PREDICTIVE DECOMPOSITION OF TIME SERIES WITH APPLICATIONS TO SEISMIC EXPLORATION

by

Enders Anthony Robinson

S.B. Massachusetts Institute of Technology

(1950)

S.M. Massachusetts Institute of Technology

(1952)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

(1954)

Signature of Author_____
Department of Geology and Geophysics July 26, 1954

Certified by_____
Thesis Supervisor

Accepted by_____
Chairman, Departmental Committee on Graduate Students

AKNOWLEDGEMENTS

The author wishes to express his sincere thanks to Professor G. P. Wadsworth for his help and ideas in the writing of this thesis. To a large measure, this thesis reflects what the author has learned from Professor Wadsworth during the past seven years as one of his pupils and one of his workers.

The author expresses his sincere thanks to Professor P. M. Hurley, Dr. J. G. Bryan, and Professor S. M. Simpson for their encouragement and ideas.

He is indebted to the many people who have contributed to this thesis, directly and indirectly, through their work for the Geophysical Analysis Group. He wishes to thank Dr. M. K. Smith, R. Bowman, D. E. Bowker, H. W. Briscoe, Miss Irene J. Calnan, J. F. Gilbert, D. Grine, Dr. K. E. Haq, S. Treitel, Mrs. May S. Turyn, and W. P. Walsh, who are or have been members of the Geophysical Analysis Group; Miss Virginia Woodward of Professor Wadsworth's staff; J. D. Porter and other members of the staff of the MIT Digital Computer Laboratory; Dr. R. F. Clippinger, Dr. J. H. Levin, and Dr. B. Dimsdale of the Raytheon Manufacturing Company; the members of the MIT Illustration Service and the MIT Photographic Service.

The author is grateful for the suggestions and criticisms given at the annual meetings by the members of the Advisory Committee of the oil and geophysical companies which have supported the Geophysical Analysis Group.

He wishes to thank the Office of Naval Research for making available Digital Computer Laboratory time.

The author wishes to express his sincere thanks to Miss Barbara Halpern for her patient and efficient assistance at every step in the writting of this thesis and for her careful typing of the thesis and its preparatory drafts.

TABLE OF CONTENTS

Page

ABSTRACT

PREDICTIVE DECOMPOSITION OF TIME SERIES
WITH APPLICATIONS TO SEISMIC EXPLORATION
by
Enders Anthony Robinson

Submitted to the Department of Geology and Geophysics on July 26, 1954 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This thesis presents in an expository manner a treatment of the theory of discrete stationary time series as developed by Cramér, Doob, Khintchine, Kolmogorov, Wiener, Wold, Yule, and others. The central theme deals with the development of the concept of the predictive decomposition of stationary time series from the point of view of applications. The Predictive Decomposition Theorem of Herman Wold (A Study in the Analysis of Stationary Time Series, Almqvist and Wiksells, Uppsala, 1938) states that a stationary time series (with an absolutely continuous spectral distribution) in additively composed of many overlapping wavelets, or pulses, which arrive as time progresses. These wavelets all have the same unique stable shape or form; and the arrival times and strengths of these wavelets are random and uncorrelated with each other.

Specific mathematical results of this thesis are:

(1) The theory of linear difference equations familiar to statisticians is united with the theory of linear systems familiar to electrical engineers. It is shown that the condition that a linear difference equation formed by the coefficients of a discrete linear operator be a stable difference equation is the condition that the Fourier transform of the linear operator has no singularities or zeros below the axis of real frequency. In other works, a stable difference equation has filter characteristics with minimum phase-shift characteristic. Computational formulae are given for the determination of this minimum phase-shift characteristic from the absolute gain characteristic of the desired filtering properties of a linear operator.

(2) The function which Wiener (The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications. National Defense Research Council (Section D2) MIT DIC Contract 6037, Cambridge, 1942) designates as the coefficient function to be used directly in making an optimum prediction for discrete stationary time series with absolutely continuous spectral distributions is shown to be the same function obtained by Wold (1938) for this purpose.

(3) The expression which Wiener (1942) gives for the mean square error of the optimum prediction is shown to be the same expression given by Wold (1938).

(4) The general solution of the filtering problem for discrete stationary time series given by Wiener (1942; Cybernetics, John Wiley, New York, 1948) is shown to be a direct consequence of the Predictive Decomposition Theorem.

(5) Wiener (1942) recommends that the geophysicist use a certain technical point in the computation of spectra for Wiener's general technique of discrete prediction. It is shown that if the geophysicist does follow this recommendation, the general technique of discrete prediction will fail.

The applications to seismic exploration deal with the case in which a section of seismic trace (recorded with automatic volume control) is additively composed of seismic wavelets, or Ricker wavelets, where each wavelet has the same stable shape, and where their strengths and arrival times may be considered to be random and uncorrelated with each other. For this case, the Predictive Decomposition Theorem tells us that the section of seismic trace is a section of a stationary time series. The problem of the separation of the dynamic component (the wavelet shape) from the random components (the strengths and arrival times of the wavelets) is considered.

For an infinite discrete stationary time series, the solution of this problem consists of the following steps:

(1) Average out the random components of the time series so as to yield the unique stable wavelet shape. This stable wavelet shape is shown to be the Fourier transform of the factor of the power spectrum of the time series, where this factor is required to have no singularities or zeros below the axis of real frequency.

(2) From the wavelet shape thus found, compute the inverse wavelet shape, which is shown to be the prediction operator for unit prediction distance.

(3) Compute the prediction errors by applying this prediction operator, or inverse wavelet shape, to the time series. These prediction errors are shown to represent the arrival times and strengths of the wavelets.

For finite discrete time series, such as the section of seismic trace, the solution of this problem consists of estimating the prediction operator directly by the Gauss method of least squares, and then using this prediction operator to determine estimates of the wavelet shape and prediction errors.

Thesis Supervisor: Dr. P. M. Hurley
Title: Professor of Geology

CHAPTER I

INTRODUCTION AND SUMMARY

1.1 Introduction

In exploration seismology, a charge of dynamite is exploded under controlled conditions, and the resulting vibrations at various points on the surface of the ground are detected by geophones and are recorded as seismic traces on the seismogram. The analysis of such seismic records yields valuable information about the structure of the sedimentary rock layers in potential oil producing areas, and such information is of considerable economic value in increasing the probability of locating new oil fields.

Present day techniques require the visual examination and mental interpretation of seismograms, with considerable importance placed on the detection of reflected energy or "reflections" which indicate reflecting horizons of subsurface sedimentary layers. From this information the geologic structure of an area may be estimated.

Although reflection seismology is only about a quarter of a century old, it has played an important role in the discovery of many of the world's oil fields. The credit

for much of this success belongs to the oil companies and geophysical companies which have the practical task of locating new petroleum reserves. It was the working geophysicist of these companies who developed a large part of the seismic method of today. To help him in his job, the engineer, who in many instances is the geophysicist himself, has developed the instrumentation needed for the ever increasing demands of seismic exploration. In addition, the research scientist has taken an active role in the development of the basic scientific theory of exploration seismology.

For a further discussion of the seismic method, together with references to the literature, the reader is referred to books by Dix (1952), Dobrin (1952), Heiland (1940), Jakosky (1950), and Nettleton (1940), and also to GEOPHYSICS, the quarterly journal of the Society of Exploration Geophysicists.

A large part of basic seismic research is directed toward a better understanding of the physical processes involved in the seismic method. Such an approach is fundamentally sound. From this point of view, the seismic trace is the response of the system consisting of the earth and recording apparatus to the impulsive source, the explosion. This system, although usually very complicated, is susceptible to a

deterministic (non-random) approach toward its analysis. To this end, controlled experiments may be carried out, and mathematical and physical models may be set up from the resulting data. Careful replication of the experiment and high precision of measurement can render such data very accurate.

On the other hand, large numbers of seismic records, which have as many as twenty or more traces per record, are needed to carry out an exploration program over a geographic area. This quantity of data necessarily requires the consideration of each record as a member of a larger group or ensemble of records. Thus the reliability of a single record is considerably less than the reliability of the ensemble of records in connection with the description of the geologic conditions existing in that area. Also from an economic standpoint, the amount of control in such an exploration program must be kept at the bare minimum consistent with worthwhile results. Thus, as a rule, the controlled experiment aspect of exploration seismology, although possible, falls short of the needs of a research scientist who wishes to set up a mathematical or physical model. As a result, in these cases the working geophysicist must proceed to fit his

empirical information into the larger overall framework without the aid of elaborate mathematical or physical models. In particular, he is faced with the general problems of description, analysis, and prediction (Cramér, 1946).

That is, first, the working seismologist is faced with problems of the description of the overall exploration seismic picture. In particular, he wishes to replace the mass of original data, which is of a complicated nature, by a small number of descriptive characteristics; that is, he is faced with the problem of the reduction of data.

Next, he is concerned with the problems of analysis in which he wishes to argue from the sample, the evidence from a limited number of seismograms, to the population the geologic structure of the area. In other words, from the sample data he wishes to find estimates of the true values which describe the geologic structure.

Finally, the working geophysicist is concerned with the problem of prediction, that is, from knowledge of past experience what course of action should he take in the future. In particular one of the goals of an exploration program is to determine favorable drilling sites.

Since the geologic structure is physically fixed and constant in nature, and has no intrinsic random character-

istics, any statistical approach to these problems immediately encounters difficulties which are commonly associated with Bayes' Theorem in the statistical literature (Cramér, 1946; Jeffreys, 1939). Nevertheless modern statistical theory admits the bypassing of these difficulties, although with reservation, and hence the working geophysicist may be considered to be faced with a situation which is essentially statistical. For example, a reflection which may be followed from trace to trace, record to record, usually has more value to the seismic interpreter, and hence is statistically more significant, than a reflection which appears only on a few traces. Such a procedure in picking reflections does not imply that the reflection which appears only on a few traces is necessarily spurious information, but only that economic limitations preclude further examination and experimentation which may render it in a more useful form.

This thesis deals with the analysis of seismic records from the statistical point of view. In those years in which the exploration seismic method was first being developed, the English statistician, G. Udny Yule, was developing methods of time series analysis which proved to open a new epoch in the analysis of time functions. The concept which Yule introduced is that a large class of disturbed motions are built

up in a time sequence from wavelets whose arrival times and strengths (or amplitudes) are random. Thus the principal feature of this class of disturbed movements is a continual shift of phase and change of amplitude as time progresses. Yule applied this hypothesis, with success, to empirical data, and thus the analysis of time series was freed for the first time from either the hypothesis of a strictly periodic variation or aperiodic variation, or the counter hypothesis of a purely random variation. Yule's concept was formulated on a firm axiomatic basis in the founding of the theory of stochastic processes by the Russian statisticians A. Kolmogorov (1933) and A. Khintchine (1933), and in the definition and basic work in theory of stationary processes by A. Khintchine (1934).

The Swedish statistician, Harald Cramér, taught a course on Time Series Analysis in 1933 which laid the foundation for a thesis prepared by his student, the Swedish statistician and economist, Herman Wold (Wold, 1938, Preface). This thesis was published in book form in 1938. Wold, in the Preface, describes his work as a trial to subject the fertile methods of empirical analysis proposed by Yule to an examination and a development by the use of the mathematically strict tools supplied by the modern theory of probability. In his work

Wold develops the predictive decomposition of stationary time series, about which the present thesis is centered.

