficients (see reference 16).

Since $\rho_2(x_1, x_2; \tau)$ is of class Δ for only two paths, the equations for $\psi_1(\tau)$ and $\psi_2(\tau)$ are independent, their solution is thus simplified and becomes

$$\begin{aligned}\psi_1(\tau) &= 0.7071 \delta(\tau) \\ \psi_2(\tau) &= 0.2761 \delta(\tau) + 0.2222 \rho e^{-1.1952 \rho \tau} \\ &= 0 \quad \tau < 0\end{aligned}\quad \begin{array}{l} \tau \geq 0 \\ (7.33) \end{array}$$

For three paths we obtain:

$$\begin{aligned}\psi_1(\tau) &= 0.7071 \delta(\tau) \\ \psi_2(\tau) &= 0.2766 \delta(\tau) + 0.2195 \rho e^{-1.2128 \rho \tau} \\ &\quad + 0.0008 \rho e^{-2.3717 \rho \tau} \\ &= 0 \quad \tau < 0 \\ \psi_3(\tau) &= -0.0624 \delta(\tau) - 0.0758 \rho e^{-1.2128 \rho \tau} \\ &\quad - 0.0126 \rho e^{-2.3717 \rho \tau} \\ &= 0 \quad \tau < 0\end{aligned}\quad \begin{array}{l} \tau \geq 0 \\ (7.34) \end{array}$$

R	Zero Memory		Memory	
	$\overline{e^2}$	$e^2 / \overline{e_a^2}$	$\overline{e^2}$	$\overline{e^2} / \overline{e_a^2}$
1	0.5000	1.0000	0.5000	1.0000
2	0.3750	0.7500	0.3666	0.7332
3	0.3673	0.7347	0.3583	0.7165
4	0.3659	0.7317	-	-
5	0.3621	0.7241	-	-

The block diagrams of these filters is shown in figs 14. and 15. and (5.31)

The performance index is worked out using equations (5.29) and plotted in fig.13. Again, little improvement is achieved in performance as compared with the zero memory filter.

7.5 Conclusions.

Although the analytical optimization procedure is tedious for a multipath memory filter, applications have been demonstrated where the labour and complexity might be justified; for instance, if ^{the} signal and noise have the same spectral density, although a linear noise filter is of no practical value, a multipath noise filter can produce a significant improvement in mean squared error.

Multipath zero memory noise filters also can be optimized using this procedure, which has advantages over the more direct method since, although an optimum zero memory filter function can be expressed explicitly in terms of probability function (equation 5.14), this expression involves integrals which often may be difficult to evaluate. In contrast to the labour involved in the optimization of

multipath memory filters, a multipath zero memory filter can be optimized very easily analytically (taking about 4 hours for an average problem).

Multipath memory filters may be connected in cascade so as to produce still lower mean-squared error. If an analogue computer is available, a second multipath memory filter can be optimized to operate on the output of an earlier filter; the mean-squared error generally will be reduced, because the combined filter is of higher classification, than a filter of class γ_1 , and can use more statistical information about the signal and noise. As the number of optimum cascaded filters is increased, the mean-squared filtering error will converge to either zero or some finite value which must be less than, or equal to, the original mean-squared error at the input to the cascaded chain. Thus the last of a large number of filters in the chain will not make an appreciable difference to the mean-squared error; intuitively, it seems that this filter will become a direct connection, whence in equation (5.26) and (5.27):

$$\begin{aligned} W_m(\tau_2) &= 0 \quad m \neq 1 \\ &= \sigma_y \delta(\tau_2) \quad m = 1 \\ &= \bar{y} \delta(\tau_2) \quad m = 0 \end{aligned} \tag{7.35}$$

Therefore the coefficients of the probability density expansion for the output and desired output satisfy the following equation

$$\delta_y C_{n_1}(T) = D_{n_1}(T) \delta_z$$

$$\text{also } \overline{y} = \overline{z} \quad (7.36)$$

If the input and desired output are such that they satisfy this equation, then the optimum filter of class $\mathcal{N}_{LR}$ is an attenuator.

A cascade connection of two optimum zero-memory filters is itself a zero-memory filter, so that it is to be expected that the second of the two filters will be a direct connection since the two filters cannot achieve a lower mean-square error than the first alone can achieve. But we can cascade a zero memory non-linear filter and a linear memory filter alternately and produce successively lower values of mean-square error since, after the non-linear filter, the linear filter input will not have the same correlation coefficient as its desired output. Such an arrangement may have considerable practical advantages due to the relative simplicity of the optimization procedure for both zero memory and linear memory filters.

8.0 NON-LINEAR PREDICTION.

For the optimization of a non-linear predictor, using the method of chapter 5.0, the desired output can be expressed in terms of the input by the following identity

$$z(t) = x(t + \tau) \quad (8.1)$$

where τ is the prediction time.

8.1 Problem 3. The optimization of some non-linear predictors for a smoothed random telegraph signal input.

The input signal, some properties of which are considered in section 6.33, is not a member of class Δ , so that the optimum multipath memory predictor is not expected to be linear (see section 6.4).

The second probability density function of the input is obtained from equation (6.45) after a slight alteration, which is needed to make the input variance unity

$$p_2(x_1, x_2; \tau) = \frac{1}{4} \left\{ \frac{1}{3} + \frac{\psi(1 + \frac{x_1}{\sqrt{3}})}{\sqrt{3}} \int \left[\frac{x_2}{\sqrt{3}} - \left(\frac{\psi x_1}{\sqrt{3}} + 1 - \psi \right) \right] \right. \\ \left. + \frac{\psi(1 - \frac{x_1}{\sqrt{3}})}{\sqrt{3}} \int \left[\frac{x_2}{\sqrt{3}} - \left(\frac{\psi x_1}{\sqrt{3}} - 1 + \psi \right) \right] \right\} \quad (8.2)$$

where $\psi = e^{-\tau}$ and $\tau \geq 0$

8.1.1 The optimum linear predictor.

The optimum linear predictor is given by the solution of the Wiener Hopf equation (5.32).

$$\int_0^{\infty} w(\tau_2) \phi_{xx}(\tau_1 - \tau_2) d\tau_2 = \phi_{xz}(\tau_1) \quad \tau_1 \geq 0 \quad (5.32)$$

where $\phi_{xx}(\tau) = e^{-|\tau|} (2 - e^{-|\tau|})$ see equation (8.3)
6.47

and $\phi_{xz}(\tau) = e^{-|\tau+\tau|} (2 - e^{-|\tau+\tau|})$ (8.4)

Since the equation (5.32) has to be satisfied when $\tau_1 \geq 0$ only, for convenience in the subsequent solution, we write:

$$\begin{aligned} \phi_{xz}(\tau) &= e^{-(\tau+\tau)} (2 - e^{-(\tau+\tau)}) \quad \begin{matrix} \tau \geq 0 \\ \tau > 0 \end{matrix} \quad (8.5) \\ &= 0 \quad \tau < 0 \end{aligned}$$

The solution, obtained by the method of appendix 13.3.2, is found to be:

$$w(\tau) = e^{-\tau} \left[\delta^{(2)}(\tau) + 2\delta(\tau) \right] - e^{-2\tau} \left[\delta^{(2)}(\tau) + \delta(\tau) \right] \quad (8.6)$$

where $\delta(\tau)$ is the Dirac delta function defined as

$$\begin{aligned} \delta(\tau) &= 0 \quad \tau \neq 0 \\ &= \infty \quad \tau = 0 \end{aligned}$$

so that $\int_{-\infty}^{\infty} \delta(\tau) d\tau = 1$

and $\delta^{(2)}(\tau)$ is the double delta function defined by

$$\delta^{(2)}(\tau) = \frac{d}{d\tau} \delta(\tau)$$

Some properties of higher order delta functions given in appendix 13.4.2 permit the mean-squared error to be calculated using equation (5.33) when $\delta_z = 1$

$$\overline{e^2} = 1 - \int_0^{\infty} w(\tau) \phi_{xz}(\tau) d\tau \quad (5.33)$$

The answers are

$\phi_{xz}(0)$	T	$\overline{e^2}$
0.5	1.228	0.6642
0.1	2.970	0.9853

8.1.2 An optimum multipath zero memory predictor.

It is easily shown that the first probability density of the input is of the form:

$$p(x) = \frac{1}{2\sqrt{3}} ; -\sqrt{3} \leq x \leq \sqrt{3} \quad (8.7)$$

= 0 elsewhere

and the moments are given by:

$$\overline{x^n} = \int_{-\sqrt{3}}^{\sqrt{3}} x^n \frac{1}{2\sqrt{3}} dx = \frac{\sqrt{3}^n}{n+1} \quad (8.8)$$

whence the polynomial functions for the input are obtained using equation (3.6):

$$\theta_0(x) = 1$$

$$\theta_1(x) = x$$

$$\theta_2(x) = \frac{\sqrt{5}}{2} (x^2 - 1)$$

$$\theta_3(x) = \frac{\sqrt{21}}{2} \left(\frac{x^3}{9} - x \right) \quad (8.9)$$

$$\theta_4(x) = \frac{1}{4} \left(\frac{35}{6} x^4 - 15x^2 + \frac{9}{2} \right)$$

$$\theta_5(x) = \frac{\sqrt{33}}{4} \left(\frac{7}{6} x^5 - \frac{35}{9} x^3 + \frac{5}{2} x \right)$$

The following equation, which is derived by a rather tedious piece of algebra, enables the moments $\overline{x_1^n x_2^m}$ to be obtained

$$\overline{x_1^n x_2^m} = (\sqrt{3})^{n+m} \sum_{r=0}^{n+m} \frac{{}^nC_r \psi^r (1-\psi)^{n-r}}{m+r+1} \left\{ \frac{r+1}{n-r+1} + \frac{n-2r}{n-r+1} \psi \right\} \quad (8.10)$$

for $n-r$ even   for $n+m$ even

$$\overline{x_1^n x_2^m} = 0 \quad \text{for } n + m \text{ odd} \quad (8.10)$$

From these moments and the polynomials functions, the expansion coefficients are found to be:

$$\begin{aligned} C_{11}(\tau) &= \psi(2-\psi) & C_{44}(\tau) &= \psi^4(5-4\psi) \\ C_{22}(\tau) &= \psi^2(3-2\psi) & C_{55}(\tau) &= \psi^5(6-5\psi) \\ C_{33}(\tau) &= \psi^3(4-3\psi) \end{aligned} \quad (8.11)$$

$$\begin{aligned} C_{31}(\tau) &= C_{13}(-\tau) = \frac{2\sqrt{21}}{3} \psi(1-\psi)^3 \tau > 0 \\ &= 0 \quad \tau \leq 0 \\ C_{42}(\tau) &= C_{24}(-\tau) = 6\sqrt{5} \psi^2(1-\psi)^3 \tau > 0 \\ &= 0 \quad \tau \leq 0 \\ C_{53}(\tau) &= C_{35}(-\tau) = 4\sqrt{77} \psi^3(1-\psi)^3 \tau > 0 \\ &= 0 \quad \tau \leq 0 \\ C_{51}(\tau) &= C_{15}(-\tau) = \frac{\sqrt{33}}{3} \psi(1-\psi)^3(1-3\psi)(2-3\psi) \tau > 0 \\ &= 0 \quad \tau \leq 0 \end{aligned} \quad (8.12)$$

where $\psi = e^{-|\tau|}$

These results clearly suggest that this process belongs to class Δ_2 , in fact, it probably belongs to a more restrictive class, since it seems that all

$$C_{nm}(\tau) = 0 \quad \tau \leq 0 \\ n \neq m$$

By using the relationship $D_{n_1}(\tau) = C_{n_1}(\tau + \tau)$, the values of $D_{n_1}(0)$ are found for $T = 1.228$:

$$\begin{aligned} D_{11}(0) &= 0.500 \\ D_{31}(0) &= 21 \times 0.0690 \\ D_{51}(0) &= 33 \times 0.004698 \end{aligned} \quad (8.13)$$

whence the predictor filter function is obtained:

$$\begin{aligned} y(t) &= D_{11}(0) \theta_1(x) + D_{31}(0) \theta_3(x) + D_{51}(0) \theta_5(x) \\ &= -0.1280 x(t) + 0.2512 x^3(t) + 0.0452 x^5(t) \end{aligned} \quad (8.14)$$

This function is plotted in figure 16. The mean squared error for one, three and five paths is worked out using equation (5.21) and (5.22).

R	$\overline{e^2}$
1	0.7500
3	0.6499
5	0.6492

8.1.3 An optimum multipath memory predictor.

The expansion coefficients $C_{nm}(\tau)$ and $D_{n_1}(\tau)$ are given earlier, it remains for the equations (5.27) to be solved by the method of undetermined coefficients (appendix 13.3).

The solutions are found to be:

$$\begin{aligned}
w_1(\tau) &= e^{-\tau}(2 - e^{-\tau})\delta(\tau) + e^{-\tau}(1 - e^{-\tau})\delta^{(2)}(\tau) \\
w_3(\tau) &= \frac{2\sqrt{21}}{3} e^{-\tau}(1 - e^{-\tau})^3\delta(\tau) + \frac{\sqrt{21}}{18} e^{-\tau}(1 - e^{-\tau})^2(1 - 4e^{-\tau})\delta^{(2)}(\tau) \\
w_5(\tau) &= \frac{\sqrt{33}}{3} e^{-\tau}(1 - e^{-\tau})^3(1 - 3e^{-\tau})(2 - 3e^{-\tau})\delta(\tau) \\
&\quad + \frac{\sqrt{33}}{45} e^{-\tau}(1 - e^{-\tau}) \left\{ \begin{aligned} &(1 - 3e^{-\tau})^2(2 - 3e^{-\tau}) \\ &- (1 - 2e^{-\tau}) \end{aligned} \right\} \delta^{(2)}(\tau) \quad (8.15)
\end{aligned}$$

The mean squared error is worked out using equation (5.29) (see appendix B.4.2) (see page 112) for 1, 3 and 5 paths. The performance index is plotted in fig.17 for both the zero memory and memory predictors for 1, 3 and 5 paths.

R	Zero Memory		Memory	
	$\overline{e^2}$	$\overline{e^2} / \overline{e_a^2}$	$\overline{e^2}$	$\overline{e^2} / \overline{e_a^2}$
1	0.7500	1.0000	6.6642	0.8857
3	0.6499	0.8665	0.5637	0.7516
5	0.6492	0.8656	6.5613	0.7484

In fig.18 curves are plotted showing the contribution to the reduction of mean-squared error of each path of both the zero memory and memory predictors as a function of the prediction time T .

8.2 Conclusions.

The curves of fig.18 show that the mean-squared error or performance index is not smoothly convergent as the number of paths is increased. This condition seems to exist also in problem 2, as demonstrated by the graphs of fig.13 showing a kind of oscillatory approach to the asymptotic value.

For this prediction problem there exists a particular value of τ , the prediction time for which the third filter path contributes more to the reduction of mean-squared error than the first path and also at some other prediction time the fifth path contributes more than the third; for this latter value of prediction time, a better two-path predictor can be constructed by using the fifth and first paths than by using the third and first.

Another interesting fact arising out of the solution is that the $\psi_m(\tau)$ functions are independent of each other, the reason is as follows. For this problem, all the coefficients $C_{nm}(\tau) = 0$ when $\tau = 0$ and $n \neq m$; therefore, if the solutions for $\psi_m(\tau)$ should be of the form $\psi_m(\tau) = \sum_i a_{mi} \delta^{(i)}(\tau)$ all the integrals on the left-hand sides of equations (5.27) are zero, except when $n=m$ so that the optimum $\psi_m(\tau)$ are given by the solutions of a set of independent integral equations, which are easily solved. ~~In this problem and many other~~

9.0 THE ANALYSIS OF NON LINEAR SYSTEMS

All physical systems are essentially non-linear, although in many cases, owing to the input being of sufficiently small amplitude, linear analysis can be used; in other cases linear approximation procedures are adequate, but there remains a class of systems for which more sophisticated analysis techniques must be applied. In this chapter approximate and exact analysis techniques for non-linear systems are considered and some approximate analysis techniques are proposed which may be useful where linearization procedures are inadequate.

By 'system' we mean any process, plant or operation which exists in the physical world, the variable quantities of which are capable of being represented by some mathematical scheme of dependence whether known or unknown. Some examples of systems are given below:

1. A communication link in which information is transmitted from A to B; to facilitate this operation the message may be encoded at the transmitter and decoded at the receiver in a manner which can be described mathematically.

2. A control system such as a feedback control in which an output quantity depends upon an input quantity in some linear or non-linear manner.*

---

* Some discussion of systems of this kind can be found in a series of articles by C.M.Burrell and J.K.Lubbock²¹, see particularly part 1.

3. A chemical plant in which, for example, the quality of some output product depends in a complicated non-linear manner (which is usually unknown) on a number of system variables such as temperature, pressure, flow.

Frequently the analysis of systems is complicated by the presence of unknown non-linearities and rarely is the mathematical description of the non-linearity already known. However, in both circumstances it may be possible to find a simple approximate structure, which behaves in a similar manner. The approximate structure need not be linear, although in some cases this is desirable (e.g. in feedback system analysis), because an adequate theory of linear systems already exists. The information needed to enable the approximate

structure to be designed can be obtained by experimental measurement on the response of the original system whilst working under normal operating conditions. Purely analytical methods of finding an approximate structure from a complete mathematical description of both the input to and character of, the original system, tend to be tedious if not impossible, so that experimental measurement on the actual or a simulated system or from operating records is indicated. This technique of experimental analysis of non-linear systems can be considered as an extension of the correlation function method of analysis^{22,23,24} for linear systems.

9.1 Systems with a single input

Given sufficient information about the input $x(t)$ and desired output $z(t)$ either in the form of signal samples or as statistical characteristics, an optimum filter can usually be determined by one of the procedures of chapter 5. In order to find the best approximate structure in the mean-squared error sense the same procedure may be adopted but ~~taking~~ the output of the actual system ^{becomes} the desired output $z(t)$ and its input as the filter input $x(t)$.

If the system belongs to the same class as the filter which is being optimized then an exact equivalence can be found using any input which is sufficiently random; a convenient choice of input is Gaussian white noise since then

the probability density function expansion is a single series (see equation 6.3) and in the optimization of a multipath memory filter, at least, the optimum weighting functions are each given by the solution of an independent integral equation.

If the system belongs to a more general class than the class of filters being optimized then the optimum filter response can only approximate to the system response and the optimum filter is dependent upon the input process. It is necessary therefore to carry out the optimization using the actual system input.

9.1.1 Best approximate attenuators

Booton^{25,26} has proposed an approximate method of analysis for relay systems; this method has been applied with success to other zero memory non-linear systems and consists essentially of replacing the non-linear system by an attenuator which gives the best approximation in the mean-squared error sense. The method could be used for more complex non-linear systems but the approximation may be too severe.

The equation for an attenuator is*

$$y(t) = k \left[x(t) - \bar{x} \right] + K \quad (4.20)$$

---

* This is a more general class of attenuator than that considered by Booton since we include a term K which is independent of x

where $x(t)$ is the attenuator input

$y(t)$ is the attenuator output

To find the best approximate attenuator we minimize the mean-squared error

$$\overline{e^2} = \overline{[z(t) - y(t)]^2}$$

where $z(t)$ is the non-linear system output with an input $x(t)$.

whence we obtain values for k and K as follows:

$$K = \overline{z(t)} \quad (9.1)$$

$$k = \frac{\phi_{xz}(0)}{\phi_{xx}(0)} = \frac{\sigma_z}{\sigma_x} \rho_{xz}(0) \quad (9.2)$$

The least error of the best attenuator is given by equation 5.5

$$\overline{e^2} = \sigma_z^2 - \sigma_y^2 \quad (5.5)$$

The results of equations 9.1 and 9.2 agree with that of equation 5.19 which gives the optimum attenuator as:

$$\begin{aligned} y(t) &= \sigma_z D_{11} \theta_1(x) + \overline{z} \\ &= \frac{\overline{(x(t) - \overline{x}) z(t)}}{\sigma_x} \left[\frac{x(t) - \overline{x}}{\sigma_x} \right] \\ &= \frac{\phi_{xz}(0)}{\phi_{xx}(0)} [x(t) - \overline{x}] + \overline{z} \end{aligned} \quad (9.3)$$

Since from equations 3.4 and 3.10

$$\theta_1(x) = \frac{x(t) - \bar{x}}{\sigma_x} \quad \text{and} \quad D_{11} = \frac{\overline{\theta_1[x(t)]z(t)}}{\sigma_z}$$

Booton²⁵ has shown that the correlation between the error $[z(t) - y(t)]$ and the input $x(t)$ is identically zero if the non-linear system has zero memory and the input is Gaussian. This result may be generalized for the class of input processes which are semi-separable negative.

$$\begin{aligned} \phi_{xe}(\tau) &= \overline{[x(t) - \bar{x}] \{z(t+\tau) - k[x(t+\tau) - \bar{x}(t+\tau)] - K\}} \\ &= \overline{[x(t) - \bar{x}] \{z(t+\tau) - kx(t+\tau)\}} \end{aligned} \quad (9.4)$$

$$= \phi_{xz}(\tau) - k\phi_{xx}(\tau) \quad (9.5)$$

For a stationary ergodic process which is semi-separable negative equation (9.4) becomes:

$$\begin{aligned} \phi_{zx}(\tau) &= a\phi_{xx}(\tau) \quad \tau \leq 0 \\ \text{where } a &= \frac{\phi_{zx}(0)}{\phi_{xx}(0)} = \frac{\phi_{xz}(0)}{\phi_{xx}(0)} \end{aligned} \quad (9.6)$$

Therefore $a = k$ which is Booton's equivalent gain and

$\phi_{xe}(\tau) \equiv 0$ when $\tau \geq 0$. Thus the linear attenuator has produced an error of approximation which is uncorrelated to

its input for $\tau \geq 0$ this performance is as good as that which ^{any} linear filter can be expected to ^{achieve}. It follows that if the input is separable then $\phi_{xe}(\tau) \equiv 0$ for all τ as shown by Nuttall.¹⁶

9.1.2 Best approximate linear filters

For inputs which are not semi-separable negative or for non-linear systems with memory a better approximation can probably be obtained with an approximate linear memory filter the input-output relationship for which is:

$$y(t) = \int_0^{\infty} w(\tau) \{x(t-\tau) - \bar{x}\} d\tau + K \quad (4.21)$$

The best approximate linear filter is characterized by the weighting function $w(\tau)$ and K which give minimum mean squared error.

$$\bar{e}^2 = \overline{[z(t) - y(t)]^2} = \overline{\left[z(t) - \int_0^{\infty} w(\tau) \{x(t-\tau) - \bar{x}\} d\tau + K \right]^2} \quad (9.7)$$

The optimum value for K , as before, is given by $K = \bar{z}(t)$. The optimum weighting function is given by equation 5.32:

$$\int_0^{\infty} w(\tau_2) \phi_{xx}(\tau_1 - \tau_2) d\tau_2 = \phi_{xz}(\tau_1) \quad (5.32)$$

$\tau_1 \geq 0$

This equation can usually be solved by the method of Appendix 13.3.1.*

By substituting $K = \overline{z(t)}$ and equation 5.32 in equation 9.7 it can be shown that the mean-squared error for the best approximate filter is again given by equation 5.5, this was proved generally in chapter 5.

If the non-linear system has zero memory and its input is semi separable negative then from equation 9.6

$$\phi_{zx}(\tau) = a \phi_{xx}(\tau) \quad \tau \leq 0$$

Therefore the solution to equation 5.32 is simply given by

$$\frac{w(\tau)}{\phi_x} = a \delta(\tau) \quad (9.8)$$

where $\delta(\tau)$ is the delta function.

Thus the best approximate-linear filter is a linear attenuator in this case.

9.1.3 Problem 4. The best approximate linear filter for a zero memory non-linear system

In this problem we shall consider a zero-memory non-linear system (see fig.19) defined by the equation:

---

* See also reference 27 for a clear description of the method of solution.

$$z(t) = \theta_3[x(t)] \quad (9.9)$$

where $\theta_3[x(t)] = -\frac{1}{2}\sqrt{21}x(t) + \frac{5}{18}\sqrt{21}x^3(t) \quad (9.10)$

The input $x(t)$ is the time image of the process considered in problem 3 of chapter 8 and $\theta_3[x]$ is the third order orthonormal polynomial for this input.

This particular problem is chosen because in this case Booton's equivalent gain is zero and therefore the equivalent attenuator is a poor approximate filter, however as demonstrated below the equivalent linear filter does produce a smaller mean squared error although it is still doubtful whether it is sufficiently small to be called a good approximation.

From equations 8.11 and 8.12 we see that:

$$\begin{aligned} \phi_{xz}(\tau) &= \frac{\sqrt{21}}{3} \psi \left\{ 2 - 6\psi + 6\psi^2 - 2\psi^3 \right\} \quad \tau > 0 \\ &= 0 \quad \tau \leq 0 \end{aligned}$$

whilst $\phi_{xx}(\tau) = \psi(2 - \psi)$

since $\sigma_z = \sigma_x = 1$ and $\left. \begin{aligned} \phi_{xz}(\tau) &= C_{13}(\tau) \\ \phi_{xx}(\tau) &= C_{11}(\tau) \end{aligned} \right\}$ after taking the time image.

where $\psi = e^{-|\tau|}$

∴ Booton's equivalent gain $k = \frac{\phi_{xz}(0)}{\phi_{xx}(0)} = 0$ (9.11)

whilst $K = \overline{z(t)} = 0$

Thus the mean squared error $\overline{e_a^2} = 1$

For the best linear filter approximation we have to solve the equation 5.32:

$$\int_0^{\infty} w(\tau_2) \phi_{xx}(\tau_1 - \tau_2) d\tau_2 = \phi_{xz}(\tau_1) \quad \tau_1 \geq 0$$

By following the method of solution in appendix 13.3 we get:

$$w(\tau) = \frac{\sqrt{21}}{3} \left[\delta(\tau) + 20e^{-3\tau} - 30e^{-4\tau} \right] \quad (9.12)$$

and the performance index.

$$\frac{\overline{e^2}}{\overline{e_a^2}} = 1 - \int_0^{\infty} w(\tau) \phi_{xz}(\tau) d\tau = 0.83 \quad (9.13)$$

9.1.4 Best approximate non-linear filters

Linearization of a non-linear system is useful particularly if there exists a feedback loop across the system (e.g. in Control Systems) because linear analysis methods can be used to determine the overall response,

whilst retaining the essential physical configuration of the loop. An example of the linearization of a feedback system is given in Section 9.3. The introduction of an approximate non-linear filter in place of a more complex non-linear system may be useful where linearization introduces errors of approximation which are too large and where feedback over the non-linearity is non-existent.

Given samples of the output of the non-linear system $z(t)$ and its input $x(t)$ or sufficient statistical information concerning the processes Z and X the best approximate multipath zero-memory filter can be found by the method outlined in Section 5.3.

The approximate filter of degree R is given by equation 5.19

$$f[x(t)] = \sum_{n=1}^R \sigma_z D_n \theta_n[x(t)] + \overline{z(t)} \quad (5.19)$$

By comparison with equations 9.1, 9.2 and 9.3 it is seen that the first two paths represent the best approximate attenuator.

In a similar way the best approximate multipath memory filter can be found by the method outlined in Section 5.4

9.2 Systems with two inputs

The preceding theory can be extended to enable approximate filters to be found when the input is the sum of two signals.

9.2.1 Best approximate attenuators

Somerville and Atherton²⁸ have described a method for finding the gain constants of two attenuators (one for each input) which give least mean squared error when used to replace a non-linear system with two uncorrelated inputs and a D.C. level. In the following theory we consider the case when the two signals are correlated and make allowance for a D.C. level on the output as well as the input.

Let $x(t)$ and $s(t)$ be two input signals with mean values $\bar{x}$ and $\bar{s}$ respectively.