For the past four years, the author has worked on a research program at the Massachusetts Institute of Technology to apply these statistical methods to the analysis of seismic records under the supervision of Professor G. P. Wadsworth of the Department of Mathematics, Professor P. M. Hurley of the Department of Geology and Geophysics, and Dr. J. G. Bryan of the Division of Industrial Cooperation. This program was initially supported by the Department of Mathematics. In February, 1952, the program was incorporated into the Department of Geology and Geophysics as the Geophysical Analysis Group. From that time until February, 1953, it was supported by the Magnolia Petroleum Company, and from then until the present time by fourteen companies of the petroleum industry.

Computational and theoretical results of this research program are given in Wadsworth, et al, (1953) and six reports (MIT GAG Reports No. 1 - No. 6) of the Geophysical Analysis Group, to which the reader is referred. The present thesis attempts to expound more fully on the reasoning presented there, and, in particular, the predictive techniques proposed for the study of seismic records.

In the final analysis, the potential usefulness of the statistical approach depends upon the coordination of statistical methods with knowledge of practical and theoretical seismology.

1.2 Summary of Chapters

In this section we should like to present a summary of the chapters which follow. Since detailed references to the literature are given in these chapters, we shall not state any references in this summary.

In Chapter II we discuss the properties of finite discrete linear operators. We distinguish between extrapolation or prediction type operators on the one hand, and interpolation or smoothing type operators on the other hand. We see that a prediction type operator has an inherent one-sidedness in that it operates on the past values of a time series, but not on the future values. Consequently a prediction type operator is computationally realizable, and may represent the impulsive response of a realizable linear system.

A smoothing type operator, however, operates on both past and future values of a time series, and thereby is not computationally realizable. Nevertheless, finite smoothing operators can be made realizable by introducing a time delay

in the computations so that all the necessary data is available at the time the computations are to be carried out. The fact that a seismic disturbance is recorded on paper in the form of a seismogram means that we have waited until all the pertinent information is available. Consequently the necessary time delay has been introduced to utilize finite time-delay smoothing operators, which are computationally realizable. Since a time-delay smoothing operator has the same mathematical form as a prediction type operator, in the remaining parts of this thesis we deal chiefly with operators of the prediction type.

The transfer function is defined as the Fourier transform of the linear operator, and corresponds to the transfer function, the system function, or filter characteristics of a linear system. By analytic continuation we may extend the transfer function into the complex plane, where the real axis represents real angular frequency. Then we see that the transfer function of a prediction type operator has no singularities below the axis of real frequency, which is indicative of the realizability of such an operator.

We state the condition that the linear difference equation formed by the coefficients of a finite prediction type operator be a stable difference equation, that is, its gen-

eral solution be a damped oscillation. We show that this stability condition is precisely the condition that the transfer function have zeros all of which lie above the axis of real frequency. Thus a realizable and stable prediction type operator has a transfer function with no singularities or zeros below the axis of real frequency, which is the requirement that its phase characteristic be of the minimum phase-shift type familiar to electrical engineers. Moreover, we extend this concept of stability to prediction type operators with an infinite number of coefficients.

We show that each stable prediction type operator has a unique inverse prediction type operator which is also stable, and that their respective transfer functions are reciprocals of each other. We see that the inverse operator may be readily computed in the time domain from a given linear operator.

Finally in Chapter II we show that in order to design a stable prediction type operator one should utilize only the absolute gain characteristics of the desired filtering properties, and not phase characteristics. That is, the phase characteristic of the resulting stable operator should be the minimum phase characteristic which is determined uniquely from the absolute gain characteristic. We give a direct

computational procedure which may be readily programmed for a digital computer.

In time series analysis, there are two lines of approach which we may call the non-statistical and the statistical. In the non-statistical approach the given time series is interpreted as a mathematical function, and in the statistical approach as a random specimen out of a population of mathematical functions.

In Chapter III, we treat the methodology used in the non-statistical or deterministic approach to the study of time series. Such an approach leads to a perfect functional representation of the observational or theoretical data. In particular, the methods of this chapter are applicable to the determination of linear operations on transient time functions of finite energy, such as seismic wavelets. Further we observe that even under a deterministic hypothesis it may be necessary to utilize methods of averaging for certain purposes such as approximations. Although methods of averaging may be developed without recourse to the theory of probability, in many applications, it is not until probability theory is introduced that certain averaging operations become meaningful in a physical sense. With this situation in mind, in the following chapters we consider the statistical approach to the

study of time series in which methods of averaging play a central role.

In Chapter IV, we present concepts from the theory of discrete stationary time series which represents a statistical approach to the study of time series. We consider stationary stochastic or random processes which generate stationary time series, and give properties of the autocorrelation and spectrum. In particular we consider time series which are "white noise". We see that any time series with an absolutely continuous spectral distribution is a process of moving summation.

In Chapter V, we give an heuristic exposition of the mehtod of the factorization of the power spectrum. We show how this factorization leads to the predictive decomposition of a stationary time series. The Predictive Decomposition Theorem shows that any stationary time series (with an absolutely continuous spectral distribution) can be considered to be additively composed of many overlapping wavelets. All These wavelets have the same stable shape or form; and the arrival times and strengths of these wavelets are random and uncorrelated with each other. If the wavelet shape represents the impulsive response of a stable linear system, and if the uncorrelated arrival times and strengths represent a "white

noise" input to the linear system, then the stationary time series represents the output of the linear system.

We then show that the solutions of the prediction and filtering problems for single time series follows directly from the Predictive Decomposition Theorem. We examine those stochastic processes which have power spectra which are rational functions, and see that the autoregressive process and the process of finite moving averages are special cases of such processes. We deal with the theory of multiple time series, in which we see that the concept of coherency plays an important role, and we treat the general technique of discrete prediction for multiple time series.

In Chapter VI we deal with applications to seismic exploration. In particular we consider the situation in which a given section of seismic trace (recorded with automatic volume control) is additively composed of seismic wavelets, where each wavelet has the same stable shape or form, and where the strengths and arrival times of the wavelets are random and uncorrelated with each other. Under these assumption, the Predictive Decomposition Theorem tells us that the section of trace may be considered to be a section of a stationary time series.

To illustrate the probabilistic approach, we consider the problem in which we wish to separate the dynamic component (the wavelet shape) from the random components (the arrival times and strengths of the wavelets).

The theoretical solution of this problem consists of the following steps:

(1) Average out the random components of the trace so as to yield the wavelet shape.

(2) From the wavelet shape thus found, compute the inverse wavelet shape, or prediction operator for unit prediction distance. Apply this prediction operator to the trace in order to yield the random components, which are the prediction errors. That is, the prediction operator contracts the wavelets to impulses, which are the prediction errors. If one wishes to filter the seismic trace, one further step is added, namely:

(3) Reaverage the prediction errors by means of a stable linear operator so as to approximate the desired output.

The practical solution of this problem consists of the following steps:

(1) Compute the prediction operator, or inverse wavelet shape, directly by the Gauss method of least squares which yields operator coefficients which satisfy certain

statistical optimum properties under general conditions.

(2) Use the prediction operator to determine the prediction errors which are the random unpredictable components.

Alternatively, from other considerations, one may have available the shape of the seismic wavelet. Then the prediction operator or inverse wavelet shape, may be readily computed from this wavelet shape. In summary, then, the prediction operator for unit prediction distance is the inverse seismic wavelet and the inverse prediction operator for unit prediction distance is the seismic wavelet.

Finally we note that multi-trace operators take into account the coherency between the traces, which is important in that seismic traces become more coherent at major reflections.

CHAPTER II

THEORY OF FINITE DISCRETE LINEAR OPERATORS

2.1 The Finite Discrete Linear Operator

In this chapter we wish to consider properties of the finite discrete linear operator for a single time series, that is, a linear operator which has a finite number of discrete operator coefficients which perform linear operations on single discrete time series.

In this thesis we deal almost exclusively with discrete time series. A discrete time series is a sequence of equidistant observations x_t which are associated with the discrete time parameter t. Without loss of generality we may take the spacing between each successive observation to be one unit of time, and thus we may represent the time series as

$$\ldots, x_{t-2}, x_{t-1}, x_t, x_{t+1}, x_{t+2}, \ldots \tag{2.11}$$

where t takes on integer values. As a result the minimum period which may be observed is equal to two units, and consequently the maximum frequency which may be observed is equal to $1/2$, which is an angular frequency of π.

Thus, we may require all frequencies f of the discrete time series to lie between $-1/2$ and $1/2$, and all angular frequencies $\omega = 2\pi f$ to lie between $-\pi$ and π.

Time series x_t may be finite or infinite in extent, and may be treated from a deterministic or statistical point of view. In this chapter we shall develop those properties of the finite discrete linear operator which are independent of the nature of the x_t time series, whereas in the following chapters we shall be mainly concerned with the nature of the x_t time series.

Following Kolmogorov (1939, 1941) we shall distinguish between extrapolation or prediction type operators on the one hand, and interpolation or smoothing type operators on the other hand.

2.2 Prediction Operators

The extrapolation or prediction operator (Kolmogorov, 1939, 1941) is given by

$$\hat{x}_{t+\alpha} = k_o x_t + k_1 x_{t-1} + \cdots k_M x_{t-M} = \sum_{s=o}^{M} k_s x_{t-s}, \quad \alpha \geq 0 \tag{2.21}$$

where α is the prediction distance. The operator coefficients, k_o, k_1,..., k_M are chosen so that the actual output, $\hat{x}_{t+\alpha}$, approximates the desired output, $x_{t+\alpha}$, according to some criterion. The operator is discrete since its coefficients are discrete values, and the operator is finite since there are only a finite number M of such coefficients.

In Section 5.3 we discuss the solution of the prediction problem in which the operator coefficients are determined by the least squares criterion. In general, we shall see that an infinite number of operator coefficients are required, although for autoregressive time series (see Section 5.5-B) only a finite number of operator coefficients are required.

The prediction operator (2.21) is one from which a specific time function may be generated if a sufficient number of initial values of the time function are specified. That is, if we are given the initial values

$$x_o, x_1, x_2 \cdots\cdots x_{M+\alpha-1}, \tag{2.22}$$

and if we define

$$\hat{x}_t = x_t \text{ for } t = M+\alpha,\ M+\alpha+1,\ M+\alpha+2,\ \ldots \tag{2.221}$$

then we may generate the time function x_t by means of equation (2.21). For example, let us consider the case given by

$$\hat{x}_{t+1} = .5x_t, \quad x_o = 1. \tag{2.23}$$

Then we may generate the time function

$$x_t = \hat{x}_t = .5x_{t-1} \quad \text{for } t = 1, 2, 3, \ldots \tag{2.24}$$

which is the sequence

$$(.5),(.5)^2,(.5)^3,(.5)^4,(.5)^5,\ldots(.5)^t, (.5)^{t+1},\ldots \tag{2.241}$$

In Section 2.6 we shall see that such a sequence must form a damped motion in order for the operator to be stable.

Further, the prediction operator (2.21) has the property that only the values x_t, x_{t-1}, x_{t-2},... of the time series at time t and prior to time t, and no values x_{t+1}, x_{t+2},..., subsequent to time t, are required in order to compute the actual output $\hat{x}_{t+\alpha}$. Thus a prediction operator has an inherent one-sidedness in that it operates on present and past values, but no future values, of the time series. As a result

if time t represents the present calender time, as it is the case for a meterologist who makes daily weather predictions, then only observations of the time series at the present time and at past times, and no observations at future times are required to carry out the necessary computation.

As we shall see in Section 2.5, the prediction operator coefficients represent the impulsive response of an equivalent electric network. Thus those impulses which have arrived at time t or prior to time t will make the network respond, whereas those impulses which have not yet arrived at time t, i.e., those impulses which arrive subsequent to time t, cannot make the network respond. In summary, then, we may say that a finite prediction operator is computationally realizable to the statistician, and physically realizable to the engineer.

Instead of considering the predicted values $\hat{x}_{t+\alpha}$ as the output of the prediction operator, one may consider the prediction error $\xi_{t+\alpha} = x_{t+\alpha} - \hat{x}_{t+\alpha}$ as the output (Wadsworth, et al, 1953). Then we have

$$\xi_{t+\alpha} = x_{t+\alpha} - \hat{x}_{t+\alpha} = x_{t+\alpha} - \sum_{s=0}^{M} k_s x_{t-s} \tag{2.25}$$

which may be rewritten

$$\xi_t = x_t - \sum_{s=o}^{M} k_s x_{t-s-\alpha} = \tag{2.26}$$

$$= x_t - k_o x_{t-\alpha} - k_1 x_{t-\alpha-1} - \cdots - k_M x_{t-\alpha-M}$$

Let us define $m = M+\alpha$ and

$$a_o = 1,\ a_1 = 0,\ a_2 = 0, \ldots a_{\alpha-1} = 0,\ a_\alpha = -k_o,$$

$$a_{\alpha+1} = -k_1, \ldots,\ a_{\alpha+M} = a_m = -k_M \tag{2.27}$$

Then the prediction error ξ_t may be written as

$$\xi_t = a_o x_t + a_1 x_{t-1} + \cdots + a_m x_{t-m} = \tag{2.271}$$

$$= \sum_{s=o}^{m} a_s x_{t-s},\ a_o = 1$$

In the sequel we shall be primarily concerned with the prediction operator for prediction distance α equal to one. Then equation (2.271) for the prediction error ($\alpha=1$) becomes

$$\xi_t = \sum_{s=o}^{m} a_s x_{t-s} = \tag{2.28}$$

$$= x_t + a_1 x_{t-1} + \cdots + a_m x_{t-m}, \; a_o = 1$$

and equation (2.27) becomes

$$a_o = 1, \; a_1 = -k_o, \; a_2 = -k_1, \cdots \; a_m = -k_M \tag{2.281}$$

We shall regard equation (2.28) as the basic form of the prediction type operator, the prediction error operator, or simply the prediction operator, with the coefficients a_o, a_1, $a_2, \ldots a_m$. The prediction error ξ_t shall be regarded as the actual output at time t of the operator (2.28).

In general, the operator coefficients are chosen so that the actual output approximates a certain desired output according to some criterion. In Section 5.4 we discuss the solution to the general filtering problem in which prediction type operators are used, the coefficients of which are chosen according to the least squares criterion.

In equation (2.28) we may let the operator coefficients represent the impulsive response of a network and the time series x_t the input. Then ξ_t is the actual output of the network.

The prediction operator (2.28) has the same realizability properties as operator (2.21), as may be readily verified.

2.3 Smoothing Operators

The interpolation or smoothing operator (Kolmogorov, 1939, 1941) is given by

$$\hat{x}_t = d_{-M}x_{t+M} + d_{-M+1}x_{t+M-1} + \cdots + d_{-1}x_{t-1}$$

$$+ d_1 x_{t-1} + d_2 x_{t-2} + \cdots + d_M x_{t-M}$$

$$= \sum_{\substack{s=-M \\ s\neq 0}}^{M} d_s x_{t-s} \tag{2.31}$$

We may consider the smoothing error given by

$$\gamma_t = x_t - \hat{x}_t = -\sum_{s=-M}^{1} d_s x_{t-s} + x_t - \sum_{s=1}^{M} d_s x_{t-s} \tag{2.32}$$

and by letting $m = M$, $c_o = 1$, and $c_s = -d_s$ ($s = \pm 1, \pm 2, \ldots \pm M$) we have

$$\gamma_t = \sum_{s=-m}^{m} c_s x_{t-s}, \quad c_o = 1 \tag{2.33}$$

In the sequel we shall regard this equation as the basic form of the smoothing type operator, the smoothing error operator,

or simply the smoothing operator, with operator coefficients c_{-m}, c_{-m+1},..., $c_o = 1$, c_1, ... c_m. The smoothing error γ_t is the actual output of the operator.

The smoothing operator does not have the same properties in regard to computational or physical realizability as the prediction operator. In particular, it is not possible to generate a time function from specified initial values by means of equation (2.31) as was done in the case of the prediction operator. The smoothing operator (2.33) has the property that values of the time series x_{t+m},..., x_{t+2}, x_{t+1} at times subsequent to time t, as well as values x_t, x_{t-1}, x_{t-2},... x_{t-m} at time t and prior to time t, are required to compute γ_t. Consequently, if time t represents the present calender time, as in the example of the meteorologist, then γ_t given by equation (2.33) can not be computed since it involves observations x_{t+1}, x_{t+2},... x_{t+m} at future times and which thereby are not observable at the present time t. Similarly, the network which would be equivalent to a smoothing type operator would be one which would respond to impulses which have not yet arrived at the present time t. Consequently, smoothing type operators are not computationally realizable to the statistician, or physically realizable to the engineer.

Nevertheless, a very simple trick makes finite smoothing

type operators computationally and physically realizable. That is, in order to compute

$$\gamma_t = \sum_{s=-m}^{m} c_s x_{t-s} = c_{-m} x_{t+m} + c_{-m+1} x_{t+m-1} + \cdots + c_{-1} x_{t+1}$$

$$+ c_o x_t + c_1 x_{t-1} + \cdots + c_m x_{t-m} \tag{2.33}$$

the statistician must delay his computations until time $t + m$, or later, which is a time delay of m or greater, at which time all the values needed in the computation will have occured. That is, the statistician delays his computations at least until time $t' = t + m$, and then computes

$$\gamma_t = \sum_{s=-m}^{m} c_s x_{t-s} = \sum_{s=-m}^{m} c_s x_{t'-m-s} \tag{2.34}$$

$$= c_{-m} x_{t'} + c_{-m+1} x_{t'-1} + \cdots + c_{-1} x_{t'-m-1}$$

$$+ c_o x_{t'-m} + c_1 x_{t'-m-1} + \cdots + c_m x_{t'-2m}$$

which we shall call the time-delay form of the smoothing operator with coefficients c_s ($s = 0, \pm 1, \ldots \pm m$). Such

an operator is in the form of the prediction type operator (2.28) with the present time now being t'. That is, t' is the time at which computations are to be carried out.

Similarly the engineer may introduce a time delay m, or greater, into his network, (Bode and Shannon, 1950) to transform the non-physically realizable system (at the time instant t)

$$\sum_{s=-m}^{m} c_s x_{t-s} \tag{2.35}$$

into the physically realizable system (at the time t' = t+m)

$$\sum_{s=-m}^{m} c_s x_{t'-m-s}. \tag{2.36}$$

The fact that a seismic disturbance is recorded on paper or magnetic tape means that we have waited until all the pertinent information is available. Consequently the necessary time delay has been introduced to utilize finite time-delay smoothing operators, which are computationally realizable. On the other hand if computations were to be carried out at the same time as the seismic disturbance is occuring, then the statistician would not be able to compute such smoothing type operations.

2.4 The Transfer Function or Filter Characteristics

As Wiener (1942) points out, the linear operator is the

approach from the standpoint of time to a filter which is essentially an instrument for the separation of different frequency ranges. The filtering action of the linear operator is brought out by its transfer function, which is the analogue of the transfer or system function of the linear system of which the linear operator is the unit impulse response.

Smith (1954) gives the following interpretation to the transfer function. Let the input information x_t be points from a sine wave of angular frequency ω. Since the system is linear, the output will be a sine wave of the same frequency but, in general, will differ in phase and amplitude. Using the complex notation for a sine wave, $x_t = e^{i\omega t}$, of angular frequency ω, the transfer ratio at angular frequency ω is the output of the linear operator, which is a complex sine wave of angular frequency ω, divided by the input $x_t = e^{i\omega t}$. The transfer function is the totality of these transfer ratios for $-\pi \quad \omega \quad \pi$, and represents the filter characteristics of the linear operator. As we shall now see the transfer function is the Fourier transform of the linear operator. Since the operator is discrete, the transfer function is in the form of a Fourier series rather than a Fourier integral.

For the prediction operator, equation (2.21),

$$\hat{x}_{t+\alpha} = \sum_{s=o}^{M} k_s x_{t-s} \ , \ \alpha \geq 0, \tag{2.21}$$

by letting $x_t = e^{i\omega t}$ we obtain the transfer ratio

$$\frac{\hat{x}_{t+\alpha}}{x_t} = \frac{\sum_{s=o}^{M} k_s x_{t-s}}{x_t} = \frac{\sum_{s=o}^{M} k_s e^{i\omega(t-s)}}{e^{i\omega t}} = \sum_{s=o}^{M} k_s e^{-i\omega s} \tag{2.41}$$

The totality of these transfer ratios yields the transfer function

$$K(\omega) = \sum_{s=o}^{M} k_s \ e^{-i\omega s} \tag{2.441}$$

which is the Fourier transform of the operator coefficients k_s.

By letting $x_t = e^{i\omega t}$ be the input for the operator for the prediction error, equation (2.28),

$$\xi_t = \sum_{s=o}^{m} a_s x_{t-s} \ , \ a_o = 1, \tag{2.28}$$

we obtain the transfer ratio

$$\frac{\xi_t}{x_t} = e^{-i\omega t} \sum_{s=o}^{m} a_s e^{i\omega(t-s)} = \sum_{s=o}^{m} a_s e^{-i\omega s} \qquad (2.42)$$

The transfer function is then

$$A(\omega) = \sum_{s=o}^{m} a_s e^{-i\omega s} \qquad (2.421)$$

which is the Fourier transform of the operator coefficients.