We require to find the values of k_x , k_s and K in the equation

$$y(t) = k_x [x(t) - \bar{x}] + k_s [s(t) - \bar{s}] + K \quad (9.14)$$

so that

$$\overline{e^2} = \overline{[z(t) - y(t)]^2} \quad \text{is a minimum}$$

$$\overline{e^2} = \overline{z^2(t)} - 2k_x \phi_{xz}(0) - 2k_s \phi_{sz}(0) - 2K \overline{z(t)} \\ + k_x^2 \sigma_x^2 + 2k_x k_s \phi_{xs}(0) + k_s^2 \sigma_s^2 + K^2$$

$$\frac{\partial \overline{e^2}}{\partial K} = 2K - 2 \overline{z(t)} \quad (9.15)$$

$\therefore K = \overline{z(t)}$ as before.

$$\frac{\partial \overline{e^2}}{\partial k_x} = -2 \phi_{xz}(0) + 2k_x \sigma_x^2 + 2k_s \phi_{xs}(0)$$

$$\frac{\partial \overline{e^2}}{\partial k_s} = -2 \phi_{sz}(0) + 2k_s \sigma_s^2 + 2k_x \phi_{xs}(0)$$

Therefore k_x and k_s are given by the solution of the simultaneous equations:

$$k_x \sigma_x + k_s \sigma_s \rho_{xs}(0) = \sigma_z \rho_{xz}(0) \quad (9.16)$$

$$k_s \sigma_s + k_x \sigma_x \rho_{xs}(0) = \sigma_z \rho_{sz}(0) \quad (9.17)$$

By solving these equations we obtain:

$$k_x = \frac{\sigma_z}{\sigma_x} \frac{\rho_{xz}(0) - \rho_{sz}(0) \rho_{xs}(0)}{1 - \rho_{xs}^2(0)} \quad (9.18)$$

$$k_s = \frac{\sigma_z}{\sigma_s} \frac{\rho_{sz}(0) - \rho_{xz}(0) \rho_{xs}(0)}{1 - \rho_{xs}^2(0)} \quad (9.19)$$

When the inputs are uncorrelated we obtain the equations of Somerville and Atherton:

$$k_x = \frac{\sigma_z}{\sigma_x} \rho_{xz}(0) = \frac{\phi_{xz}(0)}{\sigma_x^2} \quad (9.20)$$

$$k_s = \frac{\sigma_z}{\sigma_s} \rho_{sz}(0) = \frac{\phi_{sz}(0)}{\sigma_s^2} \quad (9.21)$$

From equation 9.15, 9.16 and 9.17 we obtain

$$\overline{e^2} = \sigma_z^2 - k_x^2 \sigma_x^2 - k_s^2 \sigma_s^2 - 2 k_x k_s \sigma_x \sigma_s \rho_{xs}(0) \quad (9.22)$$

This equation for ^{the} mean square error could have been obtained from equation 5.5 to which it is identical.

If the non-linear system had been replaced by a single attenuator then:

$$\overline{e_a^2} = \sigma_z^2 - k^2 \sigma_x^2 \quad (9.23)$$

where

$$\begin{aligned} k &= \frac{\phi_{x+s,z}(0)}{\sigma_{x+s}^2} \\ &= \frac{\phi_{xz}(0) + \phi_{sz}(0)}{\sigma_x^2 + \sigma_s^2 - 2 \phi_{xs}(0)} \end{aligned}$$

Using equations 9.18 and 9.19 we obtain

$$k = \frac{k_x \sigma_x^2 + k_s \sigma_s^2 + \phi_{xs}(0)(k_x + k_s)}{\sigma_x^2 + \sigma_s^2 + 2\phi_{xs}(0)} \quad (9.24)$$

From equations 9.22, 9.23 and 9.24 it may be shown that:

$$\overline{e_a^2} - \overline{e^2} = \frac{[\sigma_x^2 \sigma_s^2 - \phi_{xs}^2(0)](k_x - k_s)^2}{\sigma_x^2 + \sigma_s^2 + 2\phi_{xs}(0)} \quad (9.25)$$

$$\therefore \overline{e_a^2} - \overline{e^2} \geq 0 \text{ since } \phi_{xs}^2(0) \leq \sigma_x^2 \sigma_s^2 \quad (9.26)$$

$$\text{and } \overline{e_a^2} = \overline{e^2} \text{ when either } k_x = k_s \quad (9.27)$$

$$\text{or } \sigma_x^2 \sigma_s^2 = \phi_{xs}^2(0)$$

The latter condition occurs if $x(t) \equiv s(t)$

From physical considerations it is seen that in both conditions these ~~cases~~ the two attenuators can be replaced by a single attenuator without any alteration to the output.

If x and s are uncorrelated then the mean squared error for a single attenuator is no greater than the mean squared error for two attenuators if and only if, $k_x = k_s$.

These results can be generalized for the case of n inputs.

9.2.2 Problem 5. The best approximate attenuators for the zero memory filters of problems 1 and 2*

The best approximate attenuators for each of the zero memory noise filters of problems 1 and 2 have been worked out and the results are recorded below. Since the mean squared error of the approximation is small in most of these cases the relative values of k_z and k_v give some indication of the filter performance.

A. Approximate attenuators for the filters of problem 1:

- i. $\sigma_z^2 = \sigma_v^2 = \frac{1}{2}$; $y = 0.7983x - 0.1643x^3 + 0.0118x^5$
 $k_z = 0.5466$ normalized mean squared error of
 $k_v = 0.4534$ the approximation

$$= 1 - \frac{k_z^2 \sigma_z^2 - k_v^2 \sigma_v^2}{\sigma_y^2} = \underline{0.0774}$$
- ii. $\sigma_z^2 = 0.8$; $\sigma_v^2 = 0.2$; $y = 1.5530x - 0.6411x^3 + 0.08759x^5$
 $k_z = 0.9042$ normalized mean squared error of
 $k_v = 0.3830$ the approximation

$$= \underline{0.0552}$$
- iii. $\sigma_z^2 = 0.9412$; $\sigma_v^2 = 0.0588$; $y = 1.8377x - 1.0716x^3 + 0.2378x^5$
 $k_z = 0.9948$ normalized mean squared error of
 $k_v = 0.0838$ the approximation

$$= \underline{0.0048}$$

---

* The notation is changed to conform with problems 1 and 2.

B. Approximate attenuators for the filter of problem 2

$$\sigma_z = \sigma_v = 1$$

$$y = -0.2246 + 0.4774x + 0.2478x^2 + 0.0182x^3 \\ - 0.0087x^4 + 0.0005x^5$$

$$k_z = 0.6379$$

$$k_v = 0.3620$$

normalized mean squared error of
the approximation

$$\underline{0.1566}$$

9.2.3 A linear memory filter and an attenuator

A dual input non-linear system can be replaced by the best approximate linear memory filter for one of its inputs and the best approximate attenuator for the other.

The combined filter output can be expressed by:

$$y(t) = \int_0^\infty w(\tau) [x(t-\tau) - \bar{x}] d\tau + k_s [s(t) - \bar{s}] + K \quad (9.28)$$

We require to find $w(\tau)$, k_s and K such that

$$\overline{e^2} = \overline{[z(t) - y(t)]^2} \quad \text{is a minimum}$$

$$\overline{e^2} = \overline{z^2(t)} - 2 \int_0^\infty w(\tau) \phi_{xz}(\tau) d\tau - 2k_s \phi_{sz}(0) - 2K \overline{z(t)} \\ + \int_0^\infty w(\tau_1) d\tau_1 \int_0^\infty w(\tau_2) d\tau_2 \phi_{xx}(\tau_1 - \tau_2) \\ + 2k_s \int_0^\infty w(\tau) \phi_{xs}(\tau) d\tau \\ + k_s^2 \phi_{ss}(0) + K^2 \quad (9.29)$$

$$\frac{\partial \overline{e^2}}{\partial k_s} = -2 \phi_{sz}(0) + 2 \int_0^\infty \omega(\tau) \phi_{xs}(\tau) d\tau + 2k_s \phi_{ss}(0)$$

$$\therefore k_s = \frac{1}{\sigma_s^2} \left[\phi_{sz}(0) - \int_0^\infty \omega(\tau) \phi_{xs}(\tau) d\tau \right] \quad (9.30)$$

and as usual $K = \overline{z(t)}$

Substitution in equation 9.29 yields:

$$\begin{aligned} \overline{e^2} &= \overline{z^2(t)} - 2 \int_0^\infty \omega(\tau) \phi_{xz}(\tau) d\tau - \sigma_s^2 k_s^2 - \overline{z(t)}^2 \\ &\quad + \int_0^\infty \omega(\tau_1) d\tau_1 \int_0^\infty \omega(\tau_2) d\tau_2 \phi_{xx}(\tau_1 - \tau_2) \\ \overline{e^2} &= \sigma_z^2 - \phi_{sz}^2(0) + 2 \int_0^\infty \omega(\tau) \left[\phi_{sz}(0) \phi_{xs}(\tau) - \phi_{xz}(\tau) \right] d\tau \\ &\quad + \int_0^\infty \omega(\tau_1) d\tau_1 \int_0^\infty \omega(\tau_2) d\tau_2 \left[\phi_{xx}(\tau_1 - \tau_2) - \phi_{xs}(\tau_1) \phi_{xs}(\tau_2) \right] \quad (9.31) \end{aligned}$$

Minimization of this equation, using the variational calculus, yields an equation of the Wiener Hopf type.

$$\begin{aligned} &\int_0^\infty \omega(\tau_2) \left\{ \phi_{xx}(\tau_1 - \tau_2) - \phi_{xs}(\tau_1) \phi_{sx}(\tau_2) \right\} d\tau_2 \\ &= \phi_{xz}(\tau_1) - \phi_{sz}(0) \phi_{xs}(\tau_1) \quad (9.32) \\ &\tau_1 \geq 0 \end{aligned}$$

This equation can usually be solved by the method of appendix 13.3.

If x and s are uncorrelated then

$$k_s = \frac{\phi_{sz}(0)}{\sigma_s^2} \quad (9.33)$$

and $w(\tau)$ is given by the solution of

$$\int_0^\infty w(\tau_2) \phi_{xx}(\tau_1 - \tau_2) d\tau_2 = \phi_{xz}(\tau_1) \quad \tau_1 \geq 0$$

9.2.4 Two linear filters

We may choose to replace a dual input non-linear system by two linear memory filters one for each input. The combined output of the two filters is given by:

$$y(t) = \int_0^\infty w_x(\tau) [x(t-\tau) - \bar{x}] d\tau + \int_0^\infty w_s(\tau) [s(t-\tau) - \bar{s}] d\tau + K \quad (9.34)$$

The optimum function $w_x(\tau)$ and $w_s(\tau)$ to give the best approximation can be found by minimizing the mean squared error. This is the same problem as the one dealt with in Section 5.4 where the optimum weighting functions $w_m(\tau)$ had to be found for the multipath filter. The solutions for the optimum function $w_x(\tau)$ and $w_s(\tau)$ are therefore given by the two simultaneous integral equations below:

$$\begin{aligned}
& \int_0^{\infty} \psi_x(\tau_2) \phi_{xx}(\tau_1 - \tau_2) d\tau_2 + \int_0^{\infty} \psi_s(\tau_2) \phi_{xs}(\tau_1 - \tau_2) d\tau_2 \\
& \quad = \phi_{xz}(\tau_1) \quad \tau_1 \geq 0 \\
& \int_0^{\infty} \psi_x(\tau_2) \phi_{sx}(\tau_1 - \tau_2) d\tau_2 + \int_0^{\infty} \psi_s(\tau_2) \phi_{ss}(\tau_1 - \tau_2) d\tau_2 \quad (9.35) \\
& \quad = \phi_{sz}(\tau_1) \quad \tau_1 \geq 0
\end{aligned}$$

These equations can be solved by the method of undetermined coefficients as described in appendix 13.3.

If x and s are uncorrelated the equations 9.35 become two independent equations of the Wiener Hopf type.

For the general case with n inputs there will be n simultaneous integral equations.

9.2.5 A zero memory filter and an attenuator

A dual input non-linear system can be replaced by the best approximate zero memory filter for one of its inputs and the best attenuator for its other input.

The combined filter output can be expressed by

$$y(t) = \sum_{n=1}^R k_{x_n} \theta_n [x(t)] + k_s [s(t) - \bar{s}] + K \quad (9.36)$$

The mean square error is given by:

$$\overline{e^2} = \overline{\left[z(t) - \sum_{n=1}^R k_{x_n} \theta_n [x(t)] - k_s [s(t) - \bar{s}] + K \right]^2}$$

$$\begin{aligned}
 \therefore \overline{e^2} &= \overline{z^2(t)} - 2 \sum_{n=1}^R k_{x_n} \overline{\theta_n[x(t)]z(t)} - 2k_s \overline{[s(t)-\bar{s}]z(t)} \\
 &\quad - 2K \overline{z(t)} + \sum_{n=1}^R k_{x_n}^2 + 2 \sum_{n=1}^R k_{x_n} k_s \overline{\theta_n[x(t)]s(t)} \\
 &\quad + k_s^2 \sigma_s^2 + K^2
 \end{aligned}
 \tag{9.37}$$

(by using the orthonormal property of $\theta_n[x]$)

Differentiating partially with respect to K and equating to zero gives the usual result that $K = \overline{z(t)}$

$$\frac{\partial \overline{e^2}}{\partial k_s} = -2\phi_{sz}(0) + 2 \sum_{n=1}^R k_{x_n} \phi_{\theta_n s}(0) + 2k_s \sigma_s^2 = 0
 \tag{9.38}$$

$$\frac{\partial \overline{e^2}}{\partial k_{x_n}} = -2\phi_{\theta_n z}(0) + 2k_{x_n} + 2k_s \phi_{\theta_n s}(0) = 0
 \tag{9.39}$$

Solving these equations gives

$$k_s = \frac{\phi_{sz}(0) - \phi_{\theta_n z}(0) \phi_{\theta_n s}(0)}{\sigma_s^2 - \phi_{\theta_n s}^2(0)}
 \tag{9.40}$$

$$k_{x_n} = \frac{\sigma_s^2 \phi_{\theta_n z}(0) - \phi_{sz}(0) \phi_{\theta_n s}(0)}{\sigma_s^2 - \phi_{\theta_n s}^2(0)}
 \tag{9.41}$$

Equations 9.18 and 9.19 can be obtained from these equations.

9.3 Application to feedback system analysis

A non-linear feedback system (fig.20) can be analysed by the methods described in Section 9.1 and 9.2.

The insertion of an approximate filter for the non-linear box is chosen rather than the alternative approach of replacing the whole feedback system by an approximate filter. This is because it is desirable to retain the essential configuration of the loop so that the analysis does not become divorced too much from physical reality.

To permit linear analysis methods to be used it is necessary to linearize the loop; this can be done by replacing the non-linear box by the best approximate linear filter or by replacing it by two separate filters only one of which need be linear.

There exists a fundamental difficulty in applying the theory of Sections 9.1 and 9.2 to feedback systems; it is concerned with the fact that the introduction of an approximate structure within the loop disturbs one of the inputs and thereby invalidates the theory. Changes in the amplitude of the feedback signal can be allowed for, this is done in Booton's method and in the describing function method for deterministic inputs, but the design of a more complex approximate structure requires greater input information so that changes in other statistical properties

of the input should also be allowed for. If the approximate structure gives an output which is sufficiently close to the desired output then, of course, the feedback signal will not be appreciably changed by the insertion of the approximate structure and the approximation will not be degraded significantly because of this effect. A practical example of this situation is given later, but circumstances can arise where the approximation is poor and nothing can be done to improve it.

Let the non-linear box consist of a zero memory non-linearity only (fig.20) so that:

$$z(t) = f[e(t)] \quad \text{where} \quad e(t) = x(t) - s(t) \quad (9.42)$$

$$\text{then} \quad s(t) = \int_0^{\infty} w_m(\tau) f[e(t-\tau)] d\tau \quad (9.43)$$

Substituting in equation (9.42):

$$z(t) = f \left[x(t) - \int_0^{\infty} w_m(\tau) f[e(t-\tau)] d\tau \right]$$

$$\text{and} \quad s(t) = \int_0^{\infty} \int_0^{\infty} w_m(\tau_1) f \left[x(t-\tau_1) - w_m(\tau_2) f \{ e(t-\tau_1-\tau_2) \} \right] d\tau_1 d\tau_2$$

Successive substitution leads to the result:

$$s(t) = \underbrace{\int_0^\infty \int_0^\infty \dots \int_0^\infty}_{\infty} \omega_m(\tau_1) f \left[x(t-\tau_1) - \omega_m(\tau_2) f \left[x(t-\tau_1-\tau_2) - \omega_m(\tau_3) f \left[x(t-\tau_1-\tau_2-\tau_3) - \dots \right] \right] \right] d\tau_1 d\tau_2 \dots d\tau_\infty \quad (9.44)$$

Therefore, in general, the overall system is of Zadeh's class η_∞ . Since we are ^{proposing to} replace the non-linear box by a linear filter in the loop and at the most a non-linear filter of class η_1 in the input, the approximate overall system is only of class η_1 ; it is to be expected that situations will arise where the approximation may be very severe. Thus whilst the application of these approximate methods to feedback systems may be important because no other methods of analysis are available for non-linear feedback systems, it is not expected that they will always be satisfactory; for this reason, at least, methods of optimizing higher order filters is worth investigating (see chapter 11).

9.3.1 Problem 7. Linearization of a non-linear limiter

The limiter system of fig.21 which is almost identical to the system which Booton used as an example to demonstrate his method of linearization, was simulated on the Precision Analogue Computing Equipment (PACE) in Brussels. The simulated system is shown in fig.21.* The limiter was

---

* For a key to the principal symbols used in these diagrams see fig.23.

simulated by a biased diode feedback circuit which was adjusted to limit the swing of the output of amplifier B to $33 \pm 1/3$ volts. The input was chosen to have a variance of $33 \pm 1/3$ volts so that the range of the computer (0-100 v) was used fully without saturation occurring for more than 0.14% of the total time when using a Gaussian input (see Appendix 13.1). Two different values of the gain K of the limiter amplifier B and the time constant T of the integrator C were used. The following measurements were made (using the circuit of fig.22) by integrating various cross products for a chosen time interval. This time interval was chosen, rather arbitrarily, to be long enough to give a calculated variance in the mean squared value $\phi_{xx}(0)$ of less than 3% (see appendix 13.1.3); the experimental results agree approximately with this calculation.

	Limiter Gain K	Integrator T Time Constant	$\phi_{xx}(0)$	$\phi_{zz}(0)$
Case (i)	1	$\frac{1}{2}$ sec.	0.602	343.2
Case (ii)	10	$\frac{1}{20}$ sec.	0.602	654.6
			Integrated volts over 1 min.	

Although the results for $\phi_{s\theta_n}$ and $\phi_{z\theta_n}$ when n is even were measured, they are not included in the tables below, since by symmetry it can be seen that they must be zero.

	$\phi_{ss}(0)$	$\frac{100}{3} \phi_{s\theta_1}(0)$	$12.073 \phi_{s\theta_3}(0)$	$10.360 \phi_{s\theta_5}(0)$
Case (1)	23.513	-0.666	0.094	-0.018
Case (11)	266.560	-7.602	0.801	-0.269
	$\phi_{zs}(0)$	$\frac{100}{3} \phi_{z\theta_1}(0)$	$12.073 \phi_{z\theta_3}(0)$	$10.360 \phi_{z\theta_5}(0)$
Case (1)	-0.040	13.447	-3.867	-1.136
Case (11)	0.300	13.057	-2.907	-0.619

Integrated volts over 1 minute

	Case (1)			Case (11)		
	R = 1	R = 3	R = 5	R = 1	R = 3	R = 5
k_y	0.653	0.680	0.679	0.974	1.003	1.002
k_{x_1}	0.694	0.695	0.695	1.023	1.034	1.034
k_{x_3}		-0.543	-0.543		-0.512	-0.512
k_{x_5}			0.181			0.056

The filter functions (filter A in fig.20) were worked out using the equation $y = k_{x_1} \theta_1(x) + k_{x_3} \theta_3(x) + k_{x_5} \theta_5(x)$ and plotted in fig.24. In both cases they are seen to resemble a saturating element.

The approximate systems for case (1) and case (11) were simulated on P.A.C.E. as shown in fig.25; in figs.26, 27, 28 and 29 are shown multichannel pen recordings of the signals at various points in the system for both cases (1) and (11) when $R = 1$ and $R = 5$. The normalized mean-squared errors

which were calculated (assuming that the approximate filter causes no disturbance to its input) and measured experimentally, are shown in the table below:

	R	Calculated normalized m.s.e.	Measured normalized m.s.e.
Case (1)	1	0.09363	
	3	0.01864	
	5	0.01251	0.06496
Case(11)	1	0.3194	
	3	0.2846	
	5	0.2843	0.3094

Values of error at the system output were too small to be measureable.

Conclusions: It is observed that in both cases (1) and (11) when $R=1$, $\sigma_s^2 k_s \approx k_{x_1}$. This interesting result could be due to the fact that any 'follow up' system should produce a response $s(t)$ which is very nearly identical to its input $x(t)$. Thus from equations 9.40 and 9.41 we see that $\sigma_s^2 k_s \approx k_{x_1}$. The greatest divergence between k_{x_1} and $k_s \sigma_s^2$ occurs in case (1) which, due to the low gain and large time constant, did not represent such an accurate 'follow up' system as case (11).

From equations 9.35 it is seen that if $x(t) \approx s(t)$ then $\mathcal{W}_x(\tau) \approx \mathcal{W}_s(\tau)$; it can be concluded that if the filters A and B of fig. 20 belong to the same linear class then for a 'follow up' system little improvement is achieved by using a separate filter for each input.

Since for the case $R=1$ we can replace the attenuators A and B of fig. 20 by a single attenuator the approximate system is almost identical to Booton's. It is unfair to call it Booton's equivalent system because no allowance has been made to correct for changes in the variance of $s(t)$ but since Booton suggested doing this by a doubtful* analytical procedure, the results obtained may be no worse than Booton's. In case (i) a significantly better approximation is obtained when $R=5$ than when $R=1$. In case (ii) the difference between the results obtained when $R=5$ and $R=1$ are not very significant. In this latter case neither approximation seems to work very well.

---

* He assumes that the error signal $e(t) = x(t) - s(t)$ is Gaussian, this assumption seems very doubtful especially for a system response that follows the input fairly accurately in which case the error signal is small and could be significantly non-Gaussian although both $x(t)$ and $s(t)$ may be very nearly Gaussian themselves.

10.0 THE NON-LINEAR COMPENSATION OF A FEEDBACK CONTROL SYSTEM.

It has been demonstrated in earlier chapters that a non-linear system may be capable of giving a much better performance than the optimum linear system. Applications to noise filtering and prediction are almost obvious but other applications, particularly to the optimization of control systems, are severely complicated by practical considerations.

Sometimes some of the physical components of a system are chosen as a result of considerations other than their performance. This situation arises most commonly in control systems where often the motor and load are predetermined by considerations of the operation to be performed, the available power, convenience etc; also it is quite common for the essential control system configuration to be decided out of other practical considerations; for instance, feedback loops are commonly employed so as to minimize drift due to component changes and other causes. These practical considerations complicate the system optimization and it becomes necessary to compensate the system performance by the insertion of a linear or non-linear network. The use of linear networks for this purpose is very common, but non-linear compensation, though more difficult, possesses the advantage, mentioned above, that better performance may be obtained.

The response of a non-linear or linear system may be

compensated by the parallel connection of a multipath filter which has been optimized on a desired output of $z(t) = [s(t) - u(t)]$; where $u(t)$ is the output of the non-linear system and $s(t)$ is its desired output. This parallel compensation of an existing system may be satisfactory for applications in electronics such as encoders or decoders but it may be impossible in other applications, such as control systems, which are required to give a power output. A method of non-linear compensation of a feedback control system is considered in this chapter.

10.1 The theory.

The method proposed herein does not attempt to tackle the problem of the optimization of a feedback loop containing a non-linearity. This is too difficult at the moment, since, as demonstrated in chapter 9, such systems are, in general, of class γ_{∞} . Instead, the problem of improving the performance of an otherwise optimum linear feedback system is considered, not by the insertion of a non-linearity within the loop, but by connecting a feedforward loop over part of the linear loop. Fig.30 shows a linear feedback system. $F(\rho)$ is the motor and load transfer function, $H(\rho)$ is the linear compensating network transfer function. Fig.31 shows the same system with the feedforward compensating network C connected.

The following practical restrictions on the design are normally encountered:

1. The motor and load are specified, therefore $F(\rho)$ is invariant.
2. The motor and load, being non-linear, saturate at some amplitude; to permit the assumption that they are linear a restriction on the motor input mean power σ^2 is imposed so that saturation occurs on the average for a certain small percentage of the time.
3. The essential feedback configuration is decided by practical considerations.

Newtons^{29,30} modifications of Wiener's^{12,31,32} optimization procedure enables the optimum $H(\rho)$ (fig.30) to be found for any specified $F(\rho)$, σ^2 and input spectral density $\Phi_{xx}(\omega)$.

Clearly for the same power limitation* there is no object in adding a linear feedforward network for C , since no greater improvement in performance can be achieved than by the correct choice of $H(\rho)$. However although $H(\rho)$ has been chosen to give the best linear response the insertion of a non linear network C can improve the performance still further. Therefore C can be a multipath

---

* For the purpose of this chapter the possible presence of noise at the input is ignored although the theory needs no extension to include this case.

memory filter and we assume that $H(p)$ has been chosen to give optimum linear performance.* Since C contains a linear path, for mathematical convenience, though not in practice, we shall suppress the x input to $H(p)$ giving the system of fig.32.

The error unilateral Laplace transform is then given by:

$$E(p) = S(p) - U(p) = S(p) - \frac{F(p)}{1 + H(p)F(p)} Y(p) \quad (10.1)$$

where $S(p)$ is the transform of the desired system output $s(t)$ which may be equal to $x(t)$ or a component part of $x(t)$ (if the input contains superimposed noise). $E(p)$ is the transform of ^{the error} $e(t)$ and $Y(p)$ is the transform of the filter output $y(t)$.

$$\text{But } y(t) = \sum_{n=1}^R \int_0^{\infty} \theta_n [x(t-\tau)] w_n(\tau) d\tau$$

$$\therefore Y(p) = \sum_{n=1}^R \theta_n(p) W_n(p) \quad (10.2)$$

where $\theta_n(p)$ is the transform with respect to t of $\theta_n [x(t)]$

Substituting equation 10.2 in 10.1 gives:

$$E(p) = S(p) - U(p) = S(p) - \frac{F(p)}{1 + H(p)F(p)} \sum_{n=1}^R \theta_n(p) W_n(p) \quad (10.3)$$

* As explained later, the choice of $H(p)$ is subject only to practical considerations, it need not be chosen to give optimum linear performance.

The transform of the motor input is given by:

$$V(p) = \frac{1}{1 + H(p)F(p)} \sum_{n=1}^R \Theta_n(p) W_n(p) \quad (10.4)$$

At this stage one of the following three substitutions can be used. The particular choice of substitution will depend on the problem, as explained later. The remaining algebra is worked out in terms of general symbols.*

Case (i) The first substitution:

$$\begin{aligned} \mathcal{U}_n(p) &= \frac{W_n(p)}{1 + H(p)F(p)} \\ \gamma_n(p) &= F(p)\Theta_n(p) \\ \mathcal{A}_n(p) &= \Theta_n(p) \end{aligned} \quad (10.5)$$

Case (ii) The second substitution:

$$\begin{aligned} \mathcal{U}_n(p) &= \frac{F(p)W_n(p)}{1 + H(p)F(p)} \\ \gamma_n(p) &= \Theta_n(p) \\ \mathcal{A}_n(p) &= \Theta_n(p)F^{-1}(p) \end{aligned} \quad (10.6)$$

Case (iii) The third substitution

$$\begin{aligned} \mathcal{U}_n(p) &= W_n(p) \\ \gamma_n(p) &= \frac{F(p)\Theta_n(p)}{1 + H(p)F(p)} \\ \mathcal{A}_n(p) &= \frac{\Theta_n(p)}{1 + H(p)F(p)} \end{aligned} \quad (10.7)$$

---

*The symbols used are taken from the Russian alphabet:
 $\mathcal{U}$ pronounced sh; γ pronounced ch and $\mathcal{A}$ pronounced ya.