For the smoothing operator, equation (2.31),

$$\hat{x}_t = \sum_{\substack{s=-M \\ s\neq o}}^{M} d_s x_{t-s}, \qquad (2.31)$$

by letting the imput be $x_t = e^{i\omega t}$, the transfer ratio is

$$\frac{\hat{x}_t}{x_t} = e^{-i\omega t} \sum_{\substack{s=-M \\ s\neq o}}^{M} d_s e^{i\omega(t-s)} = \sum_{\substack{s=-M \\ s\neq o}}^{M} d_s e^{-i\omega s} \qquad (2.43)$$

and the transfer function is

$$D(\omega) = \sum_{\substack{s=-M \\ s\neq o}}^{M} d_s e^{-i\omega s} \qquad (2.431)$$

Similarly letting $x_t = e^{i\omega t}$ be the input for the operator for the smoothing error

$$\gamma_t = \sum_{s=-m}^{m} c_s \, x_{t-s} \tag{2.33}$$

we obtain the transfer ratio

$$\frac{\gamma_t}{x_t} = e^{-i\omega t} \sum_{s=-m}^{m} c_s e^{i\omega(t-s)} = \sum_{s=-m}^{m} c_s e^{-i\omega s} \tag{2.44}$$

so that the transfer function is

$$C(\omega) = \sum_{s=-m}^{m} c_s e^{-i\omega s} . \tag{2.441}$$

Using the same operator coefficients c_s ($s = 0, \pm 1, \ldots \pm m$) but introducing a time delay m, the time-delay smoothing operator, equation (2.34),

$$\gamma_t = \sum_{s=-m}^{m} c_s \, x_{t'-m-s} \tag{2.34}$$

is realizable at the time instant $t' = t+m$. Since t' now represents the time instant at which computations are to be carried out the input is $x_{t'} = e^{i\omega t'}$. The transfer ratio is

$$\frac{\sum_{s=-m}^{m} c_s \, e^{i\omega(t'-m-s)}}{e^{i\omega t'}} = \sum_{s=-m}^{m} c_s e^{-i\omega(s+m)} \tag{2.45}$$

and its transfer function is

$$\sum_{s=-m}^{m} c_s e^{-i\omega(s+m)} = e^{-i\omega m} \sum_{s=-m}^{m} c_s e^{-i\omega s} = e^{-i\omega m} C(\omega). \quad (2.451)$$

A summary of the various linear operator forms and their transfer functions is given in Figures 1 and 2.

2.5 Realizability of Linear Operators and their Relationship to Electric Networks

We now wish to summarize those parts of the preceding sections concerning the realizability of linear operators. It was seen that operators of the prediction type, for example, as represented by the prediction error operator (2.28)

$$\xi_t = \sum_{s=0}^{m} a_s x_{t-s}, \quad (2.28)$$

are computationally realizable at the time instant t. On the other hand, operators of the smoothing type, for example, as represented by the smoothing error operator (2.33)

$$\gamma_t = \sum_{s=-m}^{m} c_s x_{t-s} \quad (2.33)$$

are not computationally realiable at the time instant t. Nevertheless by delaying the computations at least to the

	Operator Form
Prediction Operator (2.21) (Computationally realizable at time instant t)	$x_{t+\alpha} = \sum_{s=o}^{M} k_s x_{t-s}, \ \alpha \geq 0, \ M \geq 0$
Prediction Error Operator (2.28) (Computationally realizable at time instant t)	$\xi_t = \sum_{s=o}^{m} a_s x_{t-s}, \ m \geq 0, \ a_o = 1$
Smoothing Operator (2.31) (Not computationally realizable at time instant t)	$x_t = \sum_{\substack{s=-M \\ s \neq o}}^{M} d_s x_{t-s}, \ M \geq 0$
Time Delay Smoothing Operator (Computationally realizable at time instant $t'=t+M$)	$x_t = \sum_{\substack{s=-M \\ s \neq o}}^{M} d_s x_{t'-M-s}, \ M \geq 0$
Smoothing Error Operator (2.33)(Not computationally realizable at time instant t)	$\gamma_t = \sum_{s=-m}^{m} c_s x_{t-s}, \ m \geq 0, \ c_o = 1$
Time Delay Smoothing Error Operator (2.34) (Computationally realizable at time instant $t' = t+m$)	$\gamma_t = \sum_{s=-m}^{m} c_s x_{t'-m-s}, \ c_o = 1$

Figure 1. Various Types of Finite Discrete Linear Operators for Single Time Series

	Transfer Function
Prediction Operator (2.21)	$K(\omega) = \sum_{s=o}^{M} k_s e^{-i\omega s}$
Prediction Error Operator (2.28)	$A(\omega) = \sum_{s=o}^{m} a_s e^{-i\omega s}$
Smoothing Operator (2.31)	$D(\omega) = \sum_{\substack{s=-M \\ s\neq 0}}^{M} d_s e^{-i\omega s}$
Time-delay Smoothing Operator	$\sum_{\substack{s=-M \\ s\neq o}}^{M} d_s e^{-i\omega(s+M)} = e^{-i\omega M} D(\omega)$
Smoothing Error Operator (2.33)	$C(\omega) = \sum_{s=-m}^{m} c_s e^{-i\omega s}$
Time-delay Smoothing Error Operator (2.34)	$\sum_{s=-M}^{M} c_s e^{-i\omega(s+m)} = e^{-i\omega m} C(\omega)$

Figure 2. Transfer Functions of Various Types of Finite Linear Operators Given in Figure 1.

time instant $t' = t+m$, one may compute

$$\gamma_t = \sum_{s=-m}^{m} c_s x_{t'-m-s} \cdot \qquad (2.34)$$

This later form of the smoothing operator, which is called the time-delay smoothing operator, is therefore computationally realizable, and indeed has the same form as a prediction type operator (2.28).

The realizability of these operator forms are reflected in their respective transfer functions as follows. Let us consider the complex plane $\lambda = \omega + i\,\sigma$, where we let the real ω axis denote the angular frequency ω. By analytic continuation, the transfer function of the prediction type operator becomes in the complex λ-plane

$$A(\lambda) = \sum_{s=o}^{m} a_s e^{-i\lambda s} = \sum_{s=o}^{m} a_s e^{\sigma s}\, e^{-i\omega s} \,. \qquad (2.51)$$

By examining this equation, we see that $A(\lambda)$ has no singularities in the lower half λ-plane, that is, for $\sigma < 0$, which reflects the realizability of the prediction type operator.

On the other hand, the transfer function of the smoothing type operator in the complex plane is

$$C(\lambda) = \sum_{s=-m}^{m} c_s e^{-i\lambda s} = \sum_{s=-m}^{-1} c_s e^{-i\lambda s} + \sum_{s=o}^{m} c_s e^{-i\lambda s}. \qquad (2.52)$$

By letting $r = -s$ for $s = -m, -m+1, \ldots -2, -1$, the transfer function becomes

$$C(\lambda) = \sum_{r=1}^{m} c_r e^{-\sigma r} e^{i\omega r} + \sum_{s=o}^{m} c_s e^{\sigma s} e^{-i\omega s}, \qquad (2.521)$$

which has singularities in the upper half λ-plane and in the lower half λ-plane, thereby reflecting the non-realizability of the smoothing type operator.

The transfer function of the time-delay smoothing operator,

$$e^{-i\lambda m} C(\lambda) = \sum_{s=-m}^{m} c_s e^{-i\lambda(s+m)} =$$

$$= \sum_{s+m=o}^{2m} c_s e^{\sigma(s+m)} e^{-i\omega(s+m)} \qquad (2.53)$$

has no singularities in the lower half λ-plane ($\sigma < 0$), which reflects the realizability of such an operator.

The computationally realizable linear operator corresponds to a physically realizable passive lumped element network together with a single amplifier (Bode and Shannon, 1950).

The linear operator coefficients represent the impulsive response of the network, and the transfer function represents the transfer or system function of the network (Levinson, 1947; Smith, 1954). For example, in equation (2.28) the operator coefficient a_s may be interpreted as the output obtained from an electric filter at time t+s in response to a unit impulse impressed upon the input of the filter at time t. Since $a_s = 0$ for $s < 0$ the output obtained from the filter is zero for times less than t, and since $a_s = 0$ for $s > t + m$ the output is zero for times greater than t+m, the filter is physically realizable. The transfer or system function of the electric filter is the transfer function of the linear operator,

$$A(\omega) = \sum_{s=o}^{m} a_s e^{-i\omega s}. \qquad (2.421)$$

Smoothing type operators, on the other hand, are not realizable to the statistician unless he introduces a time delay in his computations, or to the engineer unless he introduces a time delay in his network. Nevertheless, for problems in which physical calendar time is not important, for example, as in the analysis of seismic records, one may make computations based on the time-delay smoothing operator,

but consider these computations from the point of view of the smoothing operator itself. In other words, computations may be carried out with respect to the realizable transfer function

$$e^{-i\lambda m} \; C(\lambda) \tag{2.53}$$

of the time-delay smoothing operator, but considered as if they were carried out with respect to the non-realizable transfer function

$$C(\lambda) \tag{2.521}$$

of the smoothing operator with the same operator coefficients. This same procedure is available to the engineer in such cases (Bode and Shannon, 1950).

The prediction operator

$$\xi_t = \sum_{s=o}^{m} a_s x_{t-s} \tag{2.28}$$

has the transfer function

$$A(\omega) = \sum_{s=o}^{m} a_s e^{-i\omega s} \tag{2.421}$$

with real part

$$\text{Re } [A(\omega)] = \sum_{s=o}^{m} a_s \cos \omega s \qquad (2.541)$$

and imaginary part

$$\text{Im } [A(\omega)] = \sum_{s=o}^{m} a_s \sin \omega s. \qquad (2.542)$$

Since both the real and imaginary parts of the transfer function depend on the same variables, namely the operator coefficients a_o, a_1,... a_m, the real part Re $[A(\omega)]$ and the imaginary part Im $[A(\omega)]$ cannot be chosen independently (Smith, 1954). In other words, knowledge of Re $[A(\omega)]$ lets us compute the values of the operator coefficients a_s ($s = 0, 1, \ldots m$) by means of the equation

$$\begin{aligned} \frac{1}{\pi} \int_0^{\pi} \text{Re}[A(\omega)]\, e^{i\omega t}\, d\omega &= \\ &= \frac{1}{\pi} \int_0^{\pi} \sum_{s=o}^{m} a_s \cos \omega s \cos \omega t \, d\omega = \\ &= a_t, \text{ for } t = 0, 1, 2, \ldots m. \end{aligned} \qquad (2.543)$$

With the values of the operator coefficients a_o, a_1, ... a_m thus found, the imaginary part Im$[A(\omega)]$ may be computed by means of equation (2.542).

On the other hand the smoothing error operator, say

$$\gamma_t = \sum_{s=-m}^{m} c_s x_{t-s}, \quad m \geq 0 \tag{2.33}$$

has the transfer function

$$C(\omega) = \sum_{s=-m}^{m} c_s e^{-i\omega s}, \tag{2.441}$$

with real part

$$\mathrm{Re}[C(\omega)] = \sum_{s=-m}^{m} c_s \cos \omega s = c_o + \sum_{s=1}^{m} (c_s + c_{-s}) \cos \omega s, \tag{2.551}$$

and imaginary part

$$\mathrm{Im}[C(\omega)] = \sum_{s=-m}^{m} c_s \sin \omega s = \sum_{s=1}^{m} (c_s - c_{-s}) \sin \omega s. \tag{2.552}$$

Thus the real part of the transfer function depends only on the symmetric component, $(c_s + c_{-s})$, of the smoothing operator, and the imaginary part of the transfer function depends only on the antisymmetric component, $(c_s - c_{-s})$, of the smoothing operator. Since the symmetric component, $(c_s + c_{-s})$, is linearly independent of the antisymmetric component, $(c_s - c_{-s})$, for each value of $s = 0,1,2,\ldots m$, we see that the real part,

Re[C(ω)], may be chosen independently of the imaginary part, Im[C(ω)], of the transfer function (Smith, 1954).