By using one of these substitutions in equation 10.3 and taking the inverse transform we get:

$$e(t) = s(t) - \sum_{n=1}^R \int_0^{\infty} y_n(t-\tau) u_n(\tau) d\tau \quad (10.8)$$

where $y_n(t)$ is the inverse transform of $Y_n(p)$

and $u_n(t)$ " " " " $U_n(p)$

Also taking the inverse transform of equation 10.4 gives:

$$v(t) = \sum_{n=1}^R \int_0^{\infty} \mathcal{A}_n(t-\tau) u_n(\tau) d\tau \quad (10.9)$$

where $\mathcal{A}_n(t)$ is the inverse transform of $\mathcal{A}_n(p)$

$$\begin{aligned} \therefore \overline{e^2} &= \phi_{ss}(0) - 2 \sum_{n=1}^R \int_0^{\infty} u_n(\tau) \phi_{y_n s}(\tau) d\tau \\ &+ \sum_{n=1}^R \sum_{m=1}^R \int_0^{\infty} \int_0^{\infty} u_n(\tau_1) u_m(\tau_2) \phi_{y_n y_m}(\tau_1 - \tau_2) d\tau_1 d\tau_2 \quad (10.10) \end{aligned}$$

$$\text{also } \overline{v^2} = \sigma^2 = \sum_{n=1}^R \sum_{m=1}^R \int_0^{\infty} \int_0^{\infty} u_n(\tau_1) u_m(\tau_2) \phi_{\mathcal{A}_n \mathcal{A}_m}(\tau_1 - \tau_2) d\tau_1 d\tau_2 \quad (10.11)$$

$$\begin{aligned} \therefore \overline{e^2} + \lambda \sigma^2 &= \phi_{ss}(0) - 2 \sum_{n=1}^R \int_0^{\infty} u_n(\tau) \phi_{y_n s}(\tau) d\tau \\ &+ \sum_{n=1}^R \sum_{m=1}^R \int_0^{\infty} \int_0^{\infty} u_n(\tau_1) u_m(\tau_2) \left[\phi_{y_n y_m}(\tau_1 - \tau_2) + \lambda \phi_{\mathcal{A}_n \mathcal{A}_m}(\tau_1 - \tau_2) \right] d\tau_1 d\tau_2 \quad (10.12) \end{aligned}$$

Minimization of this equation in terms of the Lagrangian multiplier λ yields an implicit solution for $u_n(\tau)$ in terms of λ . By a trial and error procedure, a value of λ can be found, giving a set of solutions for $u_n(\tau)$ which, when substituted in equation 10.11 yields the required σ^2 .

Let Δ be the variation produced in $\bar{e}^2 + \lambda \sigma^2$ when $u_n(\tau)$ is given a variation of $\epsilon \gamma_n(\tau)$ where $\gamma_n(\tau)$ is a physically realizable linear weighting function and ϵ a small number.

Then

$$\begin{aligned} \Delta = & -2 \sum_{n=1}^R \epsilon \int_0^{\infty} \gamma_n(\tau) \phi_{y_n s}(\tau) d\tau \\ & + 2 \sum_{n=1}^R \sum_{m=1}^R \epsilon \iint_0^{\infty} \gamma_n(\tau_1) u_m(\tau_2) \left[\phi_{y_n y_m}(\tau_1 - \tau_2) \right. \\ & \quad \left. + \lambda \phi_{g_n g_m}(\tau_1 - \tau_2) \right] d\tau_1 d\tau_2 \\ & + \sum_{n=1}^R \sum_{m=1}^R \epsilon^2 \iint_0^{\infty} \gamma_n(\tau_1) \gamma_m(\tau_2) \left[\phi_{y_n y_m}(\tau_1 - \tau_2) + \lambda \phi_{g_n g_m}(\tau_1 - \tau_2) \right] \\ & \quad d\tau_1 d\tau_2 \end{aligned}$$

Equating $\frac{\partial \Delta}{\partial \epsilon} = 0$ when $\epsilon = 0$ gives

$$\begin{aligned} \sum_{n=1}^R \gamma_n(\tau_1) \left\{ \sum_{m=1}^R \int_0^{\infty} u_m(\tau_2) \left[\phi_{y_n y_m}(\tau_1 - \tau_2) + \lambda \phi_{g_n g_m}(\tau_1 - \tau_2) \right] d\tau_2 \right. \\ \left. - \phi_{y_n s}(\tau_1) \right\} = 0 \end{aligned} \quad (10.13)$$

Using the fundamental lemma of variational calculus¹¹ we get:

$$\sum_{m=1}^R \int_0^{\infty} u_m(\tau_2) \left[\phi_{y_n y_m}(\tau_1 - \tau_2) + \lambda \phi_{g_n g_m}(\tau_1 - \tau_2) \right] d\tau_2 \quad (10.14)$$

$$= \phi_{y_n s}(\tau_1) \quad \tau_1 \geq 0$$

$$n = 1, 2, \dots, R$$

The equations, which give the optimum $u_m(\tau)$ are of the same form as equation (5.27) and they can be solved by a similar method (see appendix 13.3).

Since $H(\rho)$ is unspecified by the solution of equations 10.14 it can be chosen arbitrarily provided the following practical points are attended to:

1. An unwise choice of $H(\rho)$ may cause some or all of $W_m(\rho)$ to be ^{un}non-physically/realizable.
2. A choice of $H(\rho) = 0$ is often impractical since some feedback is required to minimize drift. For this reason it may be desirable also to reconnect the input to $H(\rho)$ as in fig.33 and modify accordingly the linear path optimum transfer function $W'_1(\rho)$ in the compensating network C' as below:

$$W_1(\rho) = W'_1(\rho) - \sigma_x H(\rho) \quad (10.15)$$

since $\theta_1(x) = \frac{x}{\sigma_x}$, (assuming $\bar{x} = 0$)

3. It may be possible by a sensible choice of $H(\rho)$ not only to satisfy the practical restrictions 1 and 2 above but also to save on the number of components in the synthesis of all the $W_m(\rho)$ transfer functions, this situation arises if the function $U_m(\rho)$ (which are transforms of the optimum $u_m(\tau)$ given by equation 10.14) contain some common factors.

All of the above practical considerations concerning the choice of $H(\rho)$ require that it shall be chosen after the optimization has been done and the optimum $U_m(\rho)$ functions are known. This is easy for cases (i) and (ii) but the use of the substitutions of case (iii) necessitates a prior choice of $H(\rho)$ so that the correlation coefficients $\phi_{y_n y_m}(\tau)$, $\phi_{y_n y_m}(\tau)$ and $\phi_{y_n s}(\tau)$ can be determined, however, this choice can be modified after the optimization, provided all the optimum weighting functions $W_i(\tau)$ are modified to correspond.

The correlation coefficients needed in the optimization can be obtained either experimentally by analogue measurements on long samples of the input $x(t)$ and desired output $s(t)$ or analytically from a knowledge of the statistics of $x(t)$ and $s(t)$. Fig. 34 shows the block diagram of the measuring equipment needed for the experimental determination of the correlation coefficients for case (i); it is seen from this diagram that it is necessary to simulate not only the polynomial functions $\theta_n(x)$ but also the motor and load.

For the analytical determination of these correlation coefficients the following equations may be used:

Case (i)

$$\phi_{y_n y_m}(\tau) = \int_0^\infty \int_0^\infty f(t_1) f(t_2) \phi_{x_n x_m}(\tau + t_1 - t_2) dt_1 dt_2$$

and

$$\phi_{x_n x_m}(\tau) = C_{nm}(\tau)$$

where $f(t)$ is the inverse transform of $F(p)$

$$\text{also } \phi_{y_n s}(\tau) = \int_0^\infty f(t) \phi_{x_n s}(\tau + t) dt$$

$$\text{where } \phi_{x_n s}(\tau) = \delta_s D_{n1}(\tau)$$

(10.16)

Case (ii)

$$\phi_{x_n x_m}(\tau) = \int_0^\infty \int_0^\infty f^{-1}(t_1) f^{-1}(t_2) \phi_{y_n y_m}(\tau + t_1 - t_2) dt_1 dt_2$$

and

$$\phi_{y_n y_m}(\tau) = C_{nm}(\tau)$$

(10.17)

where

$$f^{-1}(t) \text{ is the inverse transform of } \frac{1}{F(p)}$$

also

$$\phi_{y_n s}(\tau) = \delta_s D_{n1}(\tau)$$

Case (iii)

$$\phi_{y_n y_m}(\tau) = \int_0^\infty \int_0^\infty f(t_1) f(t_2) \phi_{y_n y_m}(\tau + t_1 - t_2) dt_1 dt_2$$

and

$$\phi_{y_n y_m}(\tau) = \int_0^\infty \int_0^\infty k(t_1) k(t_2) C_{nm}(\tau + t_1 - t_2) dt_1 dt_2$$

(10.18)

where $f(t)$ is the inverse transform of $F(p)$

and $k(t)$ " " " " $K(p) = \frac{1}{1 + H(p) F(p)}$

also

$$\phi_{y_n s}(\tau) = \int_0^\infty g(t) \sigma_s D_{n1}(\tau - t) dt$$

where $g(t)$ is the inverse transform of $\frac{F(p)}{1 + H(p) F(p)} = G(p)$

It is seen from the above equations that the substitutions of case (iii) leads to considerably more work in the analytical computation of the correlation coefficients than in the other cases. Case (ii) is the best ^{set of} substitutions in this respect. Another factor which may influence the choice of substitution is the non existence of some of the correlation coefficients, e.g. when the motor and load transfer function $F(p)$ contains a factor of the form $\frac{1}{p}$ representing integration, then its output auto-correlation coefficient usually does not exist since its output is a

divergent non-stationary process for a bounded non-zero input; therefore all the other correlation coefficients $\phi_{y_n y_m}(\tau)$ of case (i) may not exist. The correlation coefficients $\phi_{\dot{y}_n \dot{y}_m}(\tau)$ of case (ii) may not exist either because the derivatives of the signal $\dot{y}_n [x(t)]$ may not exist. For these reasons the substitutions of case (iii) may have to be used.

10.2 Problem 7.

THE UNITED STATES OF AMERICA

11. LIMITATIONS AND EXTENSIONS OF THE THEORY.

In this thesis some attempt has been made to simplify non-linear filter optimization so that it can be performed by an almost routine procedure, which is generally applicable to engineering problems. This has been done by considering only a relatively simple class of filters instead of the more general and complex filters considered by others^{6,8}. A number of questions, some of which remain unanswered, have arisen as a result of this study and are considered in this chapter.

11.1 Available information.

It has been implicitly assumed hitherto that the information needed for the optimization of a non-linear filter of class γ_i is readily available, either in the form of statistical data or as long samples of the input and desired output. It is not difficult to imagine practical situations where this is not so, particularly in the application to noise filtering, where only limited information about the desired output may be available. The best that can be done in such circumstances is to choose a class of filters to be optimized, which use just that information which is available and no more. For example, if only spectral information is available, then the required class is the linear class. An approach of this kind demands first that a lucid classification of non-

linear filters be established, based upon information needed in the optimization; so that the appropriate class of filters for the problem can be chosen a priori, depending on the available information. An elucidate classification is not available at the present time.

11.2 Non-stationary signals.

A second assumption which has been made is that the input and desired output of the filter are both stationary and ergodic. Unfortunately, this assumption is more serious than the former, since a very high proportion of signals in practice are non-stationary.

The theory of chapters 3 and 6 is presented in a general manner for non-stationary signals and the optimization procedures of chapter 5 could be generalized for time-dependent filters with non-stationary inputs. However, there remains the almost insurmountable practical difficulty of how to obtain sufficient data concerning the input and desired output when only a single sample of a non-stationary ensemble is available. Special cases could be tackled, such as that considered by Beutler³³, but since few signals occurring in practical situations have been adequately analyzed, it is not yet possible to decide which special cases are of practical importance; some classification of processes based upon their non-stationary properties seems to be needed.

11.3 Other performance criteria.

The third basic assumption, which has been made, is that the mean-squared error criterion is acceptable for the application in question. It is not difficult to visualize applications where the mean-squared error criterion is not as appropriate as some other criterion. Zadeh³⁴ has pointed out that there are two approaches to the optimization problem. In one, a class of filters is chosen and an optimum filter within this class is found by minimizing the mean-squared error. In the other, an appropriate statistical criterion is chosen and the optimum filter is found for this criterion without restriction as to its class. This latter approach is more difficult but some problems have been tackled using other criteria.³⁵

11.4 Classification of processes.

The optimum non-linear filter of some class, under certain conditions of input, may belong to a subclass, which is more easily optimized and synthesized. This situation can often be foreseen, thereby simplifying the optimization, if a logical classification of processes and their properties is available. For instance, referring to the theorem in section 6.4, we see that a linear filter is as good as a multipath non-linear filter, if the input is separable, both with respect to itself and with

respect to the desired output; separability is easily tested using the invariance property so that the need to measure the expansion coefficients can be obviated in this case.

In this thesis, some extension of the classification of processes has been made by the introduction of the semi-separable class and its properties which have some importance in non-linear filtering, but there seems to be a need for further extension of the still limited classification of processes which is available.

11.5 Generalizations.

The problem of the analysis of non-linear feedback systems was discussed in chapter 9, in which it was shown that, in general, a non-linear feedback system cannot be exactly simulated except by a filter of Zadeh's class η_{∞} . Thus exact analysis and the optimization of such systems cannot be performed by the methods proposed herein. Also, outside the scope of this thesis comes the exact analysis and optimization of multi-input non-linear systems. However, some generalizations are proposed here which may lead to the development of methods, which are capable of dealing with these problems.

11.6 Generalization of the orthonormal polynomials.

So that the optimization procedures for multipath single input filters can be extended to multi-input filters, it is

necessary to define a set of generalized orthonormal polynomials.

For the set of stationary signals* $x_1, x_2, \dots, x_n$, let $\phi_{p,q,\dots,w}(x_1, x_2, \dots, x_n)$ represent a set of orthonormal polynomials with respect to the joint instantaneous probability function $p_n(x_1, x_2, \dots, x_n)$ over the whole range of $x_1, x_2, \dots, x_n$ i.e.

$$\underbrace{\int \int \dots \int}_n \phi_{p,q,\dots,w}(x_1, x_2, \dots, x_n) \phi_{r,s,\dots,y}(x_1, x_2, \dots, x_n) \times p_n(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

$$= 1 \quad \begin{array}{l} p = r \\ q = s \\ \dots \\ w = y \end{array} \quad (11.1)$$

$$= 0 \quad \text{otherwise.}$$

On the assumption that these polynomials exist, we can derive a recurrence equation for them, which is a generalization of equation (3.6).

---

* Only stationary signals are considered here for simplicity.

Let

$$\begin{aligned} & \phi_{p,q,\dots,\omega}(x_1, x_2, \dots, x_n) \\ &= A_{p,q,\dots,\omega} x_1^p x_2^q \dots x_n^\omega - \sum_{a=0}^p \sum_{b=0}^q \dots \sum_{k=0}^\omega B_{a,b,\dots,k} \phi_{a,b,\dots,k}(x_1, x_2, \dots, x_n) \\ & \quad \left(\begin{array}{c} \text{excluding} \\ a=p; b=q; \dots k=\omega \end{array} \right) \end{aligned} \quad (11.2)$$

where $\phi_{a,b,\dots,k}$ is of lower order than $\phi_{p,q,\dots,\omega}$ (i.e. at least one of its suffixes is of lower order and none are greater).

For simplicity, we write equation (11.2) using a vector notation as follows:

$$\Phi_\alpha[x] = A_\alpha [x]^\alpha - \sum_{\beta=0}^{\alpha-1} B_\beta \Phi_\beta[x] \quad (11.3)$$

and the orthonormal condition is written:

$$\int \Phi_\alpha[x] \Phi_\beta[x] P[x] dx = 1 \quad \alpha=\beta \quad (11.4)$$

$$= 0$$

otherwise

Multiplying (11.3) by $\Phi_\beta[x]$ and $\Phi_\alpha[x]$ and using the orthonormal property, gives the following two equations respectively, in the unknowns A_α and B_β

$$B_\beta = A_\alpha \overline{[x]^\alpha \Phi_\beta [x]} \quad (11.5)$$

$$1 = A_\alpha^2 \overline{[x]^{2\alpha}} - 2 \sum_{\beta=0}^{\alpha-1} B_\beta A_\alpha \overline{[x]^\alpha \Phi_\beta [x]} + \sum_{\beta=0}^{\alpha-1} B_\beta^2 \quad (11.6)$$

Substituting (11.5) in (11.6) gives

$$A_\alpha = \frac{1}{\left[\overline{[x]^{2\alpha}} - \sum_{\beta=0}^{\alpha-1} \overline{[x]^\alpha \Phi_\beta [x]}^2 \right]^{1/2}} \quad (11.7)$$

Therefore substituting (11.5) and (11.7) in (11.3) gives
 recurrence equation for
 the ~~form~~ of the polynomials,

$$\Phi_\alpha [x] = \frac{[x]^\alpha - \sum_{\beta=0}^{\alpha-1} \overline{[x]^\alpha \Phi_\beta [x]} \Phi_\beta [x]}{\left[\overline{[x]^{2\alpha}} - \sum_{\beta=0}^{\alpha-1} \overline{[x]^\alpha \Phi_\beta [x]}^2 \right]^{1/2}} \quad (11.8)$$

These polynomials will exist if $[x]^\alpha = \overline{x_1^p x_2^q \dots x_n^w}$ exist for all α (i.e. all $p, q, \dots, w$) and the denominator of (11.8) is non-zero. They can be obtained by an experimental procedure similar to that described in chapter 3 for the first order case.

It is interesting to note that if any or all the signals $x_1, x_2, \dots, x_n$ are statistically independent of the rest, then the orthonormal conditions of equation (11.4) are satisfied by a function which separates, i.e.

$$\phi_{p,q,\dots,w}(x_1, x_2, \dots, x_n) = \theta_p(x_1) \phi'_{q,\dots,w}(x_2, \dots, x_n)$$

where $\theta_p(x_1)$ is orthonormal with respect to $p(x_1)$ and x_1 is statistically independent of $x_2, x_3, \dots, x_n$

Assuming that multipliers, adding amplifiers and attenuators alone, are used to synthesize these polynomials (by a similar procedure to that recommended in chapter 3), then it can be shown that the number of multipliers needed for an n th order polynomial with S inputs is given by:

$$\underline{S}_n \quad \text{where} \quad \underline{S} = S + (S-1) + (S-2) + \dots + 1$$

$$\text{and} \quad \underline{S}_n = \underbrace{\left\{ \begin{array}{c} S \\ S \\ S \\ \vdots \\ S \end{array} \right\}}_n$$

For $\left. \begin{array}{l} S=5 \\ n=3 \end{array} \right\}$ 70 multipliers are needed.

$\left. \begin{array}{l} S=3 \\ n=3 \end{array} \right\}$ 15 multipliers are needed.

11.7 Generalizations of probability density expansions.

Two expansions of higher order probability density functions are proposed here with very little attempt at demonstrating either their limitations or their properties:

11.7.1 The first generalization:

We postulate that there exists a class of n th order probability density functions for which the following expansion is possible:

$$p_n(x_1, x_2, \dots, x_n; \tau_1, \tau_2, \dots, \tau_{n-1}) = \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \dots \sum_{w=0}^{\infty} C_{p,q,\dots,w}(\tau_1, \tau_2, \dots, \tau_{n-1}) \theta_p(x_1) \gamma_q(x_2) \dots \psi_w(x_n) \times \rho(x_1) \rho(x_2) \dots \rho(x_n) \quad (11.9)$$

where

$$C_{p,q,\dots,w}(\tau_1, \tau_2, \dots, \tau_{n-1}) = \iint \dots \int \theta_p(x_1) \gamma_q(x_2) \dots \psi_w(x_n) \times p_n(x_1, x_2, \dots, x_n; \tau_1, \tau_2, \dots, \tau_{n-1}) dx_1 dx_2 \dots dx_n \quad (11.10)$$

and

$\theta_p(x_1)$		are orthonormal with respect to	$p(x_1)$
$\gamma_q(x_2)$	"	"	$p(x_2)$
$\psi_w(x_n)$	"	"	$p(x_n)$

11.7.2 The second generalization:

We also postulate that there exists a class of n th order probability density functions, for which the following expansion is possible:

$$\begin{aligned}
& p_n(x_1, x_2, \dots, x_n; \tau_1, \tau_2, \dots, \tau_{n-1}) \\
&= \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \dots \sum_{r=0}^{\infty} C_{p,q,\dots,r}(\tau_1, \tau_2, \dots, \tau_{n-1}) \\
&\quad \times \phi_{p,q,\dots,r}(x_1, x_2, \dots, x_n) p_n(x_1, x_2, \dots, x_n)
\end{aligned} \tag{11.11}$$

In vector notation this becomes:

$$P_n[x, \tau] = \sum_{\alpha=0}^{\infty} C_{\alpha}[\tau] \bar{\Phi}_{\alpha}[x] P_n[x] \tag{11.12}$$

where the polynomials $\bar{\Phi}_{\alpha}[x]$ are defined in section 11.6, and

$$C_{\alpha}[\tau] = \int \bar{\Phi}_{\alpha}[x] P_n[x, \tau] dx \tag{11.13}$$

Since $P_n[x, 0] = P_n[x]$ in equation (11.13) we have

$$C_{\alpha}[0] = \int \bar{\Phi}_{\alpha}[x] P_n[x] dx = 0 \quad \alpha \neq 0 \tag{11.14}$$

and $C_0[0] = 1$

Also by substitution of (11.8) in (11.13) we get

$$C_\alpha[\tau] = \frac{\lambda_\alpha[\tau] - \sum_{\beta=0}^{\alpha-1} \overline{[x]^\alpha \Phi_\beta[x]} C_\beta[\tau]}{\left[\overline{[x]^{2\alpha}} - \sum_{\beta=0}^{\alpha-1} \overline{[x]^\alpha \Phi_\beta[x]^2} \right]^{1/2}} \quad (11.15)$$

$$\text{where } \lambda_\alpha[\tau] = \int [x]^\alpha P_n[x, \tau] dx$$

This result also holds for the two-dimensional case treated in chapter 3.

11.8 Higher Order filters.

Some generalizations of the optimization procedure discussed in the earlier chapters are proposed here for filters of higher order. The limitations of these generalizations are not discussed.

11.8.1 The first canonical form.

We postulate that there exists a class of Kernel function of Zadeh's filters of class γ_n for which the following expansion is possible:

$$\begin{aligned} & K[x_1, x_2, \dots, x_n; \tau_1, \tau_2, \dots, \tau_n] \\ &= \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \dots \sum_{k=0}^{\infty} w_{a,b,\dots,k}(\tau_1, \tau_2, \dots, \tau_n) \theta_a(x_1) \gamma_b(x_2) \dots \psi_k(x_n) \end{aligned} \quad (11.16)$$

where $\theta_a(x_1)$; $\gamma_b(x_2)$ and $\psi_k(x_n)$ are defined in section 3.1.1.

A class of non-linear filters can be defined by equation (11.16), optimization of which involves the solution of k simultaneous integral equations of n th order.

11.8.2 The second canonical form.

We postulate that there exists a class of Kernel function of Zadeh's class η_n for which the following expansion is possible:

$$K[x_1, x_2, \dots, x_n; \tau_1, \tau_2, \dots, \tau_n] \\ = \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \dots \sum_{k=0}^{\infty} w_{a,b,\dots,k}(\tau_1, \tau_2, \dots, \tau_n) \phi_{a,b,\dots,k}(x_1, x_2, \dots, x_n) \quad (11.17)$$

Optimization of the class of filters defined by equation (11.17) again involves the solution of k simultaneous ^{integral} equations of n th order.

11.8.3 Some multi-input filters of class η_n .

For a multi-input filter of class η_n , the Kernel function is of the form $K[x_1, x_2, \dots, x_n; \tau]$ and each of the expansions (11.16) and (11.17) yield different canonical forms.

Optimization is no more difficult than in the first order single input case, but the complexity of the Φ_α functions is naturally greater than in the single input

case. However, it seems reasonable to suppose that this generalization, unlike the rest, can be applied directly to a practical problem without further simplification, provided the number of inputs is not too great, by using the same method as in the single input case.

11.8.4 Wiener's filter using generalized orthonormal polynomials.

The orthonormal polynomials of equation (11.8) can be used in the Wiener filter of fig.4 and possess all the mathematical advantages of Bose's gate functions (see chapter 4). Whilst the complexity of these polynomial function generators does become great, for a large number of coefficients, U_5 and high order polynomials, it is considered that by using continuous orthonormal polynomials, a more rapid convergence of the filter function expansion will be obtained than by using discontinuous gate functions. Therefore, fewer terms will be needed for the same accuracy. It must be admitted, however, that the gate functions possess certain advantages from considerations of physical synthesis, since products of gate functions are more easily obtained.

12.0 REFERENCES.

- (1) ZADEH, L.A. "A Contribution to the Theory of Non-Linear Systems", Journal of the Franklin Institute, 1953, 255, p.387.
- (2) ZADEH, L.A. "Optimum Non-Linear Filters", Journal of Applied Physics, 1953, 24, p.396.
- (3) ZADEH, L.A. "On the Representation of Non-Linear Operators", I.R.E. WESCON Convention Record, 1957, 1, Pt.2, p.105-13.
- (4) WIENER, N. "Non-linear Problems in Random Theory", 1958, Chapman & Hall.
- (5) LAMPARD, D.G. "A New Method of Determining Correlation Functions of Stationary Time Series", Proc. I.E.E., 1955, 102, part C, p.35.
- (6) LANING, J.H. "Random Processes in Automatic Control, 1956, and BATTIN, H.B. McGraw-Hill.
- (7) SINGLETON, H.A. "Theory of Non-linear Transducers", 1950, M.I.T. Technical Report No. 160.
- (8) BOSE, A.G. "A Theory of Non-linear Systems", 1956, M.I.T. Technical Report no. 309.
- (9) SHANNON, C.E. "Communication in the Presense of Noise", Proc. I.R.E., 1949, 37, 1, p.10-21.
- (10) FULLER, A.T. "Design of Control Systems which contain Saturating Elements", 1959, Ph.D. Thesis, Cambridge University.
- (11) COURANT, R. "Differential and Integral Calculus", 1936, Blackie.

- (12) WIENER, N. "Extrapolation, Interpolation and Smoothing of Stationary Time Series", 1949, Wiley.
- (13) BARRETT, J.F.
and LAMPARD, D.G. "An Expansion for some Second Order Probability Distributions and its Application to Noise Problems", I.R.E. Trans., 1955, IT-1, p.10.
- (14) BARRETT J.F.
and COALES J.F. "An Introduction to the Analysis of Non-linear Control Systems with Random Inputs", Proc. I.E.E. 1956, 103, part C, p.190.
- (15) BATEMAN "Higher Transcendental Functions", Vol.3, McGraw-Hill.
- (16) NUTTALL, A.H. "Theory and Application of the Separable class of Random Processes", 1958, M.I.T. Technical Report, No. 343.
- (17) BROWN, J.L. "On a Cross-Correlation Property for Stationary Random Processes", I.R.E. Trans., 1957, IT-3, p.28.
- (18) BUSSGANG, J.J. "Cross-Correlation Functions of Amplitude Distorted Gaussian Signals", 1952, M.I.T. Technical Report No. 216.
- (19) WONHAM, W.M. "Transition Probability Densities of the Smoothed Random Telegraph Signal", Journal of Electronics and Control, 1959, 6, 4, pp. 376 - 384.

- (20) RICE, S.O. "Mathematical Analysis of Random Noise",
Bell System Technical Journal, 1945,
23, 24, p.78 and 138 respectively.
- (21) BURRELL, C.M. "What is Control Engineering", Part 1-7,
and Control, 1958 and 59, Vols. 1 and 2,
LUBBOCK, J.K. July to January inclusive.
- (22) GOODMAN, T.P. "Determination of System Characteristics
and from Normal Operating Records", A.S.M.E.
RESWICK, J.B. Trans., 1956, 78, p.259.
- (23) WESTCOTT, J.H. "Process Dynamics from Normal Disturbance
Records", S.I.T. Conference, Cambridge,
1956.
- (24) WOODROW, R.A. "Closed-Loop Dynamics from Normal
Operating Records, Trans. S.I.T., 1958,
10, 3, p.101.
- (25) BOOTON, R.C. "Non-Linear Control Systems with
Statistical Inputs", 1952, M.I.T. Report
No. 61.
- (26) BOOTON, R.C., "Nonlinear Servomechanisms with Random
MATTHEWS, M.V. Inputs", 1953, M.I.T. Report No. 70.
and
SEIFERT, W.W.
- (27) TRUXAL, J.G. "Control System Synthesis", 1955,
McGraw-Hill.
- (28) SOMERVILLE, M.J. "Multi-Gain Representation for a Single-
and Valued Non-Linearity with Several Inputs",
ATHERTON, D.P. and the Evaluation of their Equivalent
Gains by a Cursor Method".

- (28) contd. Proc. I.E.E., 1958, 105, part C, p.537-49.
- (29) NEWTON, G.C. "Compensation of Feedback Control Systems subject to Saturation", Journal of the Franklin Institute, 1952, 254, p.281-286 and 391-413.
- (30) TSIEN, H.S. "Engineering Cybernetics", 1954, McGraw-Hill.
- (31) JAMES, H.N., "The Theory of Servomechanisms", 1947, McGraw-Hill.
NICHOLS, N.B.
and
PHILLIPS, R.S.
- (32) NEWTON, G.C. "Analytical Design of Linear Feedback Controls", 1957, Wiley.
GOULD, L.A.,
and
KAISER, J.F.
- (33) BEUTLER, F.J. "Prediction and Filtering for Random Parameter Systems", I.R.E. Trans., 1958, IT-4, 4, p.166.
- (34) ZADEH, L.A. "Optimum Non-Linear Filters for the Extraction and Detection of Signals", a paper presented at I.R.E. Nat. Convention, March 25, 1953.

- (35) YOULA, D.C. "The use of the Method of Maximum Likelihood in Estimating Continuous Modulated Intelligence which has been Corrupted by Noise". I.R.E. Trans., 1954, IT-3, pp.90-105.
- (36) LUBBOCK, J.K. "Optimization of a class of Non-Linear Filters", 1959, I.R.E. Monograph No.344E.
- (37) LUBBOCK, J.K. "On a class of semi-separable processes", The Journal of Electronics and Control. (To be published).
- (38) LUBBOCK, J.K. "Non-Linear Feedforward Compensation of an otherwise Linear Feedback System". To be published.
- (39) COALES, J.F. "The Application of Information Theory to Data-Transmission Systems and the possible use of Binary Coding to increase Channel Capacity", Proc. I.E.E., 1953, 100, 3, P291.
- (40) FULLER, A.T. "Sampling Errors in the Measurement of Auto-correlation", J. Electronics and Control, 1958, 4, p551.

- (41) BENNETT   "The Generation and Measurement of low-
and   frequency Random Noise", J. Applied Physics,
FULTON   1951, 22, pp 1187-1191.

13.0 APPENDICES.

13.1 Analytical and experimental determination of the orthonormal polynomials for a clipped Gaussian signal.

All physical systems are non-linear for sufficiently large input amplitude, although linear analysis often can be used for bounded signals if the system is operated over its linear range. A most common kind of non-linearity is saturation, which prevents the output from exceeding a peak value independent of the amplitude of the input; less frequently the system may be destroyed by excessive input amplitudes without saturating; in either case the output is bounded although, in theory, the input may be unbounded.³⁹

It is concluded, therefore, that all physically realizable signals are bounded because they must be generated by a physical system.

In this appendix the effect of saturating or clipping a Gaussian signal is considered, ^{the} its moments and orthonormal polynomials are determined.

13.1.1 The analytical method.

The moments are determined using ^{the} equation:

$$\overline{x^n} = \int_{-\infty}^{\infty} x^n p(x) dx$$

where

$$\left. \begin{aligned}
 p(x) &= \frac{1}{\sqrt{2\pi} \sigma} e^{-x^2/2\sigma^2} & -\sigma < x < \sigma \\
 &= \frac{I_1 \delta(x-\sigma)}{\sqrt{2\pi} \sigma} & x = \sigma \\
 &= \frac{I_1 \delta(x+\sigma)}{\sqrt{2\pi} \sigma} & x = -\sigma
 \end{aligned} \right\} \quad (13.1.1)$$

and

$$I_1 = \int_{-\sigma}^{\infty} e^{-x^2/2\sigma^2} dx$$

since $p(x)$ is a symmetrical density function, odd moments must be zero. Even moments are given by:

$$\overline{x^{2n}} = \int_{-\sigma}^{\sigma} \frac{x^{2n}}{\sqrt{2\pi} \sigma} e^{-x^2/2\sigma^2} dx + 2 \int_0^{\infty} \frac{x^{2n}}{\sqrt{2\pi} \sigma} \delta(x-\sigma) dx I_1$$

Integrating by parts gives:

$$\overline{x^{2n}} = \frac{1}{\sqrt{2\pi} \sigma} \left[\left(\frac{x^{2n+1}}{2n+1} e^{-x^2/2\sigma^2} \right)_{-\sigma}^{\sigma} + \frac{1}{\sigma^2} \int_{-\sigma}^{\sigma} \frac{x^{2n+2}}{2n+1} e^{-x^2/2\sigma^2} dx \right] + 2 \sigma^{2n} I_1$$

$$\therefore \overline{x^{2n}} = \frac{\overline{x^{2n+2}}}{(2n+1)\sigma^2} + \frac{2\sigma^{2n+2}}{\sqrt{2\pi}\sigma} \left[\frac{1}{2n+1} e^{-y^2/2} + I_1\left(1 - \frac{y^2}{2n+1}\right) \right]$$

divide through by $\sigma^{2n}/2n+1$

$$\frac{\overline{x^{2n+2}}}{\sigma^{2n+2}} = (2n+1) \frac{\overline{x^{2n}}}{\sigma^{2n}} - \sqrt{\frac{2}{\pi}} \frac{y^{2n}}{\sigma} \left[1 e^{-y^2/2} + I_1(2n+1 - y^2) \right] \quad (13.1.2)$$

have been

The moments ~~are~~ calculated using equation (13.1.2) for $r = 3$

$$\overline{x^2} = \sigma^2 \quad \overline{x^4} = 2.9206 \sigma^4 \quad \overline{x^6} = 13.3778 \sigma^6 \quad (13.1.3)$$

$$\overline{x^8} = 78.3536 \sigma^8 \quad \overline{x^{10}} = 529.8366 \sigma^{10} \quad \overline{x^{12}} = 3910.83 \sigma^{12}$$

These should be compared with the unbounded Gaussian moments:

$$\overline{x^2} = \sigma^2 \quad \overline{x^4} = 3 \sigma^4 \quad \overline{x^6} = 15 \sigma^6 \quad (13.1.4)$$

$$\overline{x^8} = 105 \sigma^8 \quad \overline{x^{10}} = 945 \sigma^{10} \quad \overline{x^{12}} = 10395 \sigma^{12}$$

It is perhaps a little surprising at first to find such a noticeable difference in these two sets of moments, since

a Gaussian signal exceeds 3σ on the average for less than 0.14% of the total time and pen recordings of both the clipped and unclipped Gaussian signals would not seem to differ in shape very appreciably. However, the difference does occur at high amplitude so that it is to be expected that this will have a greater effect on the higher moments than on the lower ones.

The orthonormal polynomials for the clipped Gaussian signal are determined using equation (3.6).

$$\theta_0(x) = 1 \quad ; \quad \theta_1(x) = x/\sigma$$

$$\theta_2(x) = -0.7216 + 0.7216x^2/\sigma^2$$

$$\theta_3(x) = -1.3265x/\sigma + 0.4542x^3/\sigma^3 \quad (13.1.5)$$

$$\theta_4(x) = 0.7032 - 1.5167x^2/\sigma^2 + 0.2786x^4/\sigma^4$$

$$\theta_5(x) = 1.8027x/\sigma - 1.4199x^3/\sigma^3 + 0.1752x^5/\sigma^5$$

These should be compared with the orthonormal polynomials for an unclipped Gaussian signal, which are Hermite polynomials:

$$\theta_0(x) = 1 \quad ; \quad \theta_1(x) = \frac{x}{\sigma}$$

$$\theta_2(x) = -0.7071 + 0.7071 \frac{x^2}{\sigma^2}$$

$$\theta_3(x) = -1.2247 \frac{x}{\sigma} + 0.4082 \frac{x^3}{\sigma^3} \quad (13.1.6)$$

$$\theta_4(x) = 0.6124 - 1.2247 \frac{x^2}{\sigma^2} + 0.2041 \frac{x^4}{\sigma^4}$$

$$\theta_5(x) = 1.3693 \frac{x}{\sigma} - 0.9129 \frac{x^3}{\sigma^3} + 0.0913 \frac{x^5}{\sigma^5}$$

13.1.2 The experimental results.

The experimental measurement of moments of a clipped Gaussian signal was made using PACE. The output of the PACE noise generator was passed through two stages of low pass filtering (see fig.37) and then through a diode clipping circuit which had been adjusted to clip at 100 volts; ~~while~~ the mean-squared value of the signal was set to $33\frac{1}{3}$ volts. The attenuators, numbers 27-34 inclusive and 37 and 38, were adjusted to the values indicated in the table below so that the computed integrated voltage on each integrator over 15 mins. was approximately 100 volts. The experimental and analytical results are given.

POTENTIOMETERS	37/38	33/34	31/32	29/30	27/28
POTENTIOMETER SETTINGS	0.01	0.03081	0.06055	0.09304	0.1238
INTEGRATORS	19	17	16	15	14
READING	$\overline{x^2}$	$\overline{x^4}$	$\overline{x^6}$	$\overline{x^8}$	$\overline{x^{10}}$
EXPERIMENTAL RESULTS INTEGRATED VOLTS OVER 15 mins.	99.04	99.07	99.10	97.83	95.56
	99.91	99.74	101.16	101.47	100.54
	-	102.4	104.35	105.37	104.99
	99.33	99.63	99.79	99.97 ₆₆	98.72
DRIFT OVER 15 mins. WITH INPUT EARTHED	$\overset{+}{00.02}$	$\overline{00.03}$	$\overline{01.44}$	$\overset{+}{01.87}$	$\overset{+}{00.57}$
DRIFT CORRECTED READING (15 mins)	99.41	100.24	102.54	99.29	99.43
CALCULATED RESULT (15 mins)	100	99.996	100.0047	100.0008	99.976

Insufficient results have been taken to enable an estimate of their variance to be obtained, but in the next section the expected variance in some of the measured moments of a Gaussian signal are calculated. It is seen that the experimental results agree with the calculated results within a tolerance which is probably not significantly greater than would be expected from their statistical variance.

The polynomials were synthesized for $V = 33\frac{1}{3}$ volts, using multipliers, attenuators and adding amplifiers. Fig.22 shows the circuit diagram. In the synthesis a scale factor was attached to each polynomial (sometimes positive and sometimes negative, the choice of sign depending only on ease of synthesis) so that each new polynomial function thus derived gave an output voltage of 100 volts for 100 volts input, thereby utilizing the full range of the computer for greatest accuracy. The new polynomial functions were plotted on a variplotter (see fig.38). Signal samples, taken at the points indicated in the circuit of fig.22, show the output signals of each new polynomial function generator with a clipped Gaussian input of $V = 33\frac{1}{3}$ volts (see fig.39).

13.1.3 The variance of the measured values of certain joint moments over finite samples.

The finite signal samples are defined by:

$$\left. \begin{aligned} u(t) &= \begin{cases} x(t) & |t| \leq T \\ 0 & |t| > T \end{cases} \\ v(t) &= \begin{cases} y(t) & |t| \leq T \\ 0 & |t| > T \end{cases} \end{aligned} \right\} \quad (13.1.7)$$

If both x and y are ergodic, then

$$E[x^n y^m] = \lim_{A \rightarrow \infty} \frac{1}{2A} \int_{-A}^A x^n(t) y^m(t) dt$$

In practice $x^n y^m$ is averaged over finite samples only; it is shown here that the result is a satisfactory measure of the expectation of $x^n y^m$ if the sample length is long enough.

We define

$$\begin{aligned} \overline{u^n v^m} &= \frac{1}{2T} \int_{-\infty}^{\infty} u^n(t) v^m(t) dt \\ &= \frac{1}{2T} \int_{-T}^T x^n(t) y^m(t) dt \end{aligned} \quad (13.1.8)$$

Due to the finite sample lengths, some statistical variation will occur between successive readings of $\overline{u^n v^m}$ taken over different samples of the same length. However,

$$\begin{aligned}
 E[\overline{u^n v^m}] &= \lim_{T \rightarrow \infty} \overline{u^n v^m} \\
 &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x^n(t) y^m(t) dt \\
 &= \overline{x^n y^m}
 \end{aligned} \tag{13.1.9}$$

Thus it is seen that as T is increased, the measured value of $\overline{u^n v^m}$ converges to $\overline{x^n y^m}$.

The variance of the measured values of $\overline{u^n v^m}$ can be found as follows,

$$\begin{aligned}
 E[\overline{u^n v^m}^2] &= E\left[\frac{1}{4T^2} \int_{-T}^T x^n(t_1) y^m(t_1) dt_1 \int_{-T}^T x^n(t_2) y^m(t_2) dt_2\right] \\
 &= \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T E[x^n(t_1) y^m(t_1) x^n(t_2) y^m(t_2)] dt_1 dt_2
 \end{aligned} \tag{13.1.10}$$

$$\text{Let } E[x^n(t_1) x^n(t_2) y^m(t_1) y^m(t_2)] = m(t_1, t_2) = m(t_2, t_1) \tag{13.1.11}$$

Since the processes are stationary m is a function of $t_2 - t_1$ only.

$$\begin{aligned}
 \text{Let } t_2 - t_1 &= \tau \\
 \text{then } E[\overline{u^n v^m}^2] &= \frac{1}{4T^2} \int_{-T}^T dt_1 \int_{-T-t_1}^{T-t_1} m(\tau) d\tau
 \end{aligned} \tag{13.1.12}$$

The region of integration is shown in the diagram; changing the order of integration gives:

$$\begin{aligned}
 E[\overline{u^n v^m}] &= \frac{1}{4T^2} \int_{-2T}^0 d\tau \int_{-T}^{T-\tau} m(\tau) dt + \frac{1}{4T^2} \int_0^{2T} d\tau \int_{-T}^{T-\tau} m(\tau) dt \\
 &= \frac{1}{4T^2} \int_{-2T}^0 (2T+\tau) m(\tau) d\tau + \frac{1}{4T^2} \int_0^{2T} (2T-\tau) m(\tau) d\tau
 \end{aligned} \tag{13.1.13}$$

But $m(\tau) = m(-\tau)$ (from equation 13.1.11)

$$\therefore E[\overline{u^n v^m}] = \frac{1}{2T^2} \int_0^{2T} (2T-\tau) m(\tau) d\tau \tag{13.1.14}$$

Therefore the variance in the measured values of $\overline{u^n v^m}$ is given by

$$\begin{aligned}
 \sigma_{nm}^2 &= E[\overline{u^n v^m}]^2 - (E \overline{u^n v^m})^2 \\
 &= \frac{1}{2T^2} \int_0^{2T} (2T-\tau) m(\tau) d\tau - \overline{x^n y^m}^2
 \end{aligned} \tag{13.1.15}$$

(from equations 13.1.14 and 13.1.9)

An example: In section 13.1.2 the measured moments of a clipped Gaussian signal are tabulated. The variance of these moments is given by equation 13.1.15 for the case when $m = 0$ i.e.

$$\sigma_n^2 = \frac{1}{2T^2} \int_0^{2T} (2T - \tau) \overline{x^n(t)x^n(t+\tau)} d\tau - \overline{x^n}^2 \quad (13.1.16)$$

Attempts to determine $\overline{x^n(t)x^n(t+\tau)}$ for the clipped Gaussian signal have been unsuccessful but for a truly Gaussian signal with zero mean value $\overline{x^n(t)x^n(t+\tau)}$ can be found from either equations (6.6) and (3.14) or equations (3.13) and (7.18). When $n = 2$ using equations (6.6) and (3.14)

$$\frac{1}{2} \overline{He_2 \left[\frac{x(t)}{\sigma_x} \right] He_2 \left[\frac{x(t+\tau)}{\sigma_x} \right]} = \frac{1}{2} - \frac{\overline{x^2}}{\sigma_x^2} + \frac{1}{2} \frac{\overline{x^2(t)x^2(t+\tau)}}{\sigma_x^4}$$

$$\therefore \overline{x^2(t)x^2(t+\tau)} = \sigma_x^4 [2\rho^2(\tau) + 1] \quad (13.1.17)$$

We shall consider the case when $\overline{x} = 0$
 then from equation 13.1.16 $\rho(\tau) = e^{-k|\tau|}$

$$\sigma_2^2 = \frac{\sigma_x^4}{2T^2} \int_0^{2T} (2T - \tau) [2e^{-k\tau} + 1] d\tau - \sigma_x^4 \quad (13.1.18)$$

$$\therefore \sigma_2^2 = \frac{\sigma_x^4}{kT} \left[1 - \frac{1 - e^{-4kT}}{4kT} \right] \quad (13.1.18)$$

which agrees with results of Fuller⁴⁰ and of Bennett and Fulton⁴¹. (See also reference 6, page 163).

For large values of kT

$$\sigma_2^2 \simeq \frac{\sigma_x^4}{kT} \quad (13.1.19)$$

when $n=4$ it can be shown by a similar method that:

$$\overline{x^4(t)x^4(t+\tau)} = [24\rho^4(\tau) + 72\rho^2(\tau) + 9] \sigma_x^8 \quad (13.1.20)$$

Then from equations 13.1.16, when $\rho(\tau) = e^{-k|\tau|}$ it can be shown that

$$\sigma_4^2 = \frac{\sigma_x^8}{kT} \left[42 - \frac{36(1 - e^{-4kT}) + 3(1 - e^{-8kT})}{4kT} \right] \quad (13.1.21)$$

For large values of kT

$$\sigma_4^2 \simeq 42 \frac{\sigma_x^8}{kT} \quad (13.1.22)$$

In section 13.1.2 the time taken for each reading was 15 minutes, whilst 62.5 secs.

$$\therefore kT = 900 \times 62.5$$

From equation 13.1.19

$$\frac{\sigma_2}{x^2} = \frac{1}{\sqrt{kT}} \quad \text{Since } \overline{x} = 0$$

$$\therefore \frac{\sigma_2}{x^2} = \frac{1}{317}$$

$$\approx \frac{1}{3} \%$$

From equation 13.1.22

$$\frac{\sigma_4}{x^4} = \frac{\sqrt{42}}{3\sqrt{kT}} \quad \text{Since } \overline{x^4} = 3\sigma_x^4$$

$$= \frac{6.5}{3 \times 317}$$

$$= 0.7 \%$$

It is not expected that the variance in the measured moments of a clipped Gaussian signal (clipped at 3σ) would be appreciably greater than these values which have been calculated for a truly Gaussian signal.

13.2 Lampard's method for measuring correlation coefficients.

A method of measuring correlation coefficients of stationary signals proposed by Lampard⁵ obviates the need to produce finite delays. It is assumed, however, that the correlation coefficient can be expanded in the infinite series

$$\phi(\tau) = \sum_{n=0}^{\infty} a_n \psi_n(\tau) [h(\tau)]^{\gamma} \quad 0 \leq \tau < \infty \quad (13.2.1)$$

where $\psi_n(\tau)$ are an orthonormal set of functions with respect to the weighting function $h(\tau)$

$$\text{i.e.} \quad \int_0^{\infty} \psi_n(\tau) \psi_m(\tau) h(\tau) d\tau = \delta_{nm} \quad (13.2.2)$$

where

$$a_n = \int_0^{\infty} \phi(\tau) \psi_n(\tau) [h(\tau)]^{1-\gamma} d\tau \quad (13.2.3)$$

For the circuit shown in fig.40

$$y(t) = \int_0^{\infty} x(t-\tau) \omega(\tau) d\tau$$

where $\omega(\tau)$ is the delta function response of the linear network.

The multiplier output is given by

$$y(t)x_2(t) = \int_0^{\infty} x_1(t-\tau)x_2(t)\omega(\tau)d\tau$$

Averaging over time

$$\overline{y(t)x_2(t)} = \int_0^{\infty} \phi_{x_1x_2}(\tau)\omega(\tau)d\tau \quad (13.2.4)$$

$$\text{Since } \overline{x_1} = 0 \quad \overline{x_2} = 0$$

A comparison of equations (13.2.4) and (13.2.3) shows that

$$a_n = \overline{y(t)x_2(t)}$$

if

$$\omega(\tau) = \psi_n(\tau) [h(\tau)]^{1-\gamma} \quad (13.2.5)$$

Assuming for the moment that filters can be physically realized with $\psi_n(\tau) [h(\tau)]^{1-\gamma}$ as their delta function responses, then the cross-correlation ($\tau \geq 0$) between x_1 and x_2 can be obtained as an expansion of the orthonormal functions $\psi_n(\tau)$ by a simple experimental procedure. Also, since $\phi_{12}(\tau) = \phi_{21}(-\tau)$ the correlation for $\tau \leq 0$ can be obtained by interchanging x_1 and x_2 .

13.2.1 Laguerre polynomials.

The Laguerre function is defined by:

$$L_n(t) = \frac{e^t}{(n-1)!} \frac{d^{n-1}}{dt^{n-1}} (t^{n-1} e^{-t}) \quad (4.2)$$

which is orthonormal over the range $(0 - \infty)$

$$\text{i.e. } \int_0^\infty L_n(t) L_m(t) e^{-t} dt = \begin{cases} 0 & n \neq m \\ 1 & n = m \end{cases}$$

The generating function is:

$$\frac{e^{-ut/(1-u)}}{1-u} = \sum_{n=0}^{\infty} L_n(t) u^{n-1} \quad (13.2.6)$$

The first few polynomials are:

$$L_1(t) = 1 \quad L_2(t) = 1 - t$$

$$L_3(t) = 1 - 2t + \frac{1}{2}t^2 \quad L_4(t) = 1 - 3t + \frac{3}{2}t^2 - \frac{1}{6}t^3$$

$$L_5(t) = 1 - 4t + 3t^2 - \frac{2}{3}t^3 + \frac{1}{24}t^4 \quad (13.2.7)$$

In equation (13.2.5) let

$$\psi_n(\tau) = \mathcal{L}_n(\alpha\tau) ; \quad h(\tau) = e^{-\alpha\tau} \quad (13.2.8)$$

then

$$W_n(\tau) = \mathcal{L}_n(\alpha\tau) [e^{-\alpha\tau}]^{1-\gamma} \quad (13.2.9)$$

Two possible cases exist:

a. $\gamma = 0$ whence

$$W_n(\tau) = \mathcal{L}_n(\alpha\tau) e^{-\alpha\tau} \quad (13.2.10)$$

It can be shown that the unilateral transform becomes:

$$W_n(p) = \int_0^{\infty} \mathcal{L}_n(\alpha\tau) e^{-(\alpha+p)\tau} d\tau = \frac{p^{n-1}}{(p+\alpha)^n} \quad (13.2.11)$$

A network with the transfer function $\alpha W_n(p)$ is shown in fig.41a.

b. $\gamma = \frac{1}{2}$ whence

$$W_n(\tau) = \mathcal{L}_n(\alpha\tau) e^{-\frac{\alpha}{2}\tau} \quad (13.2.12)$$

It can be shown that the unilateral transform becomes:

$$W_n(p) = \int_0^{\infty} \mathcal{L}_n(\alpha\tau) e^{-(\frac{\alpha}{2}+p)\tau} d\tau = \frac{(p - \frac{\alpha}{2})^{n-1}}{(p + \frac{\alpha}{2})^n} \quad (13.2.13)$$

A network with the transfer function $\frac{\alpha}{2} W_n(\rho)$ is shown in fig.41b.

13.2.2 Another set of orthonormal functions.

Laning and Battin (reference 6, p.381) have defined a class of orthonormal functions which may be used for the expansion of correlation coefficients. These functions can be expressed as:

$$\psi_n(\alpha t) = \sum_{v=0}^n A_v e^{-v\alpha t} \quad (13.2.14)$$

such that

$$\int_0^{\infty} \psi_n(\alpha t) \psi_m(\alpha t) e^{-2\alpha t} dt = \begin{cases} 0 & n \neq m \\ 1 & n = m \end{cases} \quad (13.2.15)$$

where α is a constant which can be chosen so as to give a reasonably rapid convergence of the following expansion for the correlation coefficient:

$$\phi(\tau) = \sum_{n=1}^{\infty} a_n \psi_n(\alpha \tau) \left[e^{-2\alpha \tau} \right]^{\gamma} \quad (13.2.16)$$

$0 \leq \tau < \infty$

where

$$a_n = \int_0^{\infty} \phi(\tau) \psi_n(\alpha \tau) \left[e^{-2\alpha \tau} \right]^{1-\gamma} d\tau \quad (13.2.17)$$

For experimental measurement of the correlation

coefficient it is necessary to synthesize a network with a delta function response

$$W(\tau) = \psi_n(\alpha\tau) \left[e^{-2\alpha\tau} \right]^{1-\gamma} \quad (13.2.18)$$

The choice of γ make little difference to the method of synthesis of the linear network.

For the case $\gamma = 1/2$ it is necessary to synthesize a transfer function of the form

$$W_n(p) = \sum_{r=0}^n \frac{A_r}{p + r\alpha} \quad (13.2.19)$$

This transfer function is easily produced using a number of feedback amplifiers to produce the transfer functions $\frac{r\alpha}{p+r\alpha}$ (by the same method as that shown in fig.41) and using attenuators and an adding amplifier to form the linear summation of their outputs.

Tables of the functions $\psi_n(\alpha t)$ are worked out in reference 6. Some of the lower order functions are:

$$\begin{aligned} \psi_1(t) &= \sqrt{2\alpha} e^{-\alpha t} \\ \psi_2(t) &= \sqrt{\alpha} \left[6e^{-2\alpha t} - 4e^{-\alpha t} \right] \\ \psi_3(t) &= \sqrt{6\alpha} \left[10e^{-3\alpha t} - 12e^{-2\alpha t} + 3e^{-\alpha t} \right] \\ \psi_4(t) &= \sqrt{2\alpha} \left[70e^{-4\alpha t} - 120e^{-3\alpha t} + 60e^{-2\alpha t} - 8e^{-\alpha t} \right] \\ \psi_5(t) &= \sqrt{10\alpha} \left[126e^{-5\alpha t} - 280e^{-4\alpha t} + 210e^{-3\alpha t} - 60e^{-2\alpha t} + 5e^{-\alpha t} \right] \end{aligned} \quad (13.2.20)$$

13.3 Solutions of the Wiener Hopf equation and simultaneous integral equations of Wiener Hopf type.

In this appendix methods are described for solving the Wiener Hopf equations (5.32) and (9.33) for the optimum linear filter and the simultaneous equations (5.27) for the optimum multipath memory filter.

13.3.1 The solution of the Wiener-Hopf equation.

The optimum-linear weighting function is given by the solution of equation (5.32)

$$\int_0^{\infty} w(\tau_2) \phi_{xx}(\tau_1 - \tau_2) d\tau_2 = \phi_{xz}(\tau_1) \quad (5.32)$$

$\tau_1 \geq 0$

$$\therefore \int_0^{\infty} w(\tau_2) \phi_{xx}(\tau_1 - \tau_2) d\tau_2 - \phi_{xz}(\tau_1) = H(\tau_1)$$

where $H(\tau_1) = 0 \quad \tau_1 \geq 0$ (13.3.1)

Taking the bilateral Laplace transform of both sides gives:

$$W(p) \Phi_{xx}(p) - \Phi_{xz}(p) = H(p) \quad (13.3.2)$$

where $H(p)$ has singularities in the right hand half of the p plane (abbreviated R.H.H.) only, since $H(\tau_1) = 0 \quad \tau_1 \geq 0$ (see section 13.4.1).

In most practical cases, it is found that all the bilateral transforms $\Phi_{xx}(\rho)$, $\Phi_{xz}(\rho)$, $W(\rho)$ and $H(\rho)$ are rational algebraic functions of ρ with real coefficients; also since $\Phi_{xx}(\rho)$ is an even function of ρ^* , it can be factorized in the form

$$\Phi_{xx}(\rho) = \Phi_{xx}^+(\rho) \Phi_{xx}^-(\rho) \quad (13.3.3)$$

where $\Phi_{xx}^-(\rho)$ has poles and zeros in the R.H.H. only
and $\Phi_{xx}^+(\rho)$ " " " " " " L.H.H. only.
Then equation (13.3.2) becomes:

$$W(\rho) \Phi_{xx}^+(\rho) - \frac{\Phi_{xz}(\rho)}{\Phi_{xx}^-(\rho)} = \frac{H(\rho)}{\Phi_{xx}^-(\rho)} \quad (13.3.4)$$

where $W(\rho)$ is the transfer function of a stable system and has poles in the L.H.H. only (i.e. $W(\tau)=0$ $\tau < 0$).

A solution to equation (13.3.4) may be obtained by making the assumption that

$$\lim_{\rho \rightarrow \infty} W(\rho) \Phi_{xx}^+(\rho) = 0 \quad (13.3.5)$$

Then $W(\rho) \Phi_{xx}^+(\rho)$ contains no R.H.H. poles and its partial fraction expansion has no terms of the form $a_i \rho^i$ $i \geq 0$

* Since $\phi_{xx}(\tau)$ is an even function of τ .

also, since $\mathcal{H}_1(\rho)$ contains no L.H.H. poles we can write:

$$\omega(\rho) = \frac{1}{\Phi^+(\rho)} \left[\frac{\Phi_{xz}(\rho)}{\Phi_{xx}^-(\rho)} \right]_+ \quad (13.3.6)$$

where the suffix + indicates that only those terms of the partial fraction expansion representing poles in the L.H.H.

of the function $\left[\frac{\Phi_{xz}(\rho)}{\Phi_{xx}^-(\rho)} \right]$, are included.