Now for the smoothing operator (2.33), let us suppose that the computations are to be carried out with respect to the time t' where t' = t+p. That is, since t = t' - p, we have

$$\gamma_t = \sum_{s=-m}^{m} c_s x_{t'-p-s} = \sum_{s=-m}^{m} c_s x_{t'-(p+s)}. \qquad (2.56)$$

Letting $x_{t'} = e^{i\omega t'}$, the transfer ratio is

$$\frac{1}{e^{i\omega t'}} \sum_{s=-m}^{m} c_s e^{i\omega[t'-(p+s)]} = \sum_{s=-m}^{m} c_s e^{-i\omega(p+s)} \qquad (2.561)$$

and so the transfer function of the operator (2.56), denoted by $C_p(\omega)$, is

$$C_p(\omega) = \sum_{s=-m}^{m} c_s e^{-i\omega(s+p)}. \qquad (2.562)$$

Now when p = 0, the operator (2.56) is the smoothing operator (2.33), with transfer function (2.441), the real and imaginary parts of which are independent of each other. On the other hand, when p = m, the operator (2.56) is the time-delay smoothing operator (2.34), which is a prediction type operator, with transfer function (2.451), the real and imaginary parts of

which are entirely dependent upon each other. For those p for which $0 < p < m$, the operator (2.56), has transfer function (2.562), the real and imaginary parts of which are partially dependent upon each other, and partially independent of each other. We shall now examine to what extent they are dependent and independent.

We see that $C_p(\omega)$, given by equation (2.562), is

$$C_p(\omega) = \sum_{s=-m}^{m} c_s e^{-i\omega(s+p)} = e^{-i\omega p} \sum_{s=-m}^{m} c_s e^{-i\omega s} = e^{-i\omega p}\, C(\omega) \tag{2.563}$$

where $C(\omega)$ is the transfer function (2.441) of the smoothing operator (2.33). For $0 < p < m$, the real part of $C_p(\omega)$ is given by

$$\mathrm{Re}[C_p(\omega)] = c_{-p} + \sum_{\ell=1}^{m-p} (c_{-\ell-p}+c_{\ell-p}) \cos \omega\ell + \sum_{\ell=m-p+1}^{m+p} c_{\ell-p} \cos \omega\ell \tag{2.564}$$

and the imaginary part is given by

$$\mathrm{Im}[C_p(\omega)] = \sum_{\ell=1}^{m-p} (c_{-\ell-p}-c_{\ell-p}) \sin \omega\ell + \sum_{\ell=m-p+1}^{m+p} c_{\ell-p} \sin \omega\ell \tag{2.565}$$

Thus, given $\mathrm{Re}[C_p(\omega)]$, we may compute c_{-p}, $(c_{-\ell-p} + c_{\ell-p})$ for $\ell = 1,2,\ldots m-p$, and $c_{\ell-p}$ for $\ell = m-p+1, m-p+2,\ldots m+p$. The values $(c_{-\ell-p} - c_{\ell-p})$ for $\ell = 1,2,\ldots m-p$, which enter into equation (2.565) for the imaginary part $\mathrm{Im}[C_p(\omega)]$ are independent of the values $(c_{-\ell-p} + c_{\ell-p})$ for $\ell = 1,2,\ldots, m-p$, which were computed from the real part $\mathrm{Re}[C_p(\omega)]$, and thus reflect the partial independence of the real and imaginary parts of the transfer function. On the other hand, the values $c_{\ell-p}$ for $\ell = m-p+1, m-p+2,\ldots, m+p$, which enter into equation (2.565) for the imaginary part $\mathrm{Im}[C_p(\omega)]$ are the same values $c_{\ell-p}$ for $\ell = m-p+1$, $\ell = m-p+2,\ldots m+p$ which were computed from the real part $\mathrm{Re}[C_p(\omega)]$, and thus reflect the partial dependence of the imaginary part on the real part of the transfer function.

2.6 The Stable Prediction Operator

In the remaining sections of this chapter we shall consider only the operator form of the prediction type, namely

$$\xi_t = \sum_{s=o}^{m} a_s x_{t-s}. \tag{2.28}$$

Let us now consider the linear difference equation

$$\sum_{s=o}^{m} a_s x_{t-s} = 0 \tag{2.61}$$

obtained from equation (2.28) by requiring $\xi_t = 0$. We see that the constant coefficients a_s of the difference equation are the operator coefficients. There is no loss of generality in assuming that a_o is equal to one, in conformity with our usual convention.

The theory of difference equations is presented by various authors, and the reader is referred especially to Wold (1938) and Samuelson (1947). In the first part of this section we state the condition that the difference equation (2.61) be stable, that is, the condition that its solution x_t describes a damped oscillation. Then in the last part of this section we show that this stability condition is precisely the condition that Fourier transform of the operator (i.e. the transfer function $A(\lambda)$) has zeros and singularities, all of which lie in the lower half λ-plane, where $\lambda = \omega + i\sigma$. In other words, we show that the difference equation (2.61) is stable if the transfer function of the operator has minimum phase shift characteristic. Let us now examine the condition that the difference equation (2.61) be stable. The characteristic equation of the difference equation (2.61) is defined to be

$$P(\zeta) = a_o\zeta^m + a_1\zeta^{m-1} + \ldots + a_{m-1}\zeta + a_m = \sum_{s=o}^{m} a_s\zeta^{m-s}. \quad (2.62)$$

Since $P(\zeta)$ is a polynomial of degree m, it follows from the fundamental theorem of algebra that $P(\zeta)$ has m roots or zeros ζ_j such that

$$P(\zeta_j) = 0, \text{ for } j = 1,2,\ldots m \tag{2.621}$$

As a result $P(\zeta)$ may be written in the form

$$P(\zeta) = a_o(\zeta-\zeta_1)(\zeta-\zeta_2)\ldots(\zeta-\zeta_m) \tag{2.622}$$

Since the operator coefficients $a_s(s = 0,1,\ldots m)$ are real, the roots or zeros $\zeta_1, \zeta_2,\ldots\zeta_m$ must be real or occur in complex conjugate pairs. Let the distinct real roots of $P(\zeta)$ be represented by α_j, $j = 1,2,3\ldots h$ where each distinct root α_j is repeated γ_j times $(j = 1,2,3\ldots h)$; that is, the zero α_j is a zero of order γ_j. Let the distinct complex roots and their conjugate roots be represented by $\beta_j e^{i\Theta j}$ and $\beta_j e^{-i\Theta j}$, $(j = 1,2,\ldots k)$ where each distinct complex root is repeated ρ_j times $(j = 1,2,\ldots k)$; that is, the zero $\beta_j e^{i\Theta j}$ is a zero of order ρ_j, and the zero $\beta_j e^{-i\Theta j}$ is also a zero of order ρ_j. Here β_j represents the modulus, and Θ_j or $-\Theta_j$ represents the argument, of the complex root. Consequently, equation (2.62) becomes

$$P(\xi) = a_o\xi^m + a_1\xi^{m-1} + \ldots + a_{m-1}\,\xi + a_m$$

$$= a_o(\xi - \xi_1)(\xi - \xi_2)\ldots(\xi - \xi_m) = a \prod_{j=1}^{m} (\xi - \xi_j) \tag{2.623}$$

$$= a_o \prod_{j=1}^{h} (\xi-\alpha_j)^{\gamma_j} \prod_{j=1}^{k} (\xi-\beta_j e^{i\theta_j})^{\rho_j} (\xi-\beta_j e^{-i\theta_j})^{\rho_j}$$

$$= a_o \prod_{j=1}^{h} (\xi-\alpha_j)^{\gamma_j} \prod_{j=1}^{k} (\xi^2 - 2\beta_j \cos\theta_j\, \xi + \beta_j^2)^{\rho_j} .$$

For an arbitrary set of initial values

$$x_o,\ x_1,\ x_2,\ldots x_{t-1}, \tag{2.63}$$

the series x_t, x_{t+1}, x_{t+2},... may be generated by recursive deductions from the difference equation (2.61), and this series will form the general solution of the difference equation. Explicitly the general solution is given by

$$x_t = \sum_{j=1}^{h} P^{(1)}_{\gamma_j-1}(t)\ \ p_j^t + \sum_{j=1}^{k} [P^{(2)}_{\rho_j-1}(t) \cos\theta_j t + P^{(3)}_{\rho_j-1}(t) \sin\theta_j t]\beta_j^t \tag{2.631}$$

where the $P_r^{(s)}(t)$ denotes a polynomial of order r with arbitrary coefficients, and the h, k, α_j, β_j, γ_j, and ρ_j are

given by the characteristic equation (2.623).

The asymptotic behavior of the general solution x_t in equation (2.631) is dependent on the exponential factors α_j^t and β_j^t. A necessary and sufficient condition that

$$\sum_{t=o}^{\infty} x_t^2 \quad \text{and} \quad \sum_{t=o}^{\infty} |x_t| \tag{2.64}$$

converge for any values taken on by the $P_r^{(s)}(t)$ is that the magnitude of the roots $|\zeta_j|$, $j = 1,2,\ldots m$, of the characteristic equation (2.623) be less than one, that is, $|\zeta_j| < 1$, $j = 1,2,\ldots m$, which is

$$\begin{aligned} |\alpha_j| &< 1, \; j = 1,2,\ldots h \\ \beta_j &< 1, \; j = 1,2,\ldots k \; . \end{aligned} \tag{2.641}$$

In this case the solution x_t of the difference equation (2.61) describes a damped oscillation, and we shall call the corresponding linear operator

$$\xi_t = \sum_{s=o}^{m} a_s x_{t-s}, \; a_o = 1 \tag{2.28}$$

mathematically stable. That is, a linear operator is stable if the zeros ζ_k of its associated characteristic equation $P(\zeta)$ lie within the periphery $|\zeta| = 1$ of the unit circle, that is

$|\xi_j| < 1$.

Let us now turn our attention to the Fourier transform of the operator coefficients, that is, the transfer function,

$$A(\omega) = \sum_{s=o}^{m} a_s e^{-i\omega s} \tag{2.65}$$

of the linear operator (2.28). By analytic continuation to the complex plane $\lambda = \omega + i\,\sigma$, where the real ω axis denotes angular frequency, the transfer function becomes

$$A(\lambda) = \sum_{s=o}^{m} a_s e^{-i\lambda s} = \sum_{s=o}^{m} a_s e^{\sigma s}\, e^{-i\omega s}\,. \tag{2.651}$$

We see that $A(\lambda)$ is an entire transcendental function since it is a finite sum of entire transcendental functions $a_s e^{-i\lambda s}$ $(s = 1,2,\ldots m)$ plus a constant a_o. Consequently $A(\lambda)$ is analytic in the whole λ plane and may be represented by a power series in λ which converges in the whole plane. Outside of every circle in the λ plane, $A(\lambda)$ come arbitrarily close to every value; that is, $A(\lambda)$ has an essential singularity at the point at infinity.

Let us apply the transformation $z = e^{-i\lambda}$, for $-\pi < \omega \leq \pi$, to the transfer function $A(\lambda)$. The transformation $z = e^{-i\lambda}$ maps the strip between $\omega = -\pi$ to $\omega = \pi$ of the upper half λ

plane ($\sigma < 0$) into the exterior $|z| < 1$ of the unit circle in the z plane; it maps strip between $\omega = -\pi$ to $\omega = \pi$ of the lower half of the λ plane, ($\sigma < 0$) into the interior $|z| < 1$ of the unit circle in the z plane; and it maps the real axis $-\pi < \omega \leq \pi$ of the λ plane into the periphery $|z| = 1$ of the unit circle in the z plane. See Figure 3. Under this transformation the transfer function $A(\lambda)$, equation (2.651) becomes the polynomial $A(z)$ where

$$A(z) = \sum_{s=o}^{m} a_s z^s = a_o + a_1 z + a_2 z^2 + \ldots + a_m z^m. \qquad (2.652)$$

This polynomial or entire rational function, is analytic in the whole plane and has a pole at infinity. Let us call $A(z)$ the transfer function in the z-plane.

For a stable linear operator we have seen that the characteristic equation

$$P(\zeta) = \sum_{s=o}^{m} a_s \zeta^{m-s} = a_o \zeta^m + a_1 \zeta^{m-1} + \ldots + a_m, \; a_o = 1 \qquad (2.623)$$

has roots $\zeta_1, \zeta_2 \ldots \zeta_m$, all of which have modulus $|\zeta_k|$ less than one. Without loss of generality, we assume $a_m \neq 0$, so that $|\zeta_k| \neq 0$ for $k = 1,2,\ldots m$. Thus the characteristic equation (2.623) may be written

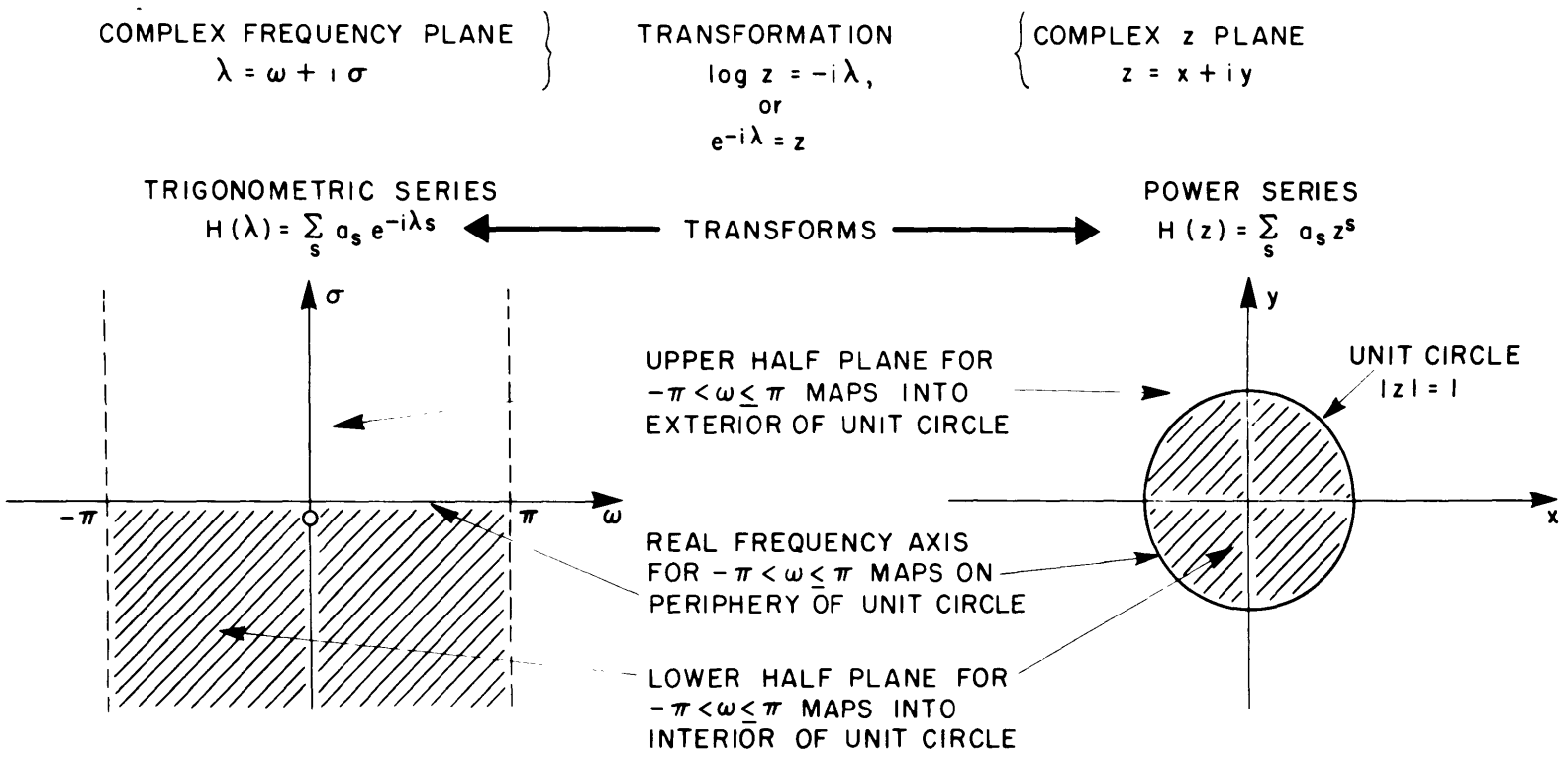


FIGURE 3 TRANSFORMATION OF TRIGONOMETRIC SERIES INTO POWER SERIES

$$P(\zeta) = (\zeta-\zeta_1)(\zeta-\zeta_2)\ldots(\zeta-\zeta_m) \tag{2.653}$$

$$= \zeta^m \zeta_1\zeta_2\ldots\zeta_m (\zeta_1^{-1} - \zeta^{-1}) (\zeta_2^{-1} - \zeta^{-1}) \ldots (\zeta_m^{-1} - \zeta^{-1})$$

$$= \zeta^m \zeta_1\zeta_2\ldots\zeta_m (-1)^m (\zeta^{-1} - \zeta_1^{-1}) (\zeta^{-1} - \zeta_2^{-1})\ldots(\zeta^{-1} - \zeta_m^{-1})$$

$$= a_m\zeta^m (\zeta^{-1} - \zeta_1^{-1}) (\zeta^{-1} - \zeta_2^{-1})\ldots(\zeta^{-1} - \zeta_m^{-1})$$

since

$$a_m = (-1)^m \zeta_1\zeta_2\ldots\zeta_m. \tag{2.654}$$

Under the transformation $z = \zeta^{-1}$, the function $P(\zeta)$, given by equation (2.653), becomes

$$P(\zeta = z^{-1}) = \sum_{s=o}^{m} a_s z^{-m+s} = z^{-m} a_m(z-z_1)(z-z_2)\ldots(z-z_m) \tag{2.655}$$

where we define

$$z_1 = \zeta_1^{-1},\ z_2 = \zeta_2^{-1},\ \ldots z_m = \zeta_m^{-1}. \tag{2.656}$$

Thus the function

$$z^m P(\zeta = z^{-1}) = \sum_{s=o}^{m} a_s z^s = a_m(z-z_1)\ (z-z_2)\ldots(z-z_m) \quad (2.657)$$

is the function A(z) in equation (2.652) which was obtained from the transfer function $A(\lambda)$ by the transformation $z = e^{-i\lambda}$. That is, the transfer function in the z-plane is

$$A(z) = \sum_{s=o}^{m} a_s z^s = a_m(z-z_1)(z-z_2)\ldots(z-z_m) \quad (2.657)$$

where the roots $z_k (k = 1,2,\ldots m)$ are given by equation (2.656). Since the stability condition is that the roots ζ_k have modulus $|\zeta_k|$ less than one, we see that this condition is that the roots z_k of A(z) have modulus

$$|z_k| = |\zeta_k^{-1}| > 1. \quad (2.658)$$

greater than one.

Thus the finite prediction operator is stable if the roots z_k of the polynomial A(z) all have modulus greater than one. That is, the operator with coefficients $a_o, a_1, \ldots, a_m$ is stable if the transfer function in the z-plane,

$$\begin{aligned} A(z) &= a_o + a_1 z + a_2 z^2 + \ldots + a_m z^m \\ &= a_o(z-z_1)(z-z_2)\ldots(z-z_m) \end{aligned} \quad (2.66)$$

has roots z_k where

$$|z_k| > 1, \quad k = 1,2,\ldots,m. \tag{2.661}$$

In other words, the stability requirement is that all the zeros of A(z) must lie exterior to the periphery $|z| = 1$ of the unit circle in the z plane.

For example, let us consider the so-called cosine operator (Simpson, 1953)

$$\begin{aligned} \xi_t &= x_t + a_1 x_{t-1} + a_2 x_{t-2} \\ &= x_t - (2 \cos \omega_o)\, x_{t-1} + x_{t-2}. \end{aligned} \tag{2.67}$$

The coefficients of this operator were chosen by the requirement that the prediction error $\xi_t = 0$ for $x_t = \cos \omega_o t$. The transfer function in the z-plane is then

$$\begin{aligned} A(z) &= 1 + a_1 z + a_2 z^2 \\ &= 1 - (2 \cos \omega_o)\, z + z^2 \\ &= (e^{i\omega_o} - z)(e^{-i\omega_o} - z) \end{aligned} \tag{2.671}$$

so that the roots of A(z) are

$$z_1 = e^{i\omega_o} \ , \ z_2 = e^{-i\omega_o} \ . \qquad (2.672)$$

Since $|z_1| = |z_2| = 1$, the cosine operator is just on the borderline of unstability.

Let us consider this stability condition (2.661) in terms of the transfer function $A(\lambda)$ in the complex λ-plane. The stability condition (2.661) becomes, under the transformation $e^{-i\omega} = z$, that the transfer function

$$A(\lambda) = \sum_{s=o}^{m} a_s e^{-i\lambda s} \qquad (2.651)$$

has zeros only in the upper half λ plane. See Figure 3. Then $A(\lambda)$ has no singularities or zeros below the axis of real frequency ω and its logarithm in that half plane is as small as possible at infinity.

A linear system has minimum phase-shift characteristic if its transfer function has no singularities or zeros in the lower half λ-plane (Bode and Shannon, 1950; Goldman, 1953). Thus if we consider $A(\lambda)$ to be a transfer function of a linear system and if $A(\lambda)$ satisfies the conditions of a minimum phase-shift characteristic, then the linear operator

$$a_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\omega) \ e^{i\omega t} d\omega \qquad (2.67)$$

is stable.

In summary, then we see that the stability condition is that the difference equation (2.28) formed by the operator coefficients a_s (s = 0,1,...m) is to have a solution which describes a damped oscillation. This stability condition is also that the transfer function $A(\lambda)$ of the linear operator is to have no singularities or zeros below the axis of real frequency ω, and that its logarithm in that half plane is to be as small as possible at infinity. Briefly, a stable difference equation yields a minimum phase network, and conversely.