Equation (13.3.6), giving an explicit solution for $\omega(\rho)$, is arrived at only after making the assumption implied by equation (13.3.5). However, this assumption is not unreasonable, since equation (13.3.5) can always be satisfied by imposing sufficiently tight restrictions on $\omega(\rho)$ of the form

$$\lim_{p \rightarrow \infty} \frac{W(p)}{p^i} = \text{constant} \quad (13.3.7)$$

Where i , known as the order of $W(p)$, is chosen to be just small enough for any particular $\Phi^+(p)$, so that equation (13.3.5) is satisfied.

When $i \geq 0$ only approximate physical filters can be realized, but the approximation need not be serious in practice if i is not too large.

13.3.2 The method of undetermined coefficients.

The method of undetermined coefficients due to Wiener (See p.106 of reference 12) is a generalization of the above method. A modification of Wiener's solution is given here.

The simultaneous integral equations (5.27) are:

$$\sum_{m=1}^R \int_0^{\infty} W_m(\tau_2) C_{nm}(\tau_1 - \tau_2) d\tau_2 - D_n(\tau_1) \delta_2 = H_n(\tau_1) \quad (13.3.8)$$

where $H_n(\tau_1) = 0 \quad \tau_1 > 0$

$$n = 1, 2, \dots, R$$

Taking the bilateral transform of both sides of equation (13.3.8) we get:

$$\sum_{m=1}^R W_m(p) C_{nm}(p) - D_n(p) = H_n(p) \quad (13.3.9)$$

Where $\mathcal{H}_n(p)$ has poles in the R.H.H. only and all the transforms involved in equation (13.3.9) are assumed to be rational algebraic functions with real coefficients.

For simplicity, the method is described for two simultaneous equations only:

Let

$$\Phi(p) = \begin{vmatrix} b_{11}(p) & b_{12}(p) \\ b_{21}(p) & b_{22}(p) \end{vmatrix} \quad (13.3.10)$$

then

$$\Phi(p) \omega_1(p) = \begin{vmatrix} D_{11} + \mathcal{H}_1(p) & b_{12}(p) \\ D_{21} + \mathcal{H}_2(p) & b_{22}(p) \end{vmatrix} \quad (13.3.11)$$

and

$$\Phi(p) \omega_2(p) = \begin{vmatrix} b_{11}(p) & D_{11}(p) + \mathcal{H}_1(p) \\ b_{21}(p) & D_{21}(p) + \mathcal{H}_2(p) \end{vmatrix} \quad (13.3.12)$$

$$\begin{aligned} \therefore \Phi(p) \omega_1(p) &= [D_{11} b_{22} - D_{21} b_{12}] \\ &\quad + \mathcal{H}_1 b_{22} + \mathcal{H}_2 b_{12} \end{aligned} \quad (13.3.13)$$

Assuming we can factorize $\Phi(p)$ such that:

$$\Phi(p) = \Phi^+(p) \Phi^-(p) \quad (13.3.14)$$

where $\Phi^+(p)$ has poles and zeros in the L.H.H. only
 $\Phi^-(p)$ " " " " " " " R.H.H. "

$$\text{then } W_1(p) \Phi^+(p) = \frac{D_{11} b_{22} - D_{21} b_{12}}{\Phi^-(p)} + \frac{H_1 b_{22} - H_2 b_{12}}{\Phi^-(p)} \quad (13.3.15)$$

but $W_1(p) \Phi^+(p)$ has poles in the L.H.H. only if $W_1(p)$ is the transfer function of a stable system (i.e. $W_1(s) = 0 \quad \tau < 0$)

As before, we assume that the order of $W_m(p)$ is chosen so that:

$$\lim_{p \rightarrow \infty} W_m(p) \Phi^+(p) = 0 \quad (13.3.16)$$

for all m

$$\therefore W_1(p) \Phi^+(p) = \left[\frac{D_{11} b_{22} - D_{21} b_{12}}{\Phi^-(p)} \right]_+ + \left[\frac{H_1 b_{22} - H_2 b_{12}}{\Phi^-(p)} \right]_+ \quad (13.3.17)$$

where the suffices indicate that we have taken only those terms of the partial fraction expansion representing poles in the L.H.H.

Since H_1 and H_2 have no poles in the L.H.H., the L.H.H. poles of $\frac{H_1 c_{22} - H_2 c_{12}}{\Phi^-(p)}$ must be poles of both c_{22} and c_{12} .

Their poles are known and we can write

$$\frac{H_1 c_{22} - H_2 c_{12}}{\Phi^-(p)} = \frac{a}{p+d} + \frac{b}{p+\rho} + \frac{c}{p+\delta} + \dots \quad (13.3.18)$$

where a, b, c are undetermined coefficients. $\omega_1(p)$ and $\omega_2(p)$ are substituted in one of the original equations, then a set of simultaneous equations in the undetermined coefficients of both $\omega_1(p)$ and $\omega_2(p)$ is obtained by equating residues of L.H.H. poles of this equation. In most problems it has been found necessary to impose tighter restrictions on the order of $\omega_m(p)$, than those required to satisfy equations (13.3.16), otherwise the undetermined coefficients are too numerous for the number of simultaneous equations.

The two noise filter problems described in chapter 6 were worked out manually, using this method, one of the problems took about 200 hours to solve, the other took only 40 hours approximately. Both of these problems involved only two simultaneous equations, but it is expected that the computational labour increases nearly proportionally to the square of the number of equations. For this reason, a programme for solving these equations has been prepared for EDSAC II, the digital computer at Cambridge University.

13.3.3 Solution in the time domain*

The simultaneous integral equations to be solved are:

$$\sum_{n=1}^R \int_0^{\infty} W_n(\tau_2) C_{nm}(\tau_1 - \tau_2) d\tau_2 = \delta_z D_n(\tau_1) \quad (5.27)$$

$\tau_1 \geq 0$
 $n = 1, 2, \dots, R$

We assume that

$$C_{nm}(\tau) \text{ is given as } \sum_{i=1}^I c_{nmi} e^{-i\alpha|\tau|} \quad (13.3.19)$$

$$\text{and } \delta_z D_n(\tau) \text{ is given as } \sum_{j=1}^J d_{nij} e^{-j\alpha\tau} \quad \text{for } \tau \geq 0 \quad (13.3.20)$$

$$\text{also } W_m(\tau) = h_{m0} \delta(\tau) + h_{m1} e^{-\beta_1 \tau} + h_{m2} e^{-\beta_2 \tau} \text{ etc. for } \tau \geq 0 \quad (13.3.21)$$

Substituting in equations (5.27) gives:

$$\begin{aligned} \sum_{m=1}^R \left\{ \int_0^{\tau_1} \left[h_{m0} \delta(\tau_2) + \sum_{\mu} h_{m\mu} e^{-\beta_{\mu} \tau_2} \right] \left[\sum_{i=1}^{I_{nm}} c_{nmi} e^{-i\alpha(\tau_1 - \tau_2)} \right] d\tau_2 \right. \\ \left. + \int_{\tau_1}^{\infty} \left[\sum_{\mu} h_{m\mu} e^{-\beta_{\mu} \tau_2} \right] \left[\sum_{i=1}^{I_{mn}} c_{mni} e^{-i\alpha(\tau_2 - \tau_1)} \right] d\tau_2 \right\} \\ = \sum_{j=1}^J d_{n1j} e^{-j\alpha\tau_1} \end{aligned} \quad (13.3.22)$$

* This solution was worked out with the assistance of Mr P.J. Lawrence, and it is the method from which the EDSAC programme was prepared by him.

Evaluating the integrals, this becomes:-

$$\sum_{n=1}^R \left\{ h_{n0} \sum_{i=1}^{I_{nm}} c_{nmi} e^{-i\alpha T_1} + \sum_{\mu} h_{n\mu} \left[\sum_{i=1}^{I_{nm}} c_{nmi} \frac{(e^{-(\beta_{\mu}-i\alpha)T_1} - 1)}{(i\alpha - \beta_{\mu})} e^{-i\alpha T_1} + \sum_{i=1}^{I_{nm}} c_{nmi} \frac{e^{-(\beta_{\mu}+i\alpha)T_1}}{\beta_{\mu}+i\alpha} e^{i\alpha T_1} \right] \right\} = \sum_{j=1}^{J_n} d_{nj} e^{-j\alpha T_1} \quad (13.3.23)$$

Simplifying and re-arranging gives: $\beta_{\mu} \neq i\alpha$ $n=1, 2 \dots R$

$$\sum_{n=1}^R \left\{ \sum_{i=1}^{I_{nm}} \left[c_{nmi} \left(h_{n0} + \sum_{\mu} h_{n\mu} \frac{1}{\beta_{\mu} - i\alpha} \right) e^{-i\alpha T_1} \right] + \sum_{\mu} \left[h_{n\mu} \left\{ \sum_{i=1}^{I_{nm}} \left(\frac{c_{nmi}}{i\alpha - \beta_{\mu}} + \frac{c_{nmi}}{\beta_{\mu} + i\alpha} \right) \right\} e^{-\beta_{\mu} T_1} \right] \right\} = \sum_{j=1}^{J_n} d_{nj} e^{-j\alpha T_1} \quad (13.3.24)$$

$n=1, 2 \dots R$

Equating coefficients of exponentials gives: $\beta_{\mu} = i\alpha$ for $e^{-i\alpha T_1}$,

$$\sum_{n=1}^R \left\{ c_{nmi} \left(h_{n0} + \sum_{\mu} h_{n\mu} \frac{1}{\beta_{\mu} - i\alpha} \right) \right\} = d_{ni} \quad (13.3.25)$$

$\beta_{\mu} \neq i\alpha$ $n=1, 2 \dots R$
 $i=1, 2 \dots I_{n_{\max}}$

where $I_{n_{\max}}$ is equal to the largest I_{nm}

for $e^{-\beta_u T_1}$

$$\sum_{m=1}^R h_{mu} \left\{ \sum_{i=1}^{I_{nm}} \left(\frac{C_{nmi}}{i\alpha - \beta_u} + \frac{C_{mni}}{i\alpha + \beta_u} \right) \right\} = 0 \quad (13.3.26)$$

$n = 1, 2, \dots, R$
 $u = 1, 2, \dots, M$ say
~~for each value of u~~

$i\alpha \neq \beta_u \quad \beta_u \neq j\alpha$

For this equation to be satisfied by non-zero h_{mu} 's it is necessary that

$$\text{Det. } \left\{ \sum_{i=1}^{I_{nm}} \left(\frac{C_{nmi}}{i\alpha - \beta_u} + \frac{C_{mni}}{i\alpha + \beta_u} \right) \right\} = 0 \quad (13.3.27)$$

This gives values of β 's say, M of them, with positive real parts. We now substitute these values of β back in equations (13.3.25) and solve for the h_{mu} 's.

This gives us $\sum_{n=1}^R I_n + M \times R$ linear equations for the h_{mu} 's, which would appear to be too many but it is found in the examples solved to date, this is not the case.

The mean-squared error.

To determine the mean-squared error, the following integral has to be evaluated:

$$\sigma_y^2 = \sum_{n=1}^R \sum_{m=1}^R \int_0^\infty \int_0^\infty W_n(T_1) W_m(T_2) C_{nm}(T_1 - T_2) dT_1 dT_2 \quad (5.28)$$

Also, to determine the mean power required to drive the motor and load in the feedback system of section 10.1, a similar integral has to be ^{evaluated} / ~~seen~~ (see equation 10.11).

Substituting for $W_m(T_2)$ and $C_{nm}(T_1 - T_2)$ using equations ^{13.3.19} (13.3.21) gives:

$$\sigma_y^2 = \sum_{n=1}^R \int_0^{\infty} W_n(T_1) \sum_{m=1}^R \left\{ \sum_{i=1}^{I_{nm}} \left[C_{nmi} \left(h_{m0} + \sum_{\mu} h_{m\mu} \frac{1}{\rho_{\mu} - i\alpha} e^{-i\alpha T_1} \right) \right] \right. \\ \left. + \sum_{\mu} \left[h_{m\mu} \left\{ \sum_{i=1}^{I_{nm}} \left(\frac{C_{nmi}}{i\alpha - \rho_{\mu}} + \frac{C_{mni}}{\rho_{\mu} + i\alpha} \right) \right\} e^{-\rho_{\mu} T_1} \right] \right\} \right\} \quad (13.328)$$

Substituting for $W_n(T_1)$ and simplifying the rest of the expression gives:-

$$\sigma_y^2 = \sum_{n=1}^R \int_0^{\infty} \left(h_{n0} \delta(T_1) + \sum_{\nu} h_{n\nu} e^{-\nu\alpha T_1} \right) \\ \times \left(k_n + \sum_{i=1}^{I_{nn}} f_{ni} e^{-i\alpha T_1} + \sum_{\mu} g_{n\mu} e^{-\rho_{\mu} T_1} \right) dT_1 \quad (13.3.29)$$

where $k_n = \sum_{m=1}^R \sum_{i=1}^{I_{nm}} C_{nmi} h_{m0}$ (13.3.30)

$$f_{ni} = \sum_{m=1}^R \sum_{\mu} C_{nmi} h_{m\mu} \frac{1}{\rho_{\mu} - i\alpha} \quad (13.3.31)$$

and
$$g_{n\mu} = \sum_{m=1}^R \sum_{i=1}^{I_{nm}} h_{m\mu} \left(\frac{C_{mni}}{id - \beta_{\mu}} + \frac{C_{mni}}{\beta_{\mu} + id} \right) \quad (13.3.32)$$

$$\begin{aligned} \therefore \sigma_y^2 = & \sum_{n=1}^R \int h_{n0} \left[k_n + \sum_i f_{ni} + \sum_{\mu} g_{n\mu} \right] \\ & + \sum_{\nu} h_{n\nu} \left[\frac{k_n}{\nu d} + \sum_{i=1}^{I_{nm}} \frac{f_{ni}}{(\nu+1)d} + \sum_{\mu} \frac{g_{n\mu}}{\nu d + \beta_{\mu}} \right] \quad (13.3.33) \end{aligned}$$

13.4 Some mathematical notes.

The following discussion on higher order delta functions and bilateral Laplace transforms is ~~not~~ found in ~~many~~ few mathematical text books and is included here for convenience.

13.4.1 Bilateral Laplace transforms.

The bilateral Laplace transform of a function is defined by the following integral:

$$F(p) = \int_{-\infty}^{\infty} f(t) e^{-pt} dt \quad (13.4.1)$$

where p is a complex variable.

It is apparent that this integral, which is a generalization of both the Fourier and unilateral Laplace transforms, may not always be convergent.

Sufficient conditions for convergence are:

(1) $f(t)$ must be bounded and continuous over the range of integration.

(2) There must exist both a positive and a negative finite value of c , however small, such that $f(t)e^{ct}$ converges to zero more rapidly than $\frac{1}{t^n}$ for $n > 1$ as t tends to $\pm\infty$ (we say $f(t)e^{ct}$ converges more rapidly than $\frac{1}{t^n}$ if $\lim_{t \rightarrow \infty} t^n f(t)e^{ct}$ is finite or zero).

The functions which are transformed in Appendix 13.3 all satisfy these conditions except the delta and higher order delta functions which are shown to have Laplace transform in section 13.4.2.

Equation (13.4.1) may be written as follows:

$$F(p) = \int_0^{\infty} f(t)e^{-pt} dt + \int_0^{\infty} f(-t)e^{pt} dt \quad (13.4.2)$$

$$\therefore F(p) = F_1(p) + F_2(p) \quad (13.4.3)$$

where $F_1(p)$ is the ordinary unilateral transform of $f(t)$ and $F_2(p)$ is the unilateral transform of $f(-t)$ with p replaced by $-p$.

e.g. Let
$$f(t) = \begin{cases} e^{-at} & t \geq 0 \\ e^{bt} & t \leq 0 \end{cases}$$

Then

$$F(p) = F_1(p) + F_2(p) = \frac{1}{p+a} + \frac{1}{-p+b} = -\frac{a+b}{(p+a)(p-b)}$$

For an even function of t equation (13.4.2) yields

$$F(p) = F_1(p) + F_1(-p)$$

(13.4.4)

Some properties of bilateral transforms are listed below (for a more detailed discussion of these properties see Balth, Van der Pol and Bremmer^{*}).

1. Poles and zeros of $F(p)$ will be symmetrically placed either side of the real axis of the complex p plane, if $F(p)$ can be expressed as a rational function of p with real coefficients.

2. If $f(t) \equiv 0$; $t < 0$ then all the poles of $F(p)$ lie in the left-hand half of the complex plane (abbreviated L.H.H.).

3. If $f(t) \equiv 0$; $t > 0$ then all the poles $F(p)$ lie in the R.H.H.

* Balth, Van der Pol and Bremmer, "Operational Calculus", 1955, Cambridge University Press.

4. If $f(t)$ is an even function of t then poles and zeros of $F(p)$ are symmetrically placed about the imaginary axis on the complex p plane.

5. If $F(p)$ is a rational function of p with constant coefficients and $f(t)$ is an even function of t we can write

$$F(p) = F^-(p) F^+(p) \quad \left. \vphantom{F(p)} \right\} \quad (13.4.5)$$

where $F^-(p) = F^+(-p)$

The plus and minus indices indicating respectively function with poles and zeros in the L.H.H. only and R.H.H. only.

Transforms of convolution integrals:

$$\text{Let } \phi(\tau) = \int_0^\infty f(t) C(\tau-t) dt$$

Taking the bilateral transform of both sides:

$$\begin{aligned} \Phi(p) &= \int_{-\infty}^{\infty} e^{-p\tau} \int_0^\infty f(t) C(\tau-t) dt d\tau \\ &= \int_0^\infty f(t) e^{-pt} \int_{-\infty}^{\infty} e^{-p(\tau-t)} C(\tau-t) d\tau dt \end{aligned}$$

(by reversing the order of integration)

$$= F_1(p) C(p)$$

(13.4.6)

Let

$$\begin{aligned}\phi(\tau) &= \int_0^\infty \int_0^\infty f(t_1) f(t_2) C(\tau + t_1 - t_2) dt_1 dt_2 \\ &= \int_0^\infty f(t_1) E(\tau + t_1) dt_1\end{aligned}$$

where $E(\tau + t_1) = \int_0^\infty f(t_2) C(\tau + t_1 - t_2) dt_2$

Taking the bilateral transform of both sides:

$$\begin{aligned}\Phi(p) &= \int_{-\infty}^\infty e^{-p\tau} \int_0^\infty f(t_1) E(\tau + t_1) dt_1 d\tau \\ &= \int_0^\infty f(t_1) e^{pt_1} \int_{-\infty}^\infty e^{-p(\tau + t_1)} E(\tau + t_1) dt_1\end{aligned}$$

(by changing the order of integration)

$$= F_1(-p) E(p)$$

(by using equation 13.4.6)

$$= F_1(-p) F_1(p) C(p)$$

(13.4.7)

where $F_1(p)$ is the unilateral transform of $f(t)$ and $C(p)$ is the bilateral transform of $C(\tau)$.

13.4.2 Higher order delta functions.

a. Definitions:

i. We define

$$\delta(t) = \begin{cases} 0 & t \neq 0 \\ \infty & t = 0 \end{cases}$$

such that

$$\int_{-\infty}^{\infty} \delta(t) dt = 1 \quad (13.4.8)$$

ii. We define

$$\delta^{(2)}(t) = \frac{d}{dt} \delta(t)$$

such that

$$\int_{-\infty}^{\infty} \delta^{(2)}(t) dt = 0 \quad (13.4.9)$$

Whilst

$$\int_{-\infty}^t \delta^{(2)}(t) dt = \delta(t) \quad \text{by definition}$$

iii. Similarly we define

$$\delta^{(n)}(t) = \frac{d^n \delta(t)}{dt^n} \quad (13.4.10)$$

b. Integrations:

Consider the integral:

$$I_1 = \int_{-t_1}^{t_2} f(t) \delta(t) dt \quad \begin{matrix} t_1 > 0 \\ t_2 > 0 \end{matrix}$$

$$I_1 = f(t) h(t) \Big|_{-t_1}^{t_2} - \int_{-t_1}^{t_2} f'(t) h(t) dt$$

(by parts)

where

$$h(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases}$$

$$\begin{aligned} \therefore I_1 &= f(t_2) - f(t_2) + f(0) \\ &= f(0) \end{aligned} \quad (13.4.11)$$

Consider next the integral:

$$\begin{aligned} I_2 &= \int_{-t_1}^{t_2} f(t) \delta^{(2)}(t) dt && \begin{matrix} t_1 > 0 \\ t_2 > 0 \end{matrix} \\ &= f(t) \delta(t) \Big|_{-t_1}^{t_2} - \int_{-t_1}^{t_2} f'(t) \delta(t) dt \\ &= -f'(0) \end{aligned} \quad \begin{matrix} \text{(by parts)} \\ (3.4.12) \end{matrix}$$

(by using equation 3.4.11)

Similarly:

$$\int_{-t_1}^{t_2} f(t) \delta^{(3)}(t) dt = f''(0) \quad \begin{matrix} t_1 > 0 \\ t_2 > 0 \end{matrix}$$

and generally:

$$\int_{-t_1}^{t_2} f(t) \delta^{(n)}(t-\tau) dt = \left[(-1)^{n-1} \frac{d^{n-1}}{dt^{n-1}} f(t) \right]_{t=\tau} \quad (13.4.13)$$

In the solution of these integrals it is presupposed that $f(t)$ and its derivatives are well-behaved at $t=\tau$ or $t=0$ as the case may be. For instance, equation (13.4.13) is meaningful only if the $(n-1)$ th derivative is defined at $t=\tau$. If

$\frac{d^{(n-1)} f(t)}{dt^{n-1}}$ should fail to be continuous at $t = \tau$ then special care must be taken to define what we mean by $\frac{d^{(n-1)} f(t)}{dt^{n-1}}$ at $t = \tau$, this can usually be done from a knowledge of the physical problem under investigation.

Equation 13.4.13 may be needed to evaluate the integrals for the mean squared error (equations 5.29 and 5.33) when the optimum weighting functions $W_n(\tau)$ contain higher order delta functions^{as}, for instance, in problem 3, chapter 8. The integrals which have to be evaluated in this problem are of the form

$$\int_0^\infty \delta^{(i)}(\tau) C_{n_i}(\tau + \tau) d\tau$$

Fortunately, $C_{n_i}(\tau + \tau)$ and its derivatives are usually continuous for all values of $\tau + \tau > 0$ so that the difficulty of definition does not arise in this problem.

Equation (13.4.10) may be used also in the analytical evaluation of the various correlation coefficients required for the design of a feedforward compensating network (see chapter 10). The evaluation of $\phi_{g_n g_m}(\tau)$ of case (ii) (equation 10.17), in particular, may involve higher order delta functions because $f'(t)$ is the inverse transform of the inverse of the motor and load transfer function. In this case, higher derivatives of $C_{nm}(\tau)$ must be defined at $\tau = 0$; to overcome this difficulty of

definition the Laplace transforms of equations (10.17) could be used for the evaluation of the required correlation (see equation 13.4.7).

c. Transforms:

$$\begin{aligned} \text{Let } \Delta^{(n)}(\rho) &= \int_{-\infty}^{\infty} e^{-\rho t} \delta^{(n)}(t) dt \\ &= (-1)^{n-1} (-\rho)^{n-1} = \rho^{n-1} \end{aligned}$$

(13.4.14)

(by using equation 13.4.13)

$$\text{e.g. } \Delta(\rho) = 1; \quad \Delta^{(2)}(\rho) = \rho; \quad \Delta^{(3)}(\rho) = \rho^2$$

(13.4.15)

FIG.1 The measurement of the expansion coefficients of a joint second probability density function.

FIG.2 Singleton's delay line filter.

FIG.3 Bose's experimental procedure for the determination of optimum non-linear memory filters.

FIG.4 Synthesis of Bose's optimum non-linear memory filter.

FIG.5 A multipath zero memory filter of degree R .

FIG.6 A multipath memory filter of degree R .

FIG.7 The cross-correlation function is invariant under non-linear transformation of the first variable.

FIG.8 Separability is preserved under linear transformation of the second variable.

FIG.9 Zero memory filter functions for problem 1.

FIG.10 The performance index plotted against numbers of paths for problem 1.

FIG.11 The optimum multipath memory filter for problem 1.

FIG.12 Zero memory filter functions for problem 2.

FIG.13 The performance index plotted against numbers of paths for problem 2.

FIG. 14 Block diagram of the two path memory filter for problem 2.

FIG. 15 Block diagram of the three path memory filter for problem 2.

FIG.16 The zero memory predictor function for problem 3.

FIG.17 The performance index plotted against number of paths for the predictor of problem 3.

FIG.18 Predictor mean-squared error as a function of prediction time T.

A ZERO MEMORY NON-LINEAR SYSTEM.

THE BEST APPROXIMATE LINEAR FILTER.

FIG.19 An equivalent linear filter for a zero memory non-linear system of problem 4.

FIG.20 Linearization of a non-linear feedback loop.

A NON - LINEAR FEEDBACK SYSTEM INCORPORATING
A LIMITER.

THE SIMULATED SYSTEM.

FIG.21 A feedback system incorporating a limiter.

ELECTRONIC ASSOCIATES, INC.
LONG BRANCH, NEW JERSEY

BY _____
DATE 4/8/59

SUBJECT THE ANALYSIS OF
NON-LINEAR FEEDBACK SYSTEMS

SHEET NO. 2 OF 3
PROJ. NO. _____

FIG.22 The simulated system of fig.21 and the measuring circuits.

ELECTRONIC ASSOCIATES, INC.

European Computation Center

43, rue de la Science,

Bruxelles, Belgium.

Tél. 11.43.69

KEY TO PRINCIPAL SYMBOLS (FIG. 23)

NOISE GENERATOR No 1

ELECTRONIC MULTIPLIER No 2

QUARTER SQUARE MULTIPLIER No 4

HIGH GAIN AMPLIFIER No 9

FEEDBACK AMPLIFIER No 34
WITH A GAIN OF 10 INTEGRATING AMPLIFIER No 11
WITH A TIME CONSTANT OF
 $\frac{1}{2}$ SEC.

POTENTIOMETER No 31

TRUNK No 25 CONNECTING THE
A AND B MACHINES

PEN RECORDER CHANNEL No 3

AN ELECTRONIC DIODE

FIG.24 The filter functions for problem 6.

ELECTRONIC ASSOCIATES, INC.
 European Computation Center
 43, rue de la Science,
 Bruxelles, Belgium.
 Tél. 11.43.69

SHEET No 3 of 3

DATE 27/8/59

FIG.25 Simulation of the approximate system.

CHANNEL 6 $10x [y(t) - z(t)]$

100V

CHANNEL 5 $x(t)$

100V

CHANNEL 4 $-u(t)$

40V

CHANNEL 3 $s(t)$

40V

CHANNEL 2 $y(t)$

100V

CHANNEL 6 $10 \times [y(t) - z(t)]$

100V

CHANNEL 5 $x(t)$

100V

CHANNEL 4 $-u(t)$

40V

CHANNEL 3 $s(t)$

40V

CHANNEL 2 $y(t)$

100V

CHANNEL 6 10x [y(t) - z(t)]

CHANNEL 5 2(t)

CHANNEL 4 -u(t)

CHANNEL 3 S(t)

CHANNEL 2 y(t)

CHANNEL 6 $10 \times [y(t) - z(t)]$

CHANNEL 5 $x(t)$

CHANNEL 4 $-u(t)$

CHANNEL 3 $s(t)$

CHANNEL 2 $y(t)$

CHANNEL 6 $10 \times [y(t) - z(t)]$

CHANNEL 5 $x(t)$

CHANNEL 4 $-u(t)$

CHANNEL 3 $s(t)$

CHANNEL 2 $y(t)$

FIG.30 An optimum linear feedback system.

FIG.32 Improvement of the performance of an optimum linear system by the addition of non-linear feedforward compensation.

FIG.32 Suppression of the input $x(t)$ to $H(p)$

FIG. 33 The final system with non-linear feedforward compensation.

FIG. 34 Experimental measurement of various correlation coefficients needed in the design of the optimum compensating network C . (case i).

ELECTRONIC ASSOCIATES, INC.
 European Computation Center
 43, rue de la Science,
 Bruxelles, Belgium.
 Tél. 11.43.69

SHEET No 1 of 3

DATE 27/7/59

FIG.37 The circuit for generating a clipped Gaussian signal and the measurement of its moments.