2.7 The Inverse Linear Operator

Let us consider the stable finite prediction operator

$$\xi_t = \sum_{s=o}^{m} a_s x_{t-s}, \quad a_o = 1 \tag{2.28}$$

so that its transfer function in the z-plane

$$A(z) = \sum_{s=o}^{m} a_s z^s \tag{2.652}$$

has zeros, all of which are exterior to the periphery $|z| = 1$ of the unit circle. In other words, the function $A(z)$ is analytic for $|z| \leq 1$, and

$$A(z) \neq 0 \text{ for } |z| \leq 1, \quad \sum_{s=o}^{m} a_s^2 < \infty. \tag{2.71}$$

Consequently the function

$$A^{-1}(z) = \frac{1}{A(z)} = \frac{1}{\sum_{s=o}^{m} a_s z^s} = B(z) \tag{2.711}$$

is a function which is analytic for $|z| \leq 1$, and has no zeros for $|z| \leq 1$. As a result we may expand this function in the power series

$$\frac{1}{A(z)} = \frac{1}{\sum_{s=o}^{m} a_s z^s} = \sum_{t=o}^{\infty} b_t z^t = B(z) \tag{2.712}$$

which converges for $|z| \leq 1$, and has no zeros for $|z| \leq 1$. The values of b_t for $t = 0, 1, 2, \ldots$ may be found by direct division of the polynomial $a_o + a_1 z + a_2 z^2 + \ldots + a_m z^m$ into unity. We shall define b_t to be those values given by equation (2.712) and define

$$b_t = 0, \text{ for } t < 0. \tag{2.713}$$

From the equation (2.712) we have

$$A(z)\ B(z) = 1 = \sum_{s=o}^{\infty} a_s z^s \sum_{t=o}^{\infty} b_t z^t \tag{2.72}$$

$$= \sum_{s=o}^{m} a_s \sum_{t=o}^{\infty} b_t z^{s+t} = \sum_{s=o}^{m} a_s \sum_{n=s}^{\infty} b_{n-s} z^n$$

where we have let $n = s+t$. Recalling that b_t is equal to zero for $t < 0$ we have

$$A(z)B(z) = 1 = \sum_{s=o}^{m} a_s \sum_{n=o}^{\infty} b_{n-s}\, \zeta^n = \sum_{n=o}^{\infty} \left(\sum_{s=o}^{m} a_s b_{n-s} \right) \zeta^n. \tag{2.721}$$

In order for this equation to hold, we see that

$$a_o b_o = 1 \tag{2.722}$$

$$\sum_{s=o}^{m} a_s b_{n-s} = 0, \text{ for } n = 1,2,3,\ldots \tag{2.723}$$

Since, we let $a_o = 1$, we have $b_o = 1$. Thus, given the operator coefficients a_o, $a_1, \ldots a_m$, the $b_1, b_2, b_3 \ldots$ may be uniquely determined by recursive deductions from the difference equation

$$\sum_{s=o}^{m} a_s b_{t-s} = 0, \quad t = 1,2,3,... \tag{2.723}$$

subject to the initial values $b_t = 0$ for $t < 0$ and $b_o = (a_o)^{-1} = 1$.

Since we have assumed that the linear operator with the coefficients a_o, a_1, $a_2 \ldots a_m$ is stable, the difference equation (2.723) with the constant coefficients a_o, a_1, ... a_m has the characteristic equation

$$P(\zeta) = \sum_{s=o}^{m} a_s \zeta^{m-s} \tag{2.73}$$

all the roots ζ_k of which lie within the periphery of the unit circle, that is $|\zeta_k| < 1$ for $k = 1,2,...m$. As a result, in virtue of equation (2.631) for the general solution of this difference equation, the solution b_t $(t = 0,1,2,...)$ describes a damped oscillation. In particular, the series

$$\sum_{t=o}^{\infty} |b_t| \text{ and } \sum_{t=o}^{\infty} b_t^2 \tag{2.731}$$

converge.

Let us now examine the linear operator, with coefficients $b_o, b_1, b_2, \ldots$ given by

$$\sum_{r=o}^{\infty} b_r \zeta_{t-r} \cdot \tag{2.74}$$

Since $b_t = 0$ for $t < 0$, this operator is of the prediction type, and since the operator has an infinite number of discrete coefficients it may be called an infinite discrete linear operator. Strictly speaking infinite linear operations may not be computed because an infinite number of multiplications would be required, and hence in this strict sense they are not computationally realizable. Nevertheless, because of equation (2.731), the operator equation (2.74) may be approximated by the partial sum

$$\sum_{r=o}^{M} b_r \zeta_{t-r} \tag{2.741}$$

to any degree of computational accuracy by choosing M sufficiently large, and in this sense the operator equation (2.74) is computationally realizable.

Since B(z), given by equation (2.711), has no singularities or zeros for $|z| \leq 1$, the transfer function of the b_t operator

$$B(\omega) = \sum_{r=o}^{\infty} b_r \, e^{-i\omega r} \tag{2.742}$$

has no singularities or zeros in the lower half λ plane where $\lambda = \omega + i\,\sigma$, (see Figure 3). Thus the linear operator b_t has minimum phase shift characteristic, and is stable.

In summary, then, the infinite linear operator b_t has the following properties:

$$b_t = 0 \text{ for } t < 0$$

$$b_o > 0 \tag{2.75}$$

$$\sum_{t=o}^{\infty} b_t^2 < \infty$$

and the

$$B(z) = \sum_{t=o}^{\infty} b_t z^t \neq 0 \text{ for } |z| \leq 1, \tag{2.751}$$

and we shall call such infinite prediction operators computationally realizable and stable.

Let us now examine the significance of the linear operator

$$\sum_{r=o}^{\infty} b_r \xi_{t-r} \tag{2.74}$$

Inserting into this equation the prediction errors ξ_t, ξ_{t-1}, ξ_{t-2},... given by the prediction operator (2.28), we have

$$\sum_{r=o}^{\infty} b_r \xi_{t-r} = \sum_{r=o}^{\infty} b_r \sum_{s=o}^{m} a_s x_{t-r-s}$$

$$= \sum_{s=o}^{m} a_s \sum_{r=o}^{\infty} b_r x_{t-r-s} \cdot \qquad (2.76)$$

Letting $n = r+s$, we have

$$\sum_{r=o}^{\infty} b_r \xi_{t-r} = \sum_{s=o}^{m} a_s \sum_{n=s}^{\infty} b_{n-s}\, x_{t-n} \qquad (2.761)$$

and recalling $b_r = 0$ for $r < 0$, we have

$$\sum_{r=o}^{\infty} b_r \xi_{t-r} = \sum_{s=o}^{m} a_s \sum_{n=o}^{\infty} b_{n-s}\, x_{t-n} = \sum_{n=o}^{\infty} \left(\sum_{s=o}^{m} a_s b_{n-s} \right) x_{t-n} \cdot \qquad (2.762)$$

Therefore, because of equations (2.722) and (2.723) we have

$$\sum_{r=o}^{\infty} b_r\, \xi_{t-r} = x_t \cdot \qquad (2.77)$$

That is, the linear operator (2.77) with coefficients $b_o, b_1, b_2 \ldots$ operates on the prediction errors $\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots$ at time t and prior to time t in order to yield the value of the time series x_t. Thus we see that the operator (2.28) with

coefficients $a_m, a_{m-1}, \ldots a_2, a_1, a_o$ operates on the time series $\ldots, x_{t-2}, x_{t-1}, x_t$ to yield the prediction errors ξ_{t-2}, ξ_{t-1}, ξ_t, whereas the operator (2.77) with coefficients $\ldots, b_2, b_1, b_o$, performs the inverse operation. Therefore the a_s operator (2.28) is called the inverse to the b_s operator (2.77), and conversely.

More generally, the infinite linear operators

$$\xi_t = \sum_{s=o}^{\infty} a_s x_{t-s} \tag{2.78}$$

$$x_t = \sum_{s=o}^{\infty} b_s \xi_{t-s} \tag{2.781}$$

are realizable, stable, and inverse to each other if the coefficients a_t satisfy

$$a_t = 0 \text{ for } t < 0$$

$$a_o > 0$$

$$\sum_{t=o}^{\infty} a_t^2 < \infty \tag{2.782}$$

$$A(z) = \sum_{t=o}^{\infty} a_t z^t \neq 0 \text{ for } |z| \leq 1,$$

if the coefficients b_t satisfy

$$b_t = 0 \text{ for } t < 0$$

$$b_o > 0$$

$$\sum_{t=o}^{\infty} b_t^2 < \infty \tag{2.783}$$

$$B(z) = \sum_{t=o}^{\infty} b_t z^t \neq 0 \text{ for } |z| \leq 1$$

and if

$$A(z)\, B(z) = 1. \tag{2.784}$$

Hence the a_t and b_t are related by

$$a_o b_o = 1$$

$$\sum_{s=o}^{t} a_s b_{t-s} = 0, \text{ for } t = 1,2,3,\ldots \tag{2.785}$$

Thus, given the set a_s, the set b_s may be uniquely determined, and vice versa. For example equation (2.785) for $t = 1,2,3,$ yields

$$a_1 = - b_1$$
$$a_2 = -b_2 + b_1^2 \qquad (2.786)$$
$$a_3 = -b_3 + 2b_1b_2 - b_1^3$$

on the one hand, and

$$b_1 = - a_1$$
$$b_2 = -a_2 + a_1^2 \qquad (2.787)$$
$$b_3 = -a_3 + 2a_1a_2 - a_1^3$$

on the other hand. Both the a_t series and the b_t series ($t = 0,1,2,\ldots$) form damped oscillations.

The transfer functions

$$A(\omega) = \sum_{s=o}^{\infty} a_s e^{-\omega s} \qquad (2.791)$$

and

$$B(\omega) = \sum_{s=o}^{\infty} b_s e^{-i\omega s} \qquad (2.792)$$

are free from singularities and zeros in the lower half λ-plane, $\lambda = \omega + i\,\sigma$, and have minimum phase shift charac-

teristic. They are related by

$$A(\omega)\ B(\omega) = 1, \tag{2.793}$$

$$\frac{1}{A(\omega)} = A^{-1}(\omega) = B(\omega), \tag{2.794}$$

$$\frac{1}{B(\omega)} = B^{-1}(\omega) = A(\omega), \tag{2.795}$$

$$A^{-1}(\omega)\ A(\omega) = 1, \tag{2.796}$$

and

$$B^{-1}(\omega)\ B(\omega) = 1. \tag{2.797}$$

2.8 The Power Transfer Function and its Stable Prediction Operator

The power transfer function $\mathbb{P}(\omega)$ is defined to be the square of the absolute value $|A(\omega)|$ of the transfer function $A(\omega)$; that is, the power transfer function is given by

$$\begin{aligned} \mathbb{P}(\omega) &= |A(\omega)|^2 = A(\omega)\ \overline{A(\omega)} \\ &= (\mathrm{Re}[A(\omega)])^2 + (\mathrm{Im}[A(\omega)])^2 \geq 0 \end{aligned} \tag{2.81}$$

where the bar indicates the complex conjugate. Here $|A(\omega)|$ is called the gain of the linear operator, and it is given by the square root of the power transfer function; that is

$$|A(\omega)| = \sqrt{\mathbb{T}(\omega)} \tag{2.811}$$

so that knowledge of the gain of the linear operator and knowledge of the power transfer function are equivalent.