-33.33

Fig. 38.

SCALE 20

$+10.3600 \theta_5(x)$ $+10.3988 \theta_4(x)$ $-12.072 \theta_3(x)$ $-17.323 \theta_2(x)$ $x(t)$

FIG.40 Measurement of correlation coefficients by Lampard's method.

FIG. 41a Synthesis of the transfer function $\frac{\alpha p^{h-1}}{(p+\alpha)^n}$

FIG. 41b Synthesis of the transfer function $\frac{\alpha/2 (p - \alpha/2)^{h-1}}{(p + \alpha/2)^n}$

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

SAVOY PLACE, LONDON, W.C.2

THE OPTIMIZATION OF A CLASS OF NON-LINEAR FILTERS

By

J. K. LUBBOCK, M.A., B.Sc.(Eng.), Graduate.

MONOGRAPH No. 344 E

November 1959

To be republished in

PART C OF THE PROCEEDINGS OF THE INSTITUTION

The Institution is not, as a body, responsible for the opinions expressed by individual authors

THE OPTIMIZATION OF A CLASS OF NON-LINEAR FILTERS

By J. K. LUBBOCK, M.A., B.Sc.(Eng.), Graduate.

(The paper was first received 23rd February, and in revised form 2nd July, 1959. It was published as an INSTITUTION MONOGRAPH in November, 1959.)

SUMMARY

An optimum linear filter, in the mean-square-error sense, is no better than the optimum attenuator if there is no dissimilarity between the spectral densities of the signal and the noise. However, non-linear filters can use more statistical information about the signal and noise, so that, although they both possess the same spectral densities, a non-linear filter may be able to introduce a significant improvement in the mean-square error (i.e. it can do better than the optimum attenuator). For this reason the use of a non-linear filter in certain circumstances may well justify the greater difficulties encountered in its optimization and physical realization.

The class of filters considered in the paper may be defined by a general expression relating output to input:

$$y(t) = \sum_{r=1}^R \int_0^{\infty} \omega_r(\tau) \theta_r[x(t-\tau)] d\tau$$

An almost routine procedure is proposed whereby the optimum set of weighting functions, $\omega_r(\tau)$, can be determined given either long enough samples of the combined input and the signal or sufficient statistical information about their characteristics. Some worked examples demonstrate that:

(a) A significant improvement in mean-square error is possible even under the condition when the signal and noise possess the same spectral densities.

(b) The class of non-linear filters under consideration can be optimized unhindered by the need to evaluate difficult integrals.

(c) Although the physical complexity of the filter increases rapidly with the value of R , the mean-square error may converge rapidly to an asymptotic value as R is increased; in one example the performance was found to be within about 5% of the asymptotic value with $R = 2$.

(1) INTRODUCTION

In communication and control systems it often happens that information transmitted by the system becomes corrupted by the addition of an unwanted component called noise. The noise may be due to a number of causes such as crosstalk, distortion or random disturbances. The relative magnitude of this effect depends on the system and its input. In some cases it is quite unimportant, in others the noise and signal may be so inextricably combined that the original signal information is veritably 'lost'. As a measure of this effect we shall use the mean-square error:

$$\bar{e}^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [y(t) - z(t)]^2 dt = \overline{[y(t) - z(t)]^2} \quad (1)$$

where $y(t)$ is the system output and $z(t)$ is the desired output, e.g. the signal component of input.

If the input to a system is already corrupted by noise, it may be possible, by correct design, to arrange for the mean-square error at its output to be less than that at its input. Such a system is called a noise filter. It may be necessary for a system to predict future values of the signal from a knowledge of its past values, in which case it is called a predictor. The use of

linear devices for these purposes is well known, but, as shown in the paper, non-linear devices can often give greater reduction in mean-square error.

A filtering device must be able to identify and use some dissimilarity between the statistical character of the noise and signal components of its input. If the statistical descriptions of the signal and noise characteristics are identical in all respects, then no significant improvement in mean-square error is possible, i.e. the optimum* filter becomes an attenuator. Furthermore, if we choose to optimize a filter subject to the restriction that it belongs to a specified class of filters, we also restrict the kind of information which the filter can identify and use. Thus an optimum filter belonging to the linear class of filters can produce a significant improvement in mean-square error, only if the spectral densities of the signal and the noise are different, since it cannot use any other information. Other dissimilarities in the characteristics of the signal and noise which the linear filter has ignored, e.g. amplitude dissimilarities, can be used by non-linear filters; thus a non-linear filter may be able to produce a significant improvement in mean-square error although a linear filter cannot do so.

(2) SOME PHYSICALLY REALIZABLE FILTERS

The following is a classification of some simple filters which can be physically realized and optimized using the theoretical method outlined in the paper.

(2.1) Attenuator

An attenuator is a device† whose output at some instant is equal to the sum of a term which is proportional to the input at the same instant and a d.c. level term. (By this definition we allow that the constant of proportionality may be greater than unity.)

$$\text{i.e.} \quad y(t) = kx(t) + K \quad . \quad . \quad . \quad (2)$$

(2.2) Zero-Memory Filter

The output of a zero-memory filter is given by an instantaneous function of the input. The class of attenuators forms a sub-class of the zero-memory class.

$$y(t) = f[x(t)] = \int_0^{\infty} f[x(t-\tau)] \delta(\tau) d\tau \quad . \quad . \quad . \quad (3)$$

where $\delta(\tau)$ is the Dirac delta function.

(2.3) Linear Filter

The output of a linear filter, in general, can be expressed as

* We define 'optimum filter' as the filter which gives the lowest value of mean-square error at its output subject to the usual restriction that it belongs to a specified class of filters.

† Although mathematically an attenuator is a special type of filter, and it can improve the mean-square error, any reduction in mean-square error that it does produce is not considered to be significant for reasons explained in Appendix 8.3.3.

a continuous linear summation of all past values of input in the form of the convolution integral

$$y(t) = \int_0^{\infty} x(t - \tau) \omega(\tau) d\tau \quad (4)$$

where $\omega(t)$ is the filter-weighting-function or delta-function response and $x(t)$ is the filter input. The class of attenuators forms a sub-class of the linear class.

(2.4) Some Non-Linear Memory Filters

A simple non-linear memory filter can be realized by connecting a zero-memory filter and a linear filter in cascade. Express-

Fig. 1.—Zero-memory filter followed by linear memory filter.

sions relating output to input are given below for several canonical forms of memory filters.

$$y(t) = \int_0^{\infty} f[x(t - \tau)] \omega(\tau) d\tau \quad (5)$$

(see Fig. 1)

$$y(t) = \sum_{r=0}^R \int_0^{\infty} f_r[x(t - \tau)] \omega_r(\tau) d\tau \quad (6)$$

Both eqns. (5) and (6) are special cases of the following equation which defines a class of filters, η_1 :

$$y(t) = \int_0^{\infty} K[x(t - \tau), \tau] d\tau \quad (7)$$

Expanding the kernel function $K[x, \tau]$ in terms of a set of linearly independent or orthogonal functions $\theta_r(\tau)$,

$$y(t) = \sum_{r=0}^{\infty} \int_0^{\infty} \theta_r[x(t - \tau)] \omega_r(\tau) d\tau \quad (8)$$

By using a finite number of terms in the expansion, the filter can be physically realized (Fig. 2) and made to approximate to

Fig. 2.—Multi-path memory filter.

any filter defined by eqn. (7) with any desired accuracy, assuming that the expansion converges. Filters for which the input and output are related by an expression of the form of eqn. (7) are members of class η_1 . The classes of linear and zero-memory filters are both sub-classes of η_1 , since eqns. (3) and (4) are of the same form as eqn. (7). A filter of class η_1 which is realized using

only R parallel paths, each path representing a term of the series expansion, is called a multi-path filter of class η_{1R} . The class η_{1R} is a sub-class of η_{1N} when $N > R$ and η_{1N} is a sub-class of η_1 .

Still more general classes of non-linear memory filters can be defined^{1, 2} by generalizing eqn. (7) and in other ways.^{3, 4}

(3) OPTIMIZATION OF FILTERS OF CLASS η_1

The optimization procedure outlined below applies not only to filters for separating signal from noise with which it has become contaminated but also to the more general case in which the desired output depends on the input or some part of it in some linear or non-linear manner, as exemplified by a predictor or a demodulator. The term 'filter' is used in the general sense to cover all such cases, and, where necessary, the term 'noise filter' is used for the special application of separating signal from noise.

The optimization procedure consists of minimizing the mean-square error defined by eqn. (1):

$$\bar{e}^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [y(t) - z(t)]^2 dt \quad (1)$$

When the filter is optimized as a noise filter, the desired output, $z(t)$, is equal to the signal component of input, and

$$x(t) = g[z(t), v(t)] \quad (9)$$

where $v(t)$ is the noise component of the input and g is a function of both z and v . Two examples will be mentioned:

(a) In transmission of intelligence (by line, radio or other means) noise may be added linearly to the signal (e.g. atmospheric noise, thermal noise, shot noise, crosstalk, etc.); thus

$$x(t) = z(t) + v(t) \quad (10)$$

where usually $z(t)$ and $v(t)$ are statistically independent.

(b) Also in transmission, a signal can become distorted due to non-linearities of the transmission system; in this case noise is added in the form of distortion components; thus

$$x(t) = z(t) + h[z(t)] \quad (11)$$

where h is a function of z , often instantaneous, containing no linear terms. In this case $z(t)$ and $h[z(t)]$ usually possess some correlation.

When the filter is optimized as a predictor $z(t)$ is the desired output of the predictor and is normally expressed as a function of $x(t + T)$:

$$z(t) = f[x(t + T)] \quad (12)$$

or, more commonly,

$$z(t) = x(t + T) \quad (13)$$

For the class of filters η_1 ,

$$y(t) = \int_0^{\infty} K[x(t - \tau), \tau] d\tau \quad (7)$$

and the mean-square error is

$$\bar{e}^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \left\{ \int_0^{\infty} K[x(t - \tau), \tau] d\tau - z(t) \right\}^2 dt$$

Changing the order of integration and expanding the integrand,

$$\begin{aligned} \bar{e}^2 = & \int_0^{\infty} \int_0^{\infty} K[x(t - \tau_1), \tau_1] K[x(t - \tau_2), \tau_2] d\tau_1 d\tau_2 \\ & - 2 \int_0^{\infty} K[x(t - \tau), \tau] z(t) d\tau + \overline{z(t)^2} \quad (14) \end{aligned}$$

If $x(t)$ and $z(t)$ are members of ensembles $X(t)$ and $Z(t)$ which are both ergodic, time averages may be replaced by ensemble averages, giving:*

$$\begin{aligned} \bar{e}^2 = & \int_0^\infty \int_0^\infty \int \int K(x_1, \tau_1) K(x_2, \tau_2) p_2(x_1, x_2; \tau_2 - \tau_1) dx_1 dx_2 d\tau_1 d\tau_2 \\ & - 2 \int_0^\infty \int \int K(x, \tau) z p_2(x, z; \tau) dx dz d\tau \\ & + \bar{z}^2 \end{aligned} \quad (15)$$

where $p_2(x_1, x_2; \tau)$ is the second probability density of the process $X(t)$, and $p_2(x, z; \tau)$ is the joint second probability density of the processes $X(t)$ and $Z(t)$. The minimization of this expression is achieved by using variational calculus.

Let $K(x, \tau)$ be changed by an amount $\epsilon H(x, \tau)$, where ϵ is a small parameter and $H(x, \tau)$ is an arbitrary kernel function subject only to the restriction that it must represent a physically realizable filter, i.e. $H(x, \tau) = 0$ $\tau < 0$.

The variation in $\bar{e}^2$ is given by†

$$\begin{aligned} \Delta \bar{e}^2 = & 2\epsilon \int_0^\infty \int_0^\infty \int \int K(x_2, \tau_2) H(x_1, \tau_1) p_2(x_1, x_2; \tau_1 - \tau_2) dx_1 dx_2 d\tau_1 d\tau_2 \\ & - 2\epsilon \int_0^\infty \int \int H(x, \tau) z p_2(x, z; \tau) dx dz d\tau \\ & + \epsilon^2 \int_0^\infty \int_0^\infty \int \int H(x_1, \tau_1) H(x_2, \tau_2) p_2(x_1, x_2; \tau_1 - \tau_2) dx_1 dx_2 d\tau_1 d\tau_2 \end{aligned} \quad (16)$$

By putting $\partial \Delta \bar{e}^2 / \partial \epsilon = 0$ when $\epsilon = 0$,

$$\begin{aligned} \int_0^\infty \int H(x_1, \tau_1) \left[\int_0^\infty \int K(x_2, \tau_2) p_2(x_1, x_2; \tau_1 - \tau_2) dx_2 d\tau_2 \right. \\ \left. - \int z p_2(x_1, z; \tau_1) dz \right] dx_1 d\tau_1 = 0 \end{aligned} \quad (17)$$

Also from eqn. (16) it can be shown that $\partial^2 \Delta \bar{e}^2 / \partial \epsilon^2$ is always positive, so that the stationary value of $\bar{e}^2$ is a minimum. Using the fundamental lemma of variational calculus,⁵ we obtain from eqn. (17) an integral equation which gives the implicit solution for the kernel function of the optimum non-linear filter:

$$\int_0^\infty \int K(x_2, \tau_2) p_2(x_1, x_2; \tau_1 - \tau_2) dx_2 d\tau_2 = \int z p_2(x_1, z; \tau_1) dz \quad (18)$$

which must be satisfied for $\tau_1 \geq 0$ only, since $H(x_1, \tau_1) = 0$ $\tau_1 < 0$. Thus the information which is required in order to optimize a filter of class η_1 is the second probability density of the input $p_2(x_1, x_2; \tau_1 - \tau_2)$ and the joint second probability density $p_2(x_1, z; \tau_1)$ between the input and desired output.

(3.1) Zero-Memory Filters

The input-output relationship for a zero-memory filter is of the form

$$y(t) = f[x(t)] = \int_0^\infty f[x(t - \tau)] \delta(\tau) d\tau \quad (3)$$

* Throughout the paper, (i) for each integral written without limits, it is implied that integration takes place over the whole range of the variable of integration, and (ii) all processes are assumed to be ergodic.

† The first term in the expression for $\Delta \bar{e}^2$ is the sum of two terms which can be combined, since $p_2(x_1, x_2; \tau_2 - \tau_1)$ is symmetrical in the pairs of variables (x_1, τ_1) and (x_2, τ_2) , i.e. $p_2(x_1, x_2; \tau_2 - \tau_1) = p_2(x_2, x_1; \tau_1 - \tau_2)$.

Eqn. (3) shows that zero-memory filters are members of the more general class η_1 and the kernel function in eqn. (7) is of the form

$$K[x(t - \tau), \tau] = f[x(t - \tau)] \delta(\tau)$$

Substitution in eqn. (17) for the kernel functions $K(x_2, \tau_2) = f(x_2) \delta(\tau_2)$ and $H(x_1, \tau_1) = h(x_1) \delta(\tau_1)$ and subsequent integration with respect to τ_1 and τ_2 leads to the equation

$$\int h(x_1) \left[\int f(x_2) p_2(x_1, x_2; 0) dx_2 - \int z p_2(x_1, z; 0) dz \right] dx_1 = 0$$

whence, using the lemma of variational calculus,

$$\int f(x_2) p_2(x_1, x_2; 0) dx_2 = \int z p_2(x_1, z; 0) dz \quad (19)$$

But

$$p_2(x_1, x_2; 0) = p(x_1) \delta(x_2 - x_1)$$

Therefore eqn. (19) can be rewritten

$$f(x) p(x) = \int z p_2(x, z) dz$$

giving an explicit solution for $f(x)$, namely

$$f(x) = \frac{\int z p_2(x, z) dz}{p(x)} \quad (20)$$

where $p_2(x, z) \delta x \delta z$ is the probability that x lies in the range x to $x + \delta x$ and z lies in the range z to $z + \delta z$ simultaneously.

Thus the optimum filter function $f(x)$ is equal to the conditional mean value of z , given x . The information needed in the optimization of a zero-memory filter is the whole information content of the first probability density function of the input and the second joint probability density function of the input and desired output with $\tau = 0$.

(3.2) Multi-Path Zero-Memory Filters

For many processes it is possible to obtain an expansion for $p_2(x, z; \tau)$ in the form [see Appendix 8.1, eqn. (58)]

$$p_2(x, z; \tau) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} D_{nm}(\tau) \theta_n(x) \gamma_m(z) p(x) q(z) \quad (58)$$

where $\theta_n(x)$ and $\gamma_m(z)$ are two sets of polynomials which are orthonormal with respect to $p(x)$ and $q(z)$ as weighting functions, over the whole range of x and z , respectively, and $p(x)$ and $q(z)$ are first probability density functions of x and z , respectively; also

$$D_{nm}(\tau) = \iint \theta_n(x) \gamma_m(z) p_2(x, z; \tau) dx dz \quad (59)$$

whence, by putting $\tau = 0$ and substituting in eqn. (20),

$$f(x) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} D_{nm} \theta_n(x) \int z \gamma_m(z) q(z) dz$$

But

$$\begin{aligned} \int z \gamma_m(z) q(z) dz &= 0 & m \neq 0; m \neq 1 \\ &= \sigma z & m = 1 \\ &= \bar{z} & m = 0 \end{aligned}$$

owing to the orthogonality of $\gamma_m(z)$.

$$\text{Therefore} \quad f(x) = \sum_{n=1}^{\infty} D_{n1} \theta_n(x) \sigma z + \bar{z} \quad (21)$$

since, as shown in Appendix 8.1, $D_{00} = 1$, $\theta_0(x) = 1$ and $D_{0n} = D_{n0} = 0$, $n \neq 0$.

It is shown in Appendix 8.3 that truncation of eqn. (21) at the R th term yields the optimum zero-memory multi-path filter of degree R . Therefore it can be synthesized approximately as shown in Fig. 3.

Fig. 3.—Multi-path zero-memory filter of degree R .

The following recurrence relations, derived in Appendix 8.2, show that the polynomials $\theta_n(x)$ can be determined from a knowledge of the set of moments of x up to and including the $2n$ th, and the coefficients $\sigma_z D_{n1}$ may be determined from a knowledge of the moments of x and the averaged cross-products between z and x^r for $r = 1, 2, \dots, n$:

$$\theta_n(x) = \frac{x^n - \sum_{r=0}^{n-1} x^n \overline{\theta_r(x)} \theta_r(x)}{\left[x^{2n} - \sum_{r=0}^{n-1} x^n \overline{\theta_r(x)}^2 \right]^{1/2}} \quad (80)$$

and, from eqns. (79), (81) and (82), putting $\tau = 0$,

$$\sigma_z D_{n1} = \overline{\theta_n(x)z} = \frac{\overline{x^n z} - \sum_{r=0}^{n-1} \overline{x^n \theta_r(x)} D_{r1} \sigma_z}{\left[\overline{x^{2n}} - \sum_{r=0}^{n-1} \overline{x^n \theta_r(x)}^2 \right]^{1/2}} \quad (22)$$

A finite number of coefficients D_{n1} [in eqn. (21)] and constants σ_z and $\bar{z}$ can be determined experimentally by analogue measurements on long samples of the input, $x(t)$, and desired output, $z(t)$. The required computing equipment consists of:

(a) A set of non-linear power-law-function generators, namely

$$x, x^2, x^3, x^4, \dots, x^{2n}$$

(A convenient device which will generate all of these functions simultaneously is a servo-type multiplier with $(2n + 1)$ potentiometers connected to its shaft.)

(b) An integrator.

(c) A set of variable attenuators.

(d) A set of adders.

(e) A multiplier.

The procedure is as follows:

(i) Measure the set of moments (up to the $2n$ th) of the input process, $X(t)$, by connecting the signal, $x(t)$, to the inputs of the function generators and averaging their outputs in turn using the integrator.

(ii) Determine the appropriate orthonormal polynomials up to the n th using the recurrence equation [eqn. (80)].

(iii) Measure the set of averaged cross-products $\overline{x^r z}$ by averaging over time using the power-law generators, the multiplier and integrator; then compute the coefficients from the recurrence equation [eqn. (22)].

Alternatively, set up the first n polynomial-function generators,

$\theta_n(x)$, using the power-law generators, the attenuators and adders and measure the coefficients $\sigma_z D_{n1}$ directly using the multiplier and integrator. For an ergodic process we can express $\sigma_z D_{n1}$ as a time average:

$$\begin{aligned} \sigma_z D_{n1} &= \iint \theta_n(x) z p_2(x, z; 0) dx dz \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \theta_n[x(t)] z(t) dt \end{aligned}$$

A block diagram of the measuring circuit is shown in Fig. 4.

Fig. 4.—Experimental method of measuring the constants D_{n1} .

(3.3) Multi-Path Non-Linear Memory Filters

The expression for the output of the multi-path filter of Fig. 2 can be obtained by substituting in eqn. (7) an expansion for the kernel function:

$$K[x(t - \tau), \tau] = \sum_{r=0}^{\infty} \omega_r(\tau) \theta_r[x(t - \tau)] \quad (23)$$

Thus by substituting for the kernel function

$$K[x_2, \tau_2] = \sum_{r=0}^{\infty} \omega_r(\tau_2) \theta_r(x_2) \quad (24)$$

in eqn. (18) for the optimum non-linear filter of class η_1 , and at the same time substituting series expansions for $p_2(x_1, x_2; \tau)$ and $p_2(x_1, z; \tau)$ (see Appendix 8.1 and 8.2), namely

$$p_2(x_1, x_2; \tau) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} C_{nm}(\tau) \theta_n(x_1) \theta_m(x_2) p(x_1) p(x_2) \quad (83)$$

$$p_2(x_1, z; \tau) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} D_{kl}(\tau) \theta_k(x_1) \gamma_l(z) p(x_1) q(z) \quad (58)$$

we obtain

$$\begin{aligned} &\sum_{r=0}^{\infty} \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \theta_n(x_1) p(x_1) \times \\ &\int_0^{\infty} C_{nm}[(\tau_1 - \tau_2)] \omega_r(\tau_2) d\tau_2 \int \theta_m(x_2) \theta_r(x_2) p(x_2) dx_2 \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} D_{kl}(\tau_1) \theta_k(x_1) p(x_1) \int z \gamma_l(z) q(z) dz \quad (25) \end{aligned}$$

$$\tau_1 \geq 0$$

$$\text{But } \int \theta_m(x_2) \theta_r(x_2) p(x_2) dx_2 = 1 \quad m = r$$

$$= 0 \quad m \neq r$$

$$\text{and } \int z \gamma_l(z) q(z) dz = 0 \quad l \neq 0; l \neq 1$$

$$= \sigma_z \quad l = 1$$

$$= \bar{z} \quad l = 0$$

Using these results and equating coefficients of $\theta_n(x_1)$ for $0 \leq n \leq R$ in eqn. (25),

$$\sum_{m=1}^R \int_0^\infty C_{nm}(\tau_1 - \tau_2) \omega_m(\tau_2) d\tau_2 = D_{n1}(\tau_1) \sigma_z \quad (26)$$

$$\tau_1 \geq 0 \quad n = 1, 2, 3 \dots R$$

and when $m = n = 0$

$$\int_0^\infty \omega_0(\tau_2) d\tau_2 = \bar{z} \quad (27)$$

In general, the optimum kernel function can be expanded only as an infinite expansion [eqn. (24)], but in order that a solution for the optimum weighting functions $\omega_m(\tau)$ can be found, truncation of eqn. (24) is necessary, thereby yielding a finite number, R , of simultaneous integral equations [eqn. (26)], in which, in general, the weighting functions $\omega_m(\tau)$ for $m \leq R$ will depend on R .

It is shown in Appendix 8.3 that the weighting functions given by eqns. (26) are those of the optimum multi-path memory filter of degree R .

The simultaneous integral equations [eqn. (26)] can be solved by the method of undetermined coefficients⁶ (see Appendix 8.4). It seems that often there is a unique set of solutions, provided that additional restrictions on the weighting functions are imposed to ensure physical realizability in a sufficiently strict sense (see Appendix 8.4). However, if a set of solutions to eqns. (26) is not unique, then, at least, it can be shown to give a mean-square error which is no greater than for any other set (Appendix 8.3).

The multi-path filter of class η_{1R} can only approximate to a general filter of class η_1 , because for physical realizability the number of paths, R , must be finite. For a large value of R more information is required to optimize the filter until in the limit as R tends to infinity the whole information content of $p_2(x_1, x_2; \tau)$ and a large part of the information in $p_2(x, z; \tau)$ is required, i.e. all the coefficients $C_{nm}(\tau)$ in the infinite expansion of $p_2(x_1, x_2; \tau)$ and $D_{n1}(\tau)$ in the infinite expansion of $p_2(x, z; \tau)$.

The coefficients $C_{nm}(\tau)$ and $\sigma_z D_{n1}(\tau)$ can be determined analytically as described in Appendix 8.2 given sufficient statistical information concerning the input and desired output, or they can be determined experimentally by making analogue measurements on long samples of the input, $x(t)$, and desired output, $z(t)$. The required computing equipment is the same as that listed in Section 3.2 except that a cross-correlator is required in place of the multiplier. A very suitable cross-correlator for this purpose is described by Lampard.⁷ The procedure for making the measurements is as follows:

(a) Measure the set of moments of the input signal, $x(t)$, up to and including the $2n$ th moment, using the power-law-function generators and the integrator.

(b) Determine the appropriate orthonormal polynomials up to the n th using eqn. (80).

(c) Set up the first n polynomial-function generators, $\theta_n(x)$, using the power-law generators, the attenuators and adders.

(d) Measure the auto- and cross-correlation functions between the outputs of the function generators taking two at a time.

These functions will be the coefficients $C_{nm}(\tau)$, since, from eqn. (84) for an ergodic process,

$$C_{nm}(\tau) = \iint \theta_n(x_1) \theta_m(x_2) p_{xx}(x_1, x_2; \tau) dx_1 dx_2$$

$$= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \theta_n[x(t)] \theta_m[x(t + \tau)] dt \quad (28)$$

(e) Measure the cross-correlation function between each output of the function generators and the desired filter output, $z(t)$. These correlation functions will be the coefficients $\sigma_z D_{n1}(\tau)$, since, from eqn. (59) for an ergodic process,

$$\sigma_z D_{n1}(\tau) = \iint \theta_n(x) z p_{xz}(x, z; \tau) dx dz$$

$$= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \theta_n[x(t)] z(t + \tau) dt \quad (29)$$

$$n \neq 0$$

The measuring circuit is shown in Fig. 5.

Fig. 5.—Experimental method of measuring the coefficients $C_{nm}(\tau)$ and $D_{n1}(\tau)$.

Although for many cases of input and desired output a non-linear filter may be capable of giving considerably smaller mean-square filtering error than can an optimum linear filter, there are situations in which the performance of an optimum non-linear filter is no better than the optimum linear filter.

$$\text{If* } \left. \begin{aligned} C_{nm}(\tau) &\equiv 0 && \text{when } n = 1 \quad m \neq 1 \\ &&& \text{and } m = 1 \quad n \neq 1 \\ D_{nm}(\tau) &\equiv 0 && \text{when } m = 1 \quad n \neq 1 \\ &&& \tau \geq 0 \end{aligned} \right\} \quad (30)$$

eqn. (26) can be satisfied by putting

$$\omega_m(\tau) \equiv 0 \quad m = 2, 3, \dots \infty$$

$$\text{and } \int_0^\infty C_{11}(\tau_1 - \tau_2) \omega_1(\tau_2) d\tau_2 = D_{11}(\tau_1) \sigma_z \quad \tau_1 \geq 0 \quad (31)$$

Therefore $\omega_1(\tau_2)$ is given by the solution of the Wiener-Hopf integral equation [eqn. (31)], in which

$$\left. \begin{aligned} C_{11}(\tau) &= \rho_{xx}(\tau) \\ D_{11}(\tau) &= \rho_{xz}(\tau) \end{aligned} \right\} \quad (32)$$

* These conditions define a class of probability density functions $p_2(x_1, x_2; \tau)$ and $p_2(x, z, \tau)$. This class is not empty and some of its properties are discussed in References 8, 9 and 10.

are the normalized correlation coefficients. Although we do not prove that this solution is unique, it does give a value of mean-square error which is as small as for any other solution (Appendix 8.3). Thus the optimum filter of class η_1 can be a linear filter, and the only information which is needed for its optimization is the auto-correlation of the input and the cross-correlation of the input and desired output.

(4) SOME NOISE-FILTER EXAMPLES

Two examples are worked out to illustrate the method. In the first example the signal and noise can have the same spectral density, but since only a zero-memory filter is optimized the spectral density (or auto-correlation coefficient) is not required; for the case in question it is found possible to determine the exact zero-memory-filter function from eqn. (20), and for comparison the multi-path zero memory filter is also derived, so that some indication of the degree of approximation is obtained. In the second example, a multi-path memory filter is optimized for the condition when the signal and noise have the same spectral density, a significant improvement in mean-square error being achieved.

Whilst examples are worked out from statistical data, it may happen that not enough information about the statistical properties of the signal and noise is available to enable the filter to be optimized; however, if long enough samples of signal and noise are available, then, using the analogue measuring technique described earlier, sufficient information can be obtained in a suitable form to allow direct optimization of a non-linear filter of class η_{1R} . Whilst there may be practical difficulties involved in obtaining sufficient accuracy of measurement, it is considered that a computer of comparable accuracy to that of Pace* should be adequate to enable these analogue measurements to be made.

Under certain circumstances it may be possible to find a simpler and better non-linear filter belonging to some class other than η_1 , in which event it would obviously not be desirable to use the optimization procedure proposed here. Although the proposed method does not always produce the simplest solution, it can be used for almost any condition of input and desired output and will always produce a physically realizable filter which is better than the best linear filter, except in the special circumstances [eqn. (30)] when the optimum filter of class η_1 becomes linear.

In all the following examples we shall optimize no more than the first five paths of the multi-path filter.

For the first example we take an input consisting of a linear addition of signal and noise, where the signal is a random square wave which can have two values $+K$ or $-K$, each occurring with equal probability (i.e. the signal has zero mean and variance K^2).

$$p_z(z) = \frac{1}{2}[\delta(z - K) + \delta(z + K)] \quad (33)$$

The noise is Gaussian with variance σ^2 and zero mean.

$$p_v(v) = \frac{1}{\sqrt{(2\pi)\sigma^2}} e^{-\frac{v^2}{2\sigma^2}} \quad (34)$$

The optimum zero-memory-filter function is found first and then the multi-path zero-memory filter is optimized and the results are compared.

Since the noise and signal are independent and added linearly,

$$\begin{aligned} p_2(x, z) &= p_v(x - z)p_z(z) \\ &= \frac{1}{2\sqrt{(2\pi)\sigma^2}} [\delta(z - K) + \delta(z + K)] e^{-\frac{(x-z)^2}{2\sigma^2}} \end{aligned}$$

* Precision Analogue Computing Equipment.

The optimum zero-memory filter is given by eqn. (20); hence

$$\begin{aligned} f(x) &= \frac{\int_{-K}^K \frac{1}{2\sqrt{(2\pi)\sigma^2}} z [\delta(z - K) + \delta(z + K)] e^{-\frac{(x-z)^2}{2\sigma^2}} dz}{\int_{-K}^K \frac{1}{2\sqrt{(2\pi)\sigma^2}} [\delta(z - K) + \delta(z + K)] e^{-\frac{(x-z)^2}{2\sigma^2}} dz} \\ &= K \frac{e^{-\frac{(x-K)^2}{2\sigma^2}} - e^{-\frac{(x+K)^2}{2\sigma^2}}}{e^{-\frac{(x-K)^2}{2\sigma^2}} + e^{-\frac{(x+K)^2}{2\sigma^2}}} \\ &= K \tanh \frac{Kx}{\sigma^2} \quad (35) \end{aligned}$$

As K/σ increases the error decreases. For small values of K/σ there is a high probability that x will be negative for a positive value of z , such a condition giving rise to large error (see Fig. 6).

Fig. 6.—Graphs of $K \tanh Kx/\sigma^2$ and the optimum multi-path zero-memory-filter function for three different values of K/σ .

$$K^2 + \sigma^2 = 1$$

The multi-path filter may be optimized as follows:

The first moments of a Gaussian signal with zero mean and variance σ^2 are given by

$$\left. \begin{aligned} \bar{v}^n &= \sigma^n \times 1 \times 3 \times 5 \times 7 \times \dots (n-1) & n \text{ even} \\ \bar{v}^n &= 0 & n \text{ odd} \end{aligned} \right\} \quad (36)$$

The first moments of a random square wave with amplitude K are $\bar{z}^n = K^n$ for n even and $\bar{z}^n = 0$ for n odd. For simplicity we will let $K^2 = \sigma^2 = \frac{1}{2}$ so that the variance of their sum is unity. Using eqn. (73) (Appendix 8.2) the moments are found from $x(t) = z(t) + v(t)$ to be

$$\bar{x}^2 = 1 \quad \bar{x}^4 = 2.5 \quad \bar{x}^6 = 9.5 \quad \bar{x}^8 = 47.75 \quad \bar{x}^{10} = 296.75 \quad (37)$$

whereas odd moments are zero.

Using the recurrence formula 80 [eqn. (80), Appendix 8.2] the polynomials are found to be

$$\left. \begin{aligned} \theta_0(x) &= 1 \\ \theta_1(x) &= x \\ \theta_2(x) &= \frac{x^2 - 1}{\left(\frac{3}{2}\right)^{1/2}} \\ \theta_3(x) &= \frac{x^3 - 2.5x}{\left(\frac{13}{4}\right)^{1/2}} \\ \theta_4(x) &= \frac{x^4 - \frac{14}{3}x^2 + \frac{13}{6}}{\left(\frac{53}{6}\right)^{1/2}} \\ \theta_5(x) &= \frac{x^5 - \frac{96}{13}x + \frac{233}{26}}{\left(\frac{1522}{52}\right)^{1/2}} \end{aligned} \right\} \dots (38)$$

The coefficients D_{n1} are then found using eqn. (81) in Appendix 8.2 for $n = 1, 3$ and 5 :

$$D_{11} = \frac{1}{\sqrt{2}} \quad D_{31} = -\frac{1}{\sqrt{26}} \quad D_{51} = \frac{9}{\sqrt{13}\sqrt{761}} \quad (39)$$

To compare this result with that obtained earlier [eqn. (35)] the polynomial represented by the expression

$$y = [D_{11}\theta_1(x) + D_{31}\theta_3(x) + D_{51}\theta_5(x)]\sigma_z$$

is worked out and plotted on the same axes as the curve $y = K \tanh Kx/\sigma^2$ when $K = 1/\sqrt{2}$ and $\sigma^2 = \frac{1}{2}$ (see Fig. 6). For these values of K and σ^2 ,

$$y = 0.7983x - 0.1643x^3 + 0.01183x^5 \quad (40)$$

Since in this case $K/\sigma (= 1)$ is relatively small, the optimum zero-memory filter is far from perfect. The performance index (Appendix 8.3) has been worked out for the multi-path zero-memory filter for 1, 3 and 5 paths (Fig. 7) using eqns. (105) and (113).

The curve of Fig. 7 suggests that an increase in the number of

Fig. 7.—Performance index for the first problem.

paths would not reduce the performance index below about 0.90 when $K/\sigma = 1$. Fig. 7 also shows the performance index plotted for two other values of K/σ . The polynomial filter functions for these cases are given below and plotted in Fig. 6:

$$K/\sigma = 2 \quad y = 1.5530x - 0.6411x^3 + 0.08759x^5 \quad (41)$$

$$K/\sigma = 4 \quad y = 1.8377x - 1.0716x^3 + 0.2378x^5 \quad (42)$$

It is to be expected that very much better performance will be obtained for larger values of K/σ .

This first example is almost trivial because the exact optimum zero-memory filter can be found using eqn. (20), and hence there is no advantage in optimizing the multi-path zero-memory filter in this case. The problem has been worked out primarily to give some indication of the nearness of the approximation (see Fig. 6) obtained by using only a few paths.

For the second case of signal and noise the optimum multi-path filters, both with and without memory, will be found. Consider a noise $v(t)$ which is Gaussian (of zero mean and unit variance) and a signal which is obtained by passing Gaussian noise through a non-linear-function generator of the form

$$z(t) = \frac{H_2[s(t)]}{\sqrt{2}} = \frac{s^2 - 1}{\sqrt{2}} \quad (125)$$

where $s(t)$ is Gaussian noise of zero mean and unit variance but uncorrelated with $v(t)$, and H_2 is the second-order Hermite polynomial.

From eqn. (125) it can be shown that $z(t)$ has zero mean and unit variance and that if the correlation coefficient of $s(t)$ is exponential so also is that of $z(t)$.* The reason for choosing this method of generating the signal lies in the latter property of non-linear Hermite-polynomial generators with respect to a Gaussian input. Since the correlation coefficient of $z(t)$ is both known and simple, we can choose the correlation coefficient of $v(t)$ to be equal to that of $z(t)$, with the object of making a comparison between the optimum non-linear filter and the optimum linear filter, which for this case becomes a simple attenuator. Unfortunately, optimization of a zero-memory filter using eqn. (20) leads to integrals which are not easy to evaluate (except by numerical methods); however, this difficulty does not prohibit us from finding the optimum multi-path zero-memory filter by the method outlined in Appendix 8.2.

The moments of Gaussian noise, $v(t)$, with zero-mean and unit variance, are

$$\left. \begin{aligned} \overline{v^n} &= 1 \times 3 \times 5 \times \dots (n-1) & n \text{ even} \\ &= 0 & n \text{ odd} \end{aligned} \right\} \quad (36)$$

and the moments of $z(t)$ (as calculated in Appendix 8.5) are

$$\left. \begin{aligned} \overline{z} &= 0 & \overline{z^2} &= 1 & \overline{z^3} &= 2.828427 & \overline{z^4} &= 15 \\ \overline{z^5} &= 96.166522 & \overline{z^6} &= 755 & \overline{z^7} &= 6983.386571 \\ \overline{z^8} &= 74417 & \overline{z^9} &= 897799.3379 & \overline{z^{10}} &= 12096873 \end{aligned} \right\} \quad (43)$$

The moments of $x(t)$ are obtained, given†

$$x(t) = \frac{v(t) + z(t)}{\sqrt{2}} \quad (44)$$

by using eqn. (73), Appendix 8.2:

$$\left. \begin{aligned} \overline{x} &= 0 & \overline{x^2} &= 1 & \overline{x^3} &= 1 & \overline{x^4} &= 6 & \overline{x^5} &= 22 & \overline{x^6} &= 130 \\ \overline{x^7} &= 822 & \overline{x^8} &= 6202 & \overline{x^9} &= 52552 & \overline{x^{10}} &= 499194 \end{aligned} \right\} \quad (45)$$

* Both the second probability density functions of the signal and noise are of class Λ . (See Appendix 8.1 and 8.5, and Reference 11.)

† $\sqrt{2}$ is introduced here for convenience so that $\overline{x^2} = 1$.

The polynomials may then be constructed using the recurrence formula [eqn. (80), Appendix 8.2] giving

$$\left. \begin{aligned} \theta_0(x) &= 1 \\ \theta_1(x) &= x \\ \theta_2(x) &= \frac{1}{2}(x^2 - x - 1) \\ \theta_3(x) &= \frac{1}{6.0622}(x^3 - 3.75x^2 - 2.25x + 2.75) \\ \theta_4(x) &= \frac{1}{24.6469}(x^4 - 8.2041x^3 + 5.2653x^2 \\ &\quad + 21.9592x - 3.0612) \\ \theta_5(x) &= \frac{1}{125.0755}(x^5 - 14.4628x^4 + 40.0826x^3 \\ &\quad + 50.9917x^2 - 103.3058x - 26.2975) \end{aligned} \right\} \quad (46)$$

The coefficients, D_{n1} , are obtained using the recurrence formula [eqn. (81), Appendix 8.2]:

$$\left. \begin{aligned} D_{11} &= 0.7071 \\ D_{21} &= 0.3536 \\ D_{31} &= -0.0875 \\ D_{41} &= -0.0386 \\ D_{51} &= 0.0617 \end{aligned} \right\} \quad (47)$$

The final polynomial then becomes

$$y = D_{11}\theta_1(x) + D_{21}\theta_2(x) + \dots + D_{51}\theta_5(x)$$

$$\text{Therefore } y = 0.0005x^5 - 0.0087x^4 + 0.0182x^3 \\ + 0.2478x^2 + 0.4774x - 0.2246 \quad (48)$$

This function has been plotted in Fig. 8. The optimum multi-path memory filter can be found by calculating the set of cross- and auto-correlation coefficients between pairs of all the outputs of the non-linear function generators, $\theta_n(x)$, and the desired output, $z(t)$. The method outlined in Appendix 8.2, although of routine nature, does take a considerable time. The results are quoted here for a maximum of only three polynomial generators.*

$$\left. \begin{aligned} C_{11}(\tau) &= \rho(\tau) \quad C_{1n}(\tau) \equiv C_{n1}(\tau) \equiv 0 \\ C_{22}(\tau) &= 0.25[\rho(\tau) + 3\rho^2(\tau)] \\ C_{23}(\tau) = C_{32}(\tau) &= \frac{0.75}{12.1244}[-\rho(\tau) + \rho^2(\tau)] \\ C_{24}(\tau) = C_{42}(\tau) &= \frac{-1.3460\rho(\tau) + 1.3460\rho^2(\tau)}{49.25} \\ C_{33}(\tau) &= \frac{0.5625\rho(\tau) + 15.1875\rho^2(\tau) + 21\rho^3(\tau)}{36.75} \\ C_{34}(\tau) = C_{43}(\tau) &= \frac{1.0088\rho(\tau) - 26.7522\rho^2(\tau) + 25.6761\rho^3(\tau)}{149.2817} \\ C_{44}(\tau) &= \frac{1.8097\rho(\tau) + 71.1243\rho^2(\tau) + 252.5364\rho^3(\tau) + 282.0875\rho^4(\tau)}{606.3906} \end{aligned} \right\} \quad (49)$$

$$\left. \begin{aligned} D_{11}(\tau) &= 0.7071\rho(\tau) \\ D_{21}(\tau) &= 0.3536\rho(\tau) \\ D_{31}(\tau) &= -0.0875\rho(\tau) \\ D_{41}(\tau) &= -0.0386\rho(\tau) \\ D_{51}(\tau) &= 0.0617\rho(\tau) \end{aligned} \right\} \quad (50)$$

* The input signal, $x(t)$, has a second probability density function for which $C_{1n}(\tau) \equiv C_{n1}(\tau) = 0$ [see eqn. (30)] although $D_{n1}(\tau) \neq 0$, so that the optimum filter is not linear, but the integral equation for $\omega_1(\tau)$ is independent of the rest of the integral equations.

where the input auto-correlation coefficient $\rho(\tau) = e^{-\beta|\tau|}$. Eqns. (26) are then solved by the method of undetermined coefficients (Appendix 8.4) to give the weighting functions $\omega_n(\tau)$.

For only two parallel paths (it so happens that the two integral equations are independent in this case),

$$\left. \begin{aligned} \omega_1(\tau) &= 0.7071\delta(\tau) \\ \omega_2(\tau) &= 0.2761\delta(\tau) + 0.2222\beta e^{-1.1952\beta\tau} \quad \tau \geq 0 \\ &= 0 \quad \tau < 0 \end{aligned} \right\} \quad (51)$$

Fig. 8.—Zero-memory-filter function for the second problem.

Fig. 9.—Optimum two-path memory filter for the second problem.

For three parallel paths,

$$\left. \begin{aligned} \omega_1(\tau) &= 0.7071\delta(\tau) \\ \omega(\tau) &= 0.2766\delta(\tau) + 0.2195\beta e^{-1.2128\beta\tau} \\ &\quad + 0.0008\beta e^{-2.3717\beta\tau} \quad \tau \geq 0 \\ &= 0 \quad \tau < 0 \\ \omega_3(\tau) &= -0.0624\delta(\tau) - 0.0758\beta e^{-1.2128\beta\tau} \\ &\quad - 0.0126\beta e^{-2.3717\beta\tau} \quad \tau \geq 0 \\ &= 0 \quad \tau < 0 \end{aligned} \right\} \quad (52)$$

The block diagrams of these filters are shown in Figs. 9 and 10.

The performance index is calculated (Fig. 11) using eqns. (105) and (113) for the zero-memory, filter and eqns. (110) and (113) for the memory filter.

Fig. 10.—Optimum three-path memory filter for the second problem.

Fig. 11.—Performance index for the second problem.

The curve of Fig. 11 suggests that the performance index cannot be reduced below about 0.72 using a zero-memory filter or 0.70 using a memory filter. These results, whilst not perfect, should be compared with the figure of unity for the best linear filter. It is interesting to note that, by using only one zero-memory non-linear filter path, $\theta_2(x)$, considerable improvement is achieved as compared with the linear filter, and little additional improvement is effected by increasing the number of paths or by using a memory filter. Thus, in this example it seems that fairly rapid convergence to the optimum filter of class η_1 is obtained; this condition, whilst common, may not necessarily occur with other types of input.

(5) CONCLUSIONS

A routine procedure, unhindered by the need to evaluate difficult integrals, is proposed for the optimization of a class of memory filters which are realized as a number of parallel paths. The physical complexity of the filter increases rapidly with the number of paths, but for many applications the performance index of the filter converges rapidly to a constant value as the

number of paths increases. Although the analytical optimization is tedious, applications can be visualized where the labour and complexity might be justified; in particular, it has been demonstrated that, if signal and noise have the same spectral density, although a linear noise filter cannot produce a significant improvement in mean-square error, a multi-path noise filter can do so.

Multi-path zero-memory filters also can be optimized using this procedure, which has advantages over the more direct method, since, although an optimum zero-memory filter function can be expressed explicitly in terms of probability functions [eqn. (20)], this expression involves integrals which often may be difficult to evaluate. In contrast to the labour involved in the optimization of the multi-path memory filter, a multi-path zero-memory filter can be optimized very easily analytically (taking about four hours for an average problem).

It is probable, in practice, that insufficient information about the input will be available and it will be necessary to obtain this information from long samples of the input and desired output. A method is proposed whereby the information can be obtained in the required form by making some measurements using an analogue computer. Although no results have been obtained in this way, no new experimental techniques are involved and the only essential requirement is a sufficiently large and accurate analogue machine.

It may be necessary to employ a digital computer as well for the solution of the set of integral equations [eqn. (26)]. The example in the text was calculated laboriously by hand, but there seems no reason why a digital machine could not be programmed to do the same task in the same way; if it should be required to solve more than two simultaneous integral equations, a digital computer would be almost indispensable.

Multi-path memory filters may be connected in cascade to produce still lower mean-square error. If an analogue computer is available, a second multi-path memory filter can be optimized to operate on the output of an earlier filter; the mean-square error generally will be reduced, because the combined filter is of higher classification^{1,2} than a filter of class η_1 , and can use more statistical information about the signal and noise. As the number of optimum cascaded filters is increased, the mean-square filtering error will converge to either zero or some finite value which must be less than, or equal to, the original mean-square error at the input to the cascaded chain. Thus the last of a large number of filters in the chain will not make an appreciable difference to the mean-square error; intuitively, it seems that this filter will become a direct connection, whence in eqns. (26) and (27),

$$\left. \begin{aligned} \omega_m(\tau_2) &= 0 & m \neq 1 \\ &= \sigma_y \delta(\tau_2) & m = 1 \\ &= \bar{y} \delta(\tau_2) & m = 0 \end{aligned} \right\} \quad \dots \quad (53)$$

Therefore the coefficients of the probability-density expansion for the output and desired output satisfy

$$\sigma_y C_{n1}(\tau) \equiv D_{n1}(\tau) \sigma_z \quad \dots \quad (54)$$

$$\bar{y} = \bar{z}$$

If the input and desired output are such that they satisfy this equation, the optimum filter of class η_{1R} is an attenuator.

A cascade connection of two optimum zero-memory filters is itself a zero-memory filter, so that it is to be expected that the second of the two filters will be a direct connection, since together they cannot produce a lower mean-square error than the first alone can achieve. However, we can cascade a zero-memory non-linear filter and a linear memory filter alternately and produce successively lower values of mean-square error. Such

an arrangement may have considerable practical advantages owing to the relative simplicity of the optimization procedure for both zero-memory and linear memory filters.

The optimization procedure proposed in the paper can be used for the analysis of non-linear systems. Given a non-linear box of class η_1 it is always possible to construct an approximate multi-path filter of class η_{1R} (approximate only in the sense that R is finite). Experimentally, the analysis of the box can be made by measuring the cross-correlations and auto-correlation of all the outputs of a set of Hermite-polynomial generators and the non-linear box output, having connected a common Gaussian input to them all (see Fig. 12). A multi-path filter is then

Fig. 12.—Experimental analysis of a non-linear system.

optimized to have an output as nearly as possible equal, in the mean-square-error sense, to the output of the non-linear box. The reason for the choice of Gaussian input and Hermite polynomials is that this produces integral equations [eqn. (26)] which are independent, thereby simplifying the mathematics, and also that a wide-band Gaussian input is sufficiently random for the unknown box to be completely characterized by the above measurements. Even if the unknown box is not of class η_1 an approximate equivalent filter of class η_{1R} can be found, and may be useful in the approximate analysis of more complicated non-linear systems. In this case the equivalence does not necessarily hold for different inputs and it may be better to find the equivalent filter under the same input conditions as those of the original non-linear box.

(6) ACKNOWLEDGMENTS

This work forms part of a research programme in control engineering, under the direction of Mr. J. F. Coales, Reader in Engineering at Cambridge University.

The help of many of the author's colleagues in the completion of this work through informal discussions is gratefully acknowledged; in particular, Mr. Coales has suggested many improvements both before and after reading the text; Mr. Loynes of the Statistical Laboratory suggested a generalization and extension in the theory; and a large part of the numerical calculations, which proved to be very time-consuming, was done by Mrs. Chambers and Mr. Lawrence of the Engineering Laboratory.

(7) REFERENCES

- (1) ZADEH, L. A.: 'A Contribution to the Theory of Non-Linear Systems', *Journal of the Franklin Institute*, 1953, **255**, p. 387.
- (2) ZADEH, L. A.: 'Optimum Non-Linear Filters', *Journal of Applied Physics*, 1953, **24**, p. 396.
- (3) SINGLETON, H. A.: 'Theory of Non-Linear Transducers', M.I.T. Technical Report No. 160, 1950.
- (4) BOSE, A. G.: 'A Theory of Non-Linear Systems', M.I.T. Technical Report No. 309, 1956.
- (5) COURANT, R.: 'Differential and Integral Calculus', Vol. 2 (Blackie, 1936).

- (6) WIENER, N.: 'Extrapolation, Interpolation and Smoothing of Stationary Time Series' (Wiley, 1949).
- (7) LAMPARD, D. G.: 'A New Method of Determining Correlation Functions of Stationary Time Series', *Proceedings I.E.E.*, Monograph No. 104 R, August, 1954 (**102 C**, p. 35).
- (8) LUBBOCK, J. K.: 'On a Semi-Separable Class of Random Processes', *Journal of Electronics and Control* (to be published).
- (9) NUTTALL, A. H.: 'Theory and Application of the Separable Class of Random Processes', M.I.T. Technical Report No. 343, 1958.
- (10) BROWN, J. L.: 'On a Cross-Correlation Property for Stationary Random Processes', *Transactions of the Institute of Radio Engineers*, 1957, **IT-3**, p. 28.
- (11) BARRETT, J. F., and LAMPARD, D. G.: 'An Expansion for some Second Order Probability Distributions and its Application to Noise Problems', *ibid.*, 1955, **IT-1**, p. 10.
- (12) BARRETT, J. F., and COALES, J. F.: 'An Introduction to the Analysis of Non-Linear Control Systems with Random Inputs', *Proceedings I.E.E.*, Monograph No. 154 M, November, 1955 (**103 C**, p. 190).
- (13) BALTH, VAN DER POL, and BREMMER: 'Operational Calculus' (Cambridge, 1955).

(8) APPENDICES

(8.1) An Expansion of Second Probability Density Functions

Let $x(t)$ and $z(t)$ be random functions of time belonging to the ensembles X and Z for which the first probability density functions $p(x)$ and $q(z)$ are known.

Suppose there exist two sets of polynomials $\theta_n(x)$ and $\gamma_m(z)$ which are orthonormal with respect to $p(x)$ and $q(z)$ respectively as weighting functions, over the whole range of x and z respectively:

$$\left. \begin{aligned} \int \theta_i(x) \theta_j(x) p(x) dx &= \delta_{ij} \\ \int \gamma_i(z) \gamma_j(z) q(z) dz &= \delta_{ij} \end{aligned} \right\} \dots \dots \dots (55)$$

where

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

The first few of these polynomials are easily identified, since

$$\left. \begin{aligned} \int p(x) dx &= 1 \\ \bar{x} &= \int x p(x) dx \\ \sigma_x^2 &= \int (x - \bar{x})^2 p(x) dx \end{aligned} \right\} \dots \dots \dots (56)$$

and

where $\bar{x}$ is the mean value of $x(t)$ and σ_x^2 is the variance of $x(t)$. Thence we may write

$$\left. \begin{aligned} \int 1 \times 1 \times p(x) dx &= 1 \\ \int \frac{x - \bar{x}}{\sigma_x} 1 \times p(x) dx &= 0 \\ \int \left(\frac{x - \bar{x}}{\sigma_x} \right)^2 p(x) dx &= 1 \end{aligned} \right\}$$

and we conclude that

$$\left. \begin{aligned} \theta_0(x) &= 1 & \theta_1(x) &= \frac{x - \bar{x}}{\sigma_x} \\ \gamma_0(z) &= 1 & \gamma_1(z) &= \frac{z - \bar{z}}{\sigma_z} \end{aligned} \right\} \dots \dots \dots (57)$$

The higher-order polynomials are all obtained from the moments of the distributions $p(x)$ and $q(z)$ (see Appendix 8.2).

The joint second probability density $p(x, z; \tau)$ may be expanded in series form:

$$p_2(x, z; \tau) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} D_{nm}(\tau) \theta_n(x) \gamma_m(z) p(x) q(z) \quad (58)$$

The restrictions on $p_2(x, z; \tau)$ to validate this formal procedure will not be enumerated.

Because of the orthogonality of $\theta_n(x)$ and $\gamma_m(z)$,

$$D_{nm}(\tau) = \iint \theta_n(x) \gamma_m(z) p_2(x, z; \tau) dx dz \quad (59)$$

The coefficients D_{nm} can all be expressed in terms of the joint moments of the form

$$\alpha_{rs}(\tau) = \overline{x^r z^s} = \iint x^r z^s p_2(x, z; \tau) dx dz \quad (60)$$

and we observe from eqns. (57) and (59) that

$$D_{00} = 1 \quad D_{0n} = D_{n0} = 0 \quad n \neq 0 \quad (61)$$

$$D_{11} = \frac{(\overline{x - \bar{x}})(\overline{z - \bar{z}})}{\sigma_x \sigma_z} = \rho_{xz}(\tau) \quad (62)$$

which is the normalized correlation function between $x(t)$ and $z(t)$. Certain other properties of these coefficients follow immediately.

$$\left. \begin{aligned} D_{nm}(\tau) &= D_{mn}(-\tau) \\ D_{nm}(0) &= \delta_{nm} \end{aligned} \right\} \quad (63)$$

provided that x and z are members of the same ensemble.

(8.1.1) Single-Series Expansion.

There exists a class of second probability density functions,¹¹ Λ , of physically realizable processes for which it is possible to expand the density function in the form of a single series:

$$p_2(x, z; \tau) = \sum_{n=0}^{\infty} D_{nn}(\tau) \theta_n(x) \gamma_n(z) p(x) q(z) \quad (64)$$

An example is the Gaussian second probability density function with zero mean value, $\bar{x} = 0$, for which we may write:

$$p_2(x_1, x_2; \tau) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp \left[-\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{2(1-\rho^2)} \right] \quad (65)$$

It can be shown in this case¹² that

$$p_2(x_1, x_2; \tau) = \frac{\exp \left(-\frac{x_1^2 + x_2^2}{2} \right)}{2\pi} \sum_{n=0}^{\infty} H_n(x_1) H_n(x_2) \frac{\rho^n}{n!} \quad (66)$$

$$\text{where } H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} \quad (67)$$

$$\text{and } \left. \begin{aligned} \int H_n(x) H_m(x) e^{-x^2/2} dx &= n! \sqrt{2\pi} \quad n=m \\ &= 0 \quad n \neq m \end{aligned} \right\} \quad (68)$$

$$\left. \begin{aligned} \text{Therefore } \theta_n(x) &= \frac{H_n(x)}{\sqrt{n!}} \\ p(x) &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \end{aligned} \right\} \quad (69)$$

and

$$D_{nn}(\tau) = \rho^n(\tau)$$

(8.2) Analytical Determination of the Coefficients of a Second Probability Density-Function Expansion

If both the first and second probability density functions of a process or the first and joint second probability density functions of a pair of processes are known, it is usually possible to obtain both the coefficients and the polynomials of the expansion of that second probability density function or joint second probability density function as follows.

(8.2.1) Moments of the Input.

To determine the appropriate polynomials for a given process (i.e. polynomials which are orthonormal with respect to the first probability density function of the process as weighting function) it is necessary first to determine the moments of that process. These can be found from the characteristic function $\psi(u)$ of the random process, which is defined as

$$\psi(u) = E(e^{jux}) = \int e^{jux} p(x) dx \quad (70)$$

Thus $\psi(u)$ is the Fourier transform of $p(x)$.

From the characteristic function the moments of the process may be determined using the following formula for the n th moment:

$$\alpha_n = \overline{x^n} = (-1)^{-n/2} \left[\frac{d^n}{du^n} \psi(u) \right]_{u=0} \quad (71)$$

(a) In filtering problems the filter input often consists of the sum of a signal $z(t)$ and noise $v(t)$:

$$x(t) = z(t) + v(t) \quad (72)$$

whence

$$\begin{aligned} \alpha_n &= \overline{x^n} = \overline{(v+z)^n} \\ &= \sum_{r=0}^n {}^nC_r \overline{v^r} \overline{z^{n-r}} \end{aligned} \quad (73)$$

when $z(t)$ and $v(t)$ are statistically independent.

(b) The input may be equal to the product of signal and noise:

$$x(t) = z(t)v(t) \quad (74)$$

In this case the moments of the input are found more easily, since

$$\overline{x^n} = \overline{z^n v^n} = \overline{z^n} \overline{v^n} \quad (75)$$

when $z(t)$ and $v(t)$ are statistically independent.

(8.2.2) Orthonormal Polynomials.

An orthonormal polynomial of order n will consist of a linear summation of the first n of the linearly independent functions

$$1, x, x^2, x^3, \dots$$

$$\text{Putting } \theta_0(x) = 1 \text{ and } \theta_1(x) = a_{11}x - a_{10}$$

such that

$$\int \theta_0(x) \theta_1(x) p(x) dx = \int (a_{11}x - a_{10}) p(x) dx = 0$$

$$\text{and } \int \theta_1^2(x) p(x) dx = \int (a_{11}x - a_{10})^2 p(x) dx = 1$$

we find that $a_{11} = 1/\sigma_x$ and $a_{10} = \bar{x}/\sigma_x$.

$$\text{Therefore } \theta_1(x) = \frac{x - \bar{x}}{\sigma_x} \quad (76)$$

$$\text{Let } \theta_2(x) = a_{22}x^2 - a_{21}\theta_1(x) - a_{20}$$

such that

$$\int \theta_2(x) \theta_1(x) p(x) dx = \int [a_{22}x^2 - a_{21}\theta_1(x) - a_{20}] \theta_1(x) p(x) dx = 0$$

$$\int \theta_2(x) \theta_0(x) p(x) dx = \int [a_{22}x^2 - a_{21}\theta_1(x) - a_{20}] p(x) dx = 0$$

$$\text{and } \int \theta_2^2(x) p(x) dx = \int [a_{22}x^2 - a_{21}\theta_1(x) - a_{20}]^2 p(x) dx = 1$$

$$\text{Then } \theta_2(x) = \frac{x^2 - \overline{x^2 \theta_1(x)} \theta_1(x) - \overline{x^2}}{\left\{ \overline{x^4} - [\overline{x^2 \theta_1(x)}]^2 - [\overline{x^2}]^2 \right\}^{1/2}} \quad (77)$$

For the general case

$$\theta_n(x) = a_{nn}x^n - \sum_{r=0}^{n-1} a_{nr}\theta_r(x) \quad (78)$$

where

$$\left. \begin{aligned} a_{nr} &= a_{nn} \overline{x^n \theta_r(x)} \\ a_{nn} &= \left\{ \overline{x^{2n}} - \sum_{r=0}^{n-1} [\overline{x^n \theta_r(x)}]^2 \right\}^{-1/2} \end{aligned} \right\} \quad (79)$$

$$\text{Therefore } \theta_n(x) = \frac{x^n - \sum_{r=0}^{n-1} \overline{x^n \theta_r(x)} \theta_r(x)}{\left\{ \overline{x^{2n}} - \sum_{r=0}^{n-1} [\overline{x^n \theta_r(x)}]^2 \right\}^{1/2}} \quad (80)$$

(8.2.3) Coefficients of the Second Probability Density-Function Expansions.

Using eqns. (79) we can obtain a recurrence formula for the coefficients $D_{n1}(\tau)$ occurring in the expansion of the joint second probability density $p_2(x, z; \tau)$ as follows:

$$\begin{aligned} \sigma_z D_{n1}(\tau) &= \sigma_z \overline{\theta_n(x) \gamma_1(z)} = \overline{\theta_n(x) z} \quad n \neq 0 \\ &= 0 \quad n = 0 \end{aligned}$$

$$\text{Therefore } \sigma_z D_{n1}(\tau) = a_{nn}x^n z - \sum_{r=1}^{n-1} a_{nr}\theta_r(x)z$$

$$\text{i.e. } \sigma_z D_{n1}(\tau) = a_{nn}\lambda_{n1} - \sum_{r=1}^{n-1} a_{nr}\sigma_z D_{r1}(\tau) \quad (81)$$

$$\text{where } \lambda_{n1}(\tau) = \iint x^n z p_{xz}(x, z; \tau) dx dz \quad (82)$$

A similar but more complicated expression can be found for the coefficients $C_{nm}(\tau)$ occurring in the expansion of the second probability density:

$$p_2(x_1, x_2; \tau) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} C_{nm}(\tau) \theta_n(x_1) \theta_m(x_2) p(x_1) p(x_2) \quad (83)$$

$$\text{where } C_{nm}(\tau) = \iint \theta_n(x_1) \theta_m(x_2) p_2(x_1, x_2; \tau) dx_1 dx_2 \quad (84)$$

For convenience we choose to express eqn. (80) in the form

$$\theta_n(x) = b_n x^n + b_{n-1} x^{n-1} + \dots + b_r x^r + \dots + b_0 \quad (85)$$

$$\text{Then } C_{nm}(\tau) = \sum_{r=0}^n \sum_{s=0}^m b_r b_s \alpha_{rs}(\tau) \quad (86)$$

$$\text{where } \alpha_{rs}(\tau) = \iint x_1^r x_2^s p_{xz}(x_1, x_2; \tau) dx_1 dx_2 \quad (87)$$

Therefore all the coefficients $C_{nm}(\tau)$ and $D_{n1}(\tau)$ needed for the optimization can be expressed in terms of the joint moments

$\alpha_{rs}(\tau)$ and $\lambda_{n1}(\tau)$ defined by eqns. (87) and (82) respectively. These joint moments can be found from the joint characteristic functions; in particular

$$\lambda_{nm}(\tau) = \overline{x^n z^m} = (-1)^{-\frac{(n+m)}{2}} \frac{\partial^n}{\partial u_1^n} \frac{\partial^m}{\partial u_2^m} \psi_{xz}(u_1, u_2; \tau) \quad (88)$$

$$\text{where } \psi_{xz}(u_1, u_2; \tau) = \iint e^{j(u_1 x + u_2 z)} p_{xz}(x, z; \tau) dx dz \quad (89)$$

(a) When the input is equal to the sum of signal, $z(t)$ and noise, $v(t)$,

$$\begin{aligned} \alpha_{nm} &= \overline{x_1^n x_2^m} = \overline{(v_1 + z_1)^n (v_2 + z_2)^m} \\ &= \sum_{r=0}^n \sum_{s=0}^m {}^nC_r {}^mC_s \overline{v_1^r z_1^{n-r} v_2^s z_2^{m-s}} \end{aligned}$$

and when $z(t)$ and $v(t)$ are statistically independent,

$$\alpha_{nm} = \sum_{r=0}^n \sum_{s=0}^m {}^nC_r {}^mC_s \overline{v_1^{n-r} z_1^{r-n} v_2^s z_2^{m-s}} \overline{v_1^r v_2^s} \quad (90)$$

$$\begin{aligned} \lambda_{n1} &= \overline{x^n z} = \overline{(v_1 + z_1)^n z_2} \\ &= \sum_{r=0}^n {}^nC_r \overline{v_1^r z_1^{n-r} z_2} \end{aligned}$$

and when $z(t)$ and $v(t)$ are statistically independent,

$$\lambda_{n1} = \sum_{r=0}^n {}^nC_r \overline{v_1^r z_1^{n-r} z_2} \quad (91)$$

(b) When the input is equal to the product of signal and noise the joint moments are easily found, since

$$\begin{aligned} \alpha_{nm}(\tau) &= \overline{x_1^n x_2^m} = \overline{z_1^n v_1^n z_2^m v_2^m} \\ &= \overline{z_1^n z_2^m} \overline{v_1^n v_2^m} \end{aligned} \quad (92)$$

provided that $z(t)$ and $v(t)$ are statistically independent, and

$$\lambda_{n1}(\tau) = \overline{x_1^n z_2} = \overline{z_1^n z_2} \overline{v_1^n} \quad (93)$$

provided that $z(t)$ and $v(t)$ are statistically independent.

It will be appreciated that the labour involved may be very great in order to compute first the polynomials and then the coefficients of an expansion for the second or joint probability density function from a knowledge of either the second or joint characteristic function or the probability density function itself. Also, when the expansion corresponds to a process which itself is the sum of two other processes whose probability densities are given [as in (a)] the work is greater still. One such problem, the solution to which is given earlier, took about 120 hours to calculate the expansion coefficients on a desk calculator.

(8.3) Mean-Square Filtering Error

The mean-square filtering error is given by eqn. (1):

$$\overline{e^2} = \overline{[y(t) - z(t)]^2} = \overline{y(t)^2} + \overline{z(t)^2} - 2\overline{y(t)z(t)} \quad (1)$$

Let $y(t) = g[x(t)]$, where g represents any linear or non-linear operation, of finite mean square, on the whole, or part, of the past input.

The variation in $\overline{e^2}$ produced by a variation in $g[x]$ of $\epsilon h[x]$, where ϵ is a small constant and $h[x]$ again depends only on the past of x , is given by

$$\Delta \overline{e^2} = -2\epsilon \overline{h(x)[g(x) - z]} + \epsilon^2 \overline{h(x)^2}$$

For $g(x)$ to be optimum,

$$\left[\frac{\partial \Delta e^2}{\partial \epsilon} \right]_{\epsilon=0} = -2\overline{h(x)[g(x) - z]} = 0$$

$$\text{Therefore } \overline{h(x)[g(x) - z]} = 0 \quad (94)$$

Thence the following two conditions are derived:

For a chosen variation $\epsilon h(x) = \epsilon g(x)$

$$\begin{aligned} \overline{g(x)^2} &= \overline{zg(x)} \\ \text{i.e. } \overline{y(t)^2} &= \overline{z(t)y(t)} \end{aligned} \quad (95)$$

For a chosen variation $\epsilon h(x) = \epsilon \times 1$

$$\begin{aligned} \overline{g(x)} &= \overline{z} \\ \text{i.e. } \overline{y(t)} &= \overline{z(t)} \end{aligned} \quad (96)$$

Eqns. (95) and (96) are seen to be necessary if $g(x)$ represents an optimum filter belonging to any class.

Substitution of eqns. (95) and (96) in eqn. (1) yields

$$\overline{e^2} = \sigma_z^2 - \sigma_y^2 \quad (97)$$

That this result is true for each of the classes of filters previously considered is easily verified.

Furthermore, suppose there should exist more than one optimum solution for $g(x)$; by taking any two solutions $g_1(x)$ and $g_2(x)$ we may write, using eqn. (94),

$$\overline{g_1(x)[g_2(x) - z]} = 0$$

$$\text{Similarly } \overline{g_2(x)[g_1(x) - z]} = 0$$

$$\text{whence } \overline{g_1(x)z} = \overline{g_2(x)z} \quad (98)$$

From eqns. (95) and (98) it is seen that the mean-square error [eqn. (1)] is the same for each of the two solutions $g_1(x)$ and $g_2(x)$. Thus although a particular solution may not be unique we can be sure that it is no worse than any other solution in so far as it gives the same mean-square error.

(8.3.1) The Class of Multi-Path Zero-Memory Filters of Degree R .

Eqn. (94) gives the sufficient conditions for $g(x)$ to represent an optimum filter of any specified class where $h(x)$ represents any physically realizable filter of the same class.

For the class of multi-path zero-memory filters of degree R ,

$$g(x) = \sum_{m=0}^R k_{gm} \theta_m(x)$$

The most general form for $h(x)$ is

$$h(x) = \sum_{n=0}^R k_{hn} \theta_n(x)$$

Substitution in eqn. (94) gives

$$\sum_{n=0}^R k_{hn} \left[\sum_{m=0}^R k_{gm} \overline{\theta_n(x) \theta_m(x)} - \overline{\theta_n(x) z} \right] = 0 \quad (99)$$

$$\begin{aligned} \overline{\theta_n(x) \theta_m(x)} &= 0 \quad m \neq n \\ &= 1 \quad m = n \end{aligned}$$

$$\text{we obtain } \sum_{n=0}^R k_{hn} k_{gn} = \sum_{n=0}^R k_{hn} \overline{\theta_n(x) z} \quad (100)$$

Since the set of constants k_{hn} can be given any pre-assigned set of values only one solution exists, namely

$$k_{gn} = \overline{\theta_n(x) z} \quad n = 0, 1, 2 \dots R \quad (102)$$

Also, since $\theta_0(x) = 1$,

$$k_{g0} = \overline{g(x)} = \overline{y(t)} = \overline{z(t)} \quad (103)$$

and the optimum multi-path zero-memory filter of degree R is given by

$$f[x(t)] = \sum_{n=1}^R \sigma_z D_{n1} \theta_n[x(t)] + \overline{z(t)} \quad (104)$$

This equation is identical with the equation obtained by truncating the expansion (21). From eqn. (104)

$$\sigma_y^2 = \overline{\{f[x(t)] - \overline{z(t)}\}^2} = \sum_{n=1}^R \sigma_z^2 D_{n1}^2$$

$$\begin{aligned} \overline{\theta_n(x) \theta_m(x)} &= 0 \quad m \neq n \\ &= 1 \quad m = n \end{aligned}$$

Substitution in eqn. (97) gives

$$\overline{e^2} = \sigma_z^2 \left(1 - \frac{\sigma_y^2}{\sigma_z^2} \right) = \sigma_z^2 \left(1 - \sum_{n=1}^R D_{n1}^2 \right) \quad (105)$$

(8.3.2) The Class of Multi-Path Memory Filters of Degree R .

$$g(x) = \sum_{m=0}^R \int_0^\infty \omega_{gm}(\tau) \theta_m[x(t - \tau)] d\tau$$

The most general form for $h(x)$ is

$$h(x) = \sum_{n=0}^R \int_0^\infty \omega_{hn}(\tau) \theta_n[x(t - \tau)] d\tau$$

Substitution in eqn. (94) gives

$$\begin{aligned} \sum_{n=0}^R \int_0^\infty \omega_{hn}(\tau_1) \left\{ \sum_{m=0}^R \int_0^\infty \omega_{gm}(\tau_2) C_{nm}[(\tau_1 - \tau_2)] d\tau_2 \right. \\ \left. - \overline{\theta_n[x(t - \tau_1)] z(t)} \right\} d\tau_1 = 0 \quad (106) \end{aligned}$$

Using the fundamental lemma of variation calculus⁵ and noting that all $\omega_{hn}(\tau)$ are arbitrary but physically realizable weighting functions of linear networks [i.e. $\omega_{hn}(\tau) = 0$ for $\tau < 0$], we obtain

$$\begin{aligned} \sum_{m=0}^R \int_0^\infty \omega_{gm}(\tau_2) C_{nm}[(\tau_1 - \tau_2)] d\tau_2 - \overline{\theta_n[x(t - \tau_1)] z(t)} = 0 \quad (107) \\ n = 0, 1, 2 \dots R \quad \tau_1 \geq 0 \end{aligned}$$

but since $C_{0m} = 0, m \neq 0; C_{0m} = 1, m = 0$; and $\theta_0[x(t)] = 1$

$$\int_0^\infty \omega_0(\tau) d\tau = \omega_0 = \overline{z(t)} \quad (27)$$

$$\text{and } \sum_{m=1}^R \int_0^\infty \omega_{gm}(\tau_2) C_{nm}[(\tau_1 - \tau_2)] d\tau_2 = \sigma_z D_{n1}(\tau_1) \quad \tau_1 \geq 0 \quad (26)$$

From eqns. (26) and (27), it follows that

$$\sigma_y^2 = \sum_{n=1}^R \sum_{m=1}^R \int_0^\infty \int_0^\infty \omega_n(\tau_1) \omega_m(\tau_2) C_{nm}[(\tau_1 - \tau_2)] d\tau_1 d\tau_2 \quad (108)$$

$$\text{or we may write } \sigma_y^2 = \sum_{n=1}^R \int_0^\infty D_{n1}(\tau) \sigma_z \omega_n(\tau) d\tau \quad (109)$$

and by substitution in eqn. (97) obtain:

$$\bar{e}^2 = \sigma_z^2 \left(1 - \frac{\sigma_y^2}{\sigma_z^2}\right) = \sigma_z^2 - \sum_{n=1}^R \int_0^\infty D_{n1}(\tau) \sigma_z \omega_n(\tau) d\tau \quad (110)$$

(8.3.3) Performance Index for Noise Filters.

Even in the most unfavourable conditions when the statistical characteristics of the signal and noise are identical, an optimum linear or non-linear noise filter will reduce the mean-square output error given by the optimum attenuator; for example, the optimum linear filter for the condition when the input signal and noise have the same spectral densities becomes an attenuator. Therefore it is both logical and convenient to choose the mean-square error of the optimum attenuator as a datum against which to compare the mean-square errors of more complex optimum filters.

The optimum attenuator mean-square error is given by eqn. (107):

$$\bar{e}_a^2 = \sigma_z^2(1 - D_{11}^2) \quad (111)$$

and the comparative mean-square error by

$$\frac{\bar{e}^2}{\bar{e}_a^2} = \frac{1 - \sigma_y^2/\sigma_z^2}{1 - D_{11}^2} \quad (112)$$

When the signal and noise are additive and uncorrelated,

$$\frac{\bar{e}^2}{\bar{e}_a^2} = \frac{1 - \sigma_y^2/\sigma_z^2}{1 - \sigma_z^2/\sigma_x^2} \quad (113)$$

where $\bar{e}_a^2$ is the mean-square error of the optimum attenuator, $\bar{e}^2$ is the mean-square filter error and $0 \leq \bar{e}^2/\bar{e}_a^2 \leq 1$.

The two limiting values of $\bar{e}^2/\bar{e}_a^2$ represent the extreme conditions when the optimum filter is perfect and when it is no better than the optimum attenuator. If $\bar{e}^2/\bar{e}_a^2 < 1$ we say the improvement in mean-square error is 'significant'.

(8.4) Method of Undetermined Coefficients

The solution for the set of integral equations [eqn. (26)] has been worked out by Wiener,⁴ and a modification of his method is formally presented here.

The integral equations are

$$\sum_{m=1}^R \int_0^\infty C_{nm}[(\tau_1 - \tau_2)] \omega_m(\tau_2) d\tau_2 - D_{n1}(\tau_1) \sigma_z = H_n(\tau_1) \quad (26)$$

where $H_n(\tau_1) = 0 \quad \tau_1 \geq 0 \quad n = 1, 2, \dots, R$

Taking the bilateral Laplace transform¹³ of both sides of these equations,

$$\sum_{m=1}^R \mathcal{C}_{nm}(p) \mathcal{W}_m(p) - \mathcal{D}_{n1}(p) = \mathcal{H}_n(p) \quad (114)$$

where $\mathcal{H}_n(p)$ has singularities in the right-hand half of the p -plane only [since $H_n(\tau_1) = 0, \tau_1 \geq 0$].

For simplicity the method will be described for the case of two equations only.

$$\text{Let} \quad \Phi(p) = \begin{vmatrix} \mathcal{C}_{11}(p) & \mathcal{C}_{12}(p) \\ \mathcal{C}_{21}(p) & \mathcal{C}_{22}(p) \end{vmatrix} \quad (115)$$

$$\text{then} \quad \Phi(p) \mathcal{W}_1(p) = \begin{vmatrix} \mathcal{D}_{11}(p) + \mathcal{H}_1(p) & \mathcal{C}_{12}(p) \\ \mathcal{D}_{21}(p) + \mathcal{H}_2(p) & \mathcal{C}_{22}(p) \end{vmatrix} \quad (116)$$

$$\text{and} \quad \Phi(p) \mathcal{W}_2(p) = \begin{vmatrix} \mathcal{C}_{11}(p) & \mathcal{D}_{11}(p) + \mathcal{H}_1(p) \\ \mathcal{C}_{21}(p) & \mathcal{D}_{21}(p) + \mathcal{H}_2(p) \end{vmatrix} \quad (117)$$

Therefore

$$\Phi(p) \mathcal{W}_1(p) = [\mathcal{D}_{11} \mathcal{C}_{22} - \mathcal{D}_{21} \mathcal{C}_{12}] + \mathcal{H}_1 \mathcal{C}_{22} - \mathcal{H}_2 \mathcal{C}_{12} \quad (118)$$

$$\text{Suppose we can write} \quad \Phi(p) = \Phi^+(p) \Phi^-(p) \quad (119)$$

where $\Phi^+(p)$ has poles and zeros in the left-hand half of the p -plane only, and $\Phi^-(p)$ has poles and zeros in the right-hand half only. Then

$$\mathcal{W}_1(p) \Phi^+(p) = \frac{\mathcal{D}_{11} \mathcal{C}_{22} - \mathcal{D}_{21} \mathcal{C}_{12}}{\Phi^-(p)} + \frac{\mathcal{H}_1 \mathcal{C}_{22} - \mathcal{H}_2 \mathcal{C}_{12}}{\Phi^-(p)} \quad (120)$$

$\mathcal{W}_1(p) \Phi^+(p)$ has poles in the left-hand half of the p -plane only if $\mathcal{W}_1(p)$ is the transfer function of a stable system, i.e. $\omega_1(\tau) = 0, \tau < 0$. Also, it may usually be assumed that

$$\lim_{p \rightarrow \infty} \mathcal{W}_1(p) \Phi^+(p) = 0 \quad (121)$$

It may be necessary to impose some further restriction on $\mathcal{W}_m(p)$ so as to obtain a unique solution to eqns. (26). This restriction is of the form:

$$\lim_{p \rightarrow \infty} \frac{\mathcal{W}_m(p)}{p^i} = \text{constant} \quad (122)$$

where for noise-filtering problems normally $i = 0$ and in prediction problems normally $i = 1$. Only approximate physical filters for $i \geq 0$ can be realized, but the approximation is not serious in practice if i is not too large. In any particular problem i is chosen so that a unique solution is obtained.

Assuming $\mathcal{W}_1(p) \Phi^+(p)$ is a rational algebraic function, its numerator must of a lower order than its denominator.

Therefore

$$\mathcal{W}_1(p) \Phi^+(p) = \left[\frac{\mathcal{D}_{11} \mathcal{C}_{22} - \mathcal{D}_{21} \mathcal{C}_{12}}{\Phi^-(p)} \right]_+ + \left[\frac{\mathcal{H}_1 \mathcal{C}_{22} - \mathcal{H}_2 \mathcal{C}_{12}}{\Phi^-(p)} \right]_+ \quad (123)$$

where the suffixes '+' indicate that only those terms of the partial-fraction expansion representing poles in the left-hand half of the p -plane have been taken.

Since $\mathcal{H}_1$ and $\mathcal{H}_2$ have no poles in the left-hand half-plane, poles in this half-plane of $(\mathcal{H}_1 \mathcal{C}_{22} - \mathcal{H}_2 \mathcal{C}_{12})/\Phi^-(p)$ must be poles of both $\mathcal{C}_{22}$ and $\mathcal{C}_{12}$. We can write

$$\frac{\mathcal{H}_1 \mathcal{C}_{22} - \mathcal{H}_2 \mathcal{C}_{12}}{\Phi^-(p)} = \frac{a}{p + \alpha} + \frac{b}{p + \beta} + \frac{c}{p + \gamma} \quad (124)$$

where $a, b, c, \dots$ are undetermined coefficients, which may be determined by substitution for $\mathcal{W}_1(p)$ and $\mathcal{W}_2(p)$ in one of the original equations (114).

A set of simultaneous equations in the undetermined coefficients of both $\mathcal{W}_1$ and $\mathcal{W}_2$ is obtained by equating residues of left-hand half-plane poles of this equation and by using eqn. (122).

(8.5) A Second Probability Density of Class Λ

We shall consider the signal developed at the output of a zero-memory non-linear device characterized by the equation

$$z(t) = \frac{H_2[s(t)]}{\sqrt{2}} = \frac{s^2 - 1}{\sqrt{2}} \quad (125)$$

where $z(t)$ is the output, $s(t)$ the input

$$\text{and} \quad H_n(s) = (-1)^n \epsilon^{s^2/2} \frac{d^n}{ds^n} \epsilon^{-s^2/2}$$

When the input is Gaussian with zero mean and unit variance, the output is a signal whose probability density function $p_2(z_1, z_2; \tau)$ is of class Λ .

It is always possible to find a polynomial θ_m of order m such that

$$y = \theta_m[H_2(s)] = H_{2m}(s) \quad (126)$$

A proof of this statement is not given here but it can be verified for all finite values of m as follows:

Let θ_m and θ_n be two such polynomials, and write

$$\phi_{mn} = \iint \theta_m(z_1) \theta_n(z_2) p_2(z_1, z_2; \tau) dz_1 dz_2$$

where ϕ_{mn} is the cross-correlation between the outputs y_1 and y_2 of the polynomial-function generators θ_n and θ_m (see Fig. 13).

Fig. 13.—Polynomials θ_n and θ_m chosen so that $y_1 = H_{2n}[s(t)]$ and $y_2 = H_{2m}[s(t)]$ when $s(t)$ is a Gaussian signal.

Then

$$\begin{aligned} \phi_{mn} &= \iint \theta_m[H_2(s_1)] \theta_n[H_2(s_2)] p_2(s_1, s_2; \tau) ds_1 ds_2 \\ &= \iint H_{2m}(s_1) H_{2n}(s_2) p_2(s_1, s_2; \tau) ds_1 ds_2 \end{aligned}$$

whence

$$\begin{aligned} \phi_{mn} &= 0 & n \neq m \\ &= \rho_{s_1 s_2}^{2n}(\tau) & n = m \end{aligned} \quad (127)$$

where $\rho_{s_1 s_2}$ is the normalized auto-correlation coefficient of $s(t)$. But $\phi_{mn} = C_{mn}$, using the same notation as used in eqns. (86) and (87), and $p_2(z_1, z_2; \tau)$ is a member of class Λ . We also note that $\rho_{z_1 z_2}(\tau) = \rho_{s_1 s_2}^2(\tau)$.

Therefore

$$C_{nn}(\tau) = \rho_{z_1 z_2}^n(\tau) \quad (128)$$

whence, if $\rho_{s_1 s_2}(\tau) = e^{-\beta|\tau|}$, then $\rho_{z_1 z_2}(\tau) = e^{-2\beta|\tau|}$. The probability density of z can be obtained as follows:

$$p(s) = \frac{1}{2\pi} e^{-s^2/2} \text{ for zero mean and unit variance.}$$

$$s = \sqrt{(z\sqrt{2} + 1)}$$

$$2s ds = 2dz$$

Therefore

$$\begin{aligned} p(z) &= \frac{1}{2\sqrt{\pi}} e^{-\sqrt{2} + 1/2} \frac{1}{\sqrt{(z\sqrt{2} + 1)}} \quad z > -\frac{1}{\sqrt{2}} \\ &= 0 \quad z \leq -\frac{1}{\sqrt{2}} \end{aligned} \quad (129)$$

The moments of z can be found from

$$\bar{z}^n = \int z^n p(z) dz$$

or from

$$\bar{z}^n = \left(\frac{s^2 - 1}{\sqrt{2}} \right)^n \quad (130)$$

given the moments of a Gaussian process with zero mean and unit variance.

Whence $\bar{z} = 0$, $\bar{z}^2 = 1$, $\bar{z}^3 = 2.828427$, $\bar{z}^4 = 15$, $\bar{z}^5 = 96.166522$, $\bar{z}^6 = 755.0$, $\bar{z}^7 = 6983.386571$, $\bar{z}^8 = 74417$, $\bar{z}^9 = 897799.3379$ and $\bar{z}^{10} = 12096873$.