The power transfer function $\mathbb{T}(\omega)$ of a finite linear operator $a_0, a_1, \ldots a_m$ may be expressed by the finite trigonometric series

$$\mathbb{T}(\omega) = \sum_{s=0}^{m} a_s e^{-i\omega s} \sum_{t=0}^{m} a_t e^{i\omega t} = |A(\omega)|^2 \geq 0 \tag{2.812}$$

which is non-negative for $-\pi < \omega \leq \pi$. We may rewrite this expression in the following way

$$\begin{aligned} \mathbb{T}(\omega) &= \sum_{s=0}^{m} \sum_{t=0}^{m} a_s a_t e^{-i\omega(s-t)} \\ &= \sum_{\tau=-m}^{m} e^{-i\omega\tau} \sum_{t=0}^{m} a_t a_{t+\tau} \end{aligned} \tag{2.813}$$

where $\tau = s-t$. Let us define r_τ to be

$$r_\tau = \sum_{t=o}^{m} a_t a_{t+\tau}; \; r_\tau = r_{-\tau} \tag{2.814}$$

that is,

$$\begin{aligned}
r_o &= a_o^2 + a_1^2 + a_2^2 + \cdots + a_{m-2}^2 + a_{m-1}^2 + a_m^2 \\
r_1 &= r_{-1} = a_o a_1 + a_1 a_2 + a_2 a_3 + \cdots + a_{m-2} a_{m-1} + a_{m-1} a_m \\
r_2 &= r_{-2} = a_o a_2 + a_1 a_3 + a_2 a_4 + \cdots + a_{m-2} a_m \\
&\cdots \\
r_{m-1} &= r_{1-m} = a_o a_{m-1} + a_1 a_m \\
r_m &= r_{-m} = a_o a_m
\end{aligned} \tag{2.814}$$

Therefore if we are given the linear operator with coefficients $a_o, a_1, \ldots, a_m$, we may find the r_τ by means of equation (2.814) and thereby determine the power transfer function

$$\begin{aligned}
\Psi(\omega) &= \left| \sum_{s=o}^{m} a_s e^{-i\omega s} \right|^2 = \sum_{\tau=-m}^{m} r_\tau e^{-i\omega\tau} \\
&= r_o + 2 \sum_{\tau=o}^{m} r_\tau \cos \omega\tau \geq 0.
\end{aligned} \tag{2.815}$$

In this section we wish to consider the inverse problem; that is, given the power transfer function $\Psi(\omega)$, find the coefficients $a_0, a_1, \ldots a_m$ of the linear operator which yield this power transfer function. This inverse problem, as it stands, is not unique in that several different linear operators may yield the same power transfer function $\Psi(\omega)$. As we shall see, however, all of these linear operators are unstable, except one. In other words, given the power transfer function $\Psi(\omega)$ we wish to find the one, and only one, realizable, stable prediction operator which yields the power transfer function $\Psi(\omega)$. This stable prediction operator has the transfer function with gain equal to $\sqrt{\Psi(\omega)}$ and minimum phase characteristic. The import of this section resides in the fact that if one wishes to design a stable linear operator, he needs only to have information about the desired absolute gain characteristics $|A(\omega)| = \sqrt{\Psi(\omega)}$, and needs no information concerning the phase characteristics. In this section, we give direct procedure for the determination of such stable finite linear operators. This procedure may be readily programmed for automatic computation on a digital computer.

Since we wish to consider finite linear operators with the coefficients $a_0, a_1, \ldots a_m$, it is necessary to express the

power transfer function in terms of a finite trigonometric series

$$\Psi(\omega) = \sum_{\tau=-m}^{m} r_\tau e^{-i\omega\tau} = r_o + 2\sum_{\tau=1}^{m} r_\tau \cos \omega\tau, \; r_o > 0, \; r_\tau = r_{-\tau} \tag{2.82}$$

which is non-negative for $-\pi < \omega \leq \pi$, and the r_τ are real.

If the power transfer function is given by the infinite Fourier series

$$\Psi(\omega) = \rho_o + 2\sum_{\tau=1}^{\infty} \rho_\tau \cos \omega\tau \geq 0, \; -\pi < \omega \leq \pi, \tag{2.821}$$

which is non-negative for $-\pi < \omega \leq \pi$, then the Cesaro partial sum

$$\rho_o + 2\sum_{\tau=1}^{N} \left(1 - \frac{\tau}{N}\right) \rho_\tau \cos \omega\tau, \tag{2.822}$$

which is also non-negative for $\pi < \omega \leq \pi$, may be used as the finite series approximation to $\Psi(\omega)$. We then have

$$\Psi(\omega) \approx r_o + 2\sum_{\tau=1}^{m} r_\tau \cos \omega\tau, \; r_\tau = \left(1 - \frac{\tau}{N}\right) \rho_\tau \tag{2.823}$$

which may be used for equation (2.82).

Let Q(u) be the polynomial of order m obtained from

$$r_o + 2 \sum_{\tau=1}^{m} r_\tau (z^{-\tau} + z^{\tau}) \qquad (2.824)$$

by the substitution $z^{-1} + z = u$. Wold (1938, Theorem 12), in connection with his study of the process of moving averages, shows that a necessary and sufficient condition that the finite series

$$r_o + 2 \sum_{\tau=1}^{m} r_\tau (e^{i\omega\tau} + e^{-i\omega\tau}) \qquad (2.825)$$

be non-negative for $-\pi \leq \omega \leq \pi$ is that the equation $Q(u) = 0$ should have no real root of odd multiplicity in the interval $-2 < u < 2$.

The method which we give in this section was used by Wold (1938) in order to factor the power spectrum of the process of finite moving averages, and the power spectrum of the autoregressive process.

As we shall see in Section 5.5-A, the power spectrum of a process of finite moving averages is given by

$$\phi(0) + 2 \sum_{\tau=1}^{M} \phi(\tau) \cos \omega\tau \geq 0 \qquad (5.523)$$

which has the same form as the power transfer function $\overline{\Psi}(\omega)$ given by equation (2.82). Thus by the method to be presented in this section, we may factor the power spectrum into the product

$$\phi(0) + 2 \sum_{\tau=1}^{M} \phi(\tau) \cos \omega\tau = B(\omega)\, \overline{B(\omega)} \qquad (5.523)$$

where

$$B(\omega) = \sum_{s=0}^{M} b_s e^{-i\omega s} \qquad (5.524)$$

is free from singularities and zeros in the lower half λ-plane.

As we shall see in Section 5.5-B, the reciprocal of the power spectrum of an autoregressive process is given by

$$r_o + 2 \sum_{\tau=1}^{m} r_\tau \cos \omega\tau \qquad (5.542)$$

which has the same form as the power transfer function given by equation (2.82). Thus, by the method to be presented in this section, the reciprocal of the power spectrum may be factored into

$$r_o + 2 \sum_{\tau=1}^{m} r_\tau \cos \omega\tau = A(\omega)\, \overline{A(\omega)} \qquad (5.543)$$

where

$$\frac{1}{A(\omega)} = \frac{1}{\sum_{s=o}^{m} a_s e^{-i\omega s}} \tag{5.547}$$

is free of singularities and zeros in the lower half λ-plane.

Returning now to the power transfer function let us suppose that the $r_\tau = r_{-\tau}$ are such that $\Psi(\omega)$ given by equation (2.82) is non-negative for $-\pi < \omega \leq \pi$. Under the transformation $z = e^{-i\lambda}$, where $\lambda = \omega + i\,\sigma$, (see Figure 3), the power transfer fucntion $\Psi(\omega)$ becomes

$$\Psi(z) = \sum_{\tau=-m}^{m} r_\tau\, z^\tau \tag{2.83}$$

where $\Psi(z)$ is a rational function in z.

We see that

$$z^m\, \Psi(z) = \sum_{\tau=-m}^{m} r_\tau\, z^{\tau+m} \tag{2.831}$$

is a polynomial of order 2m. Expressing this polynomial in terms of its roots z_k (k=1,2,...2m) we have

$$z^m\, \Psi(z) = r_m(z-z_1)(z-z_2)\ldots(z-z_{2m}) \tag{2.832}$$

where we assume $r_m \neq 0$ so that $|z_k| \neq 0$ for k = 1,2,.., 2m. Hence the rational function $\Psi(z)$ may be expressed as

$$\Psi(z) = z^{-m} r_m (z-z_1)(z-z_2)\ldots(z-z_{2m}) \qquad (2.833)$$

and the power transfer function is therefore

$$\Psi(\omega) = e^{i\omega m} r_m(e^{-i\omega} - z_1)(e^{-i\omega} - z_2)\ldots(e^{-i\omega} - z_{2m}) \quad (2.834)$$

Since the power transfer function $\Psi(\omega)$ is a real function of ω, we have

$$\Psi(\omega) = \overline{\Psi(\omega)} = r_m e^{-i\omega m} (e^{i\omega} - \bar{z}_1)(e^{i\omega} - \bar{z}_2) \ldots (e^{i\omega} - \bar{z}_{2m})$$

$$= r_m e^{i\omega m} (1 - \bar{z}_1 e^{-i\omega})(1 - \bar{z}_2 e^{-i\omega}) \ldots (1 - \bar{z}_{2m} e^{-i\omega})$$

$$= \bar{z}_1 \bar{z}_2 \ldots \bar{z}_{2m} r_m e^{i\omega m} \left(\frac{1}{z_1} - e^{-i\omega}\right)\left(\frac{1}{z_2} - e^{-i\omega}\right) \ldots \left(\frac{1}{z_{2m}} - e^{-i\omega}\right)$$

(2.835)

Letting $z = e^{-i\lambda}$ where $\lambda = \omega + i\,\sigma$ we have

$$\overline{\Psi}(z) = \overline{z}_1\overline{z}_2\ldots\overline{z}_{2m}r_m z^{-m}(\overline{z}_1^{-1} - z)(\overline{z}_2^{-1} - z)\ldots(\overline{z}_{2m}^{-1} - z) \quad (2.836)$$

Comparing equations (2.824) and (2.831) we see that if z_k is a root of the polynomial $z^m\,\overline{\Psi}(z)$, then $\overline{z}_k^{-1}$ is also a root.

Moreover since the power transfer function $\overline{\Psi}(\omega)$ is a real even function of ω, we have

$$\overline{\Psi}(-\omega) = \overline{\overline{\Psi}(\omega)} = \overline{\Psi}(\omega) \quad (2.84)$$

which is

$$\overline{\Psi}(-\omega) = r_m e^{i\omega m}(e^{-i\omega} - \overline{z}_1)(e^{-i\omega} - \overline{z}_2)\ \ldots\ (e^{-i\omega} - \overline{z}_{2m})$$

$$= \overline{\overline{\Psi}(\omega)} = r_m e^{i\omega m}\,(e^{-i\omega} - z_1)(e^{-i\omega} - z_2)\ldots(e^{-i\omega} - z_{2m}) = \overline{\Psi}(\omega).$$

Letting $z = e^{-i\lambda}$, where $\lambda = \omega + i\,\sigma$, we have

$$\overline{\Psi}(z) = r_m z^{-m}(z-\overline{z}_1)(z-\overline{z}_2)\ldots(z-\overline{z}_{2m}) = r_m z^{-m}(z-z_1)(z-z_2)\ldots(z-z_{2m}) \quad (2.842)$$

so that if z_k is a root of the polynomial $z^m\ \overline{\Psi}(z)$ then $\overline{z}_k$ is also a root.

In summary, then, if z_k is a root of $z^m \Psi(z)$ then $\bar{z}_k$, z_k^{-1}, and $\bar{z}_k^{-1}$ are also roots. Thus if α_k is a complex root of $z^m \Psi(z)$ with modulus $|\alpha_k| \neq 1$, then α_k, α_k^{-1}, $\bar{\alpha}_k^{-1}$ are distinct from each other, and are all roots of $z^m \Psi(z)$. If β_k is a real root of $z^m \Psi(z)$ with modulus $|\beta_k| \neq 1$, then β_k and β_k^{-1} are distinct from each other, and are both roots of $z^m \Psi(z)$. If γ_k is a complex root of $z^m \Psi(z)$ with modulus $|\gamma_k| = 1$, then γ_k and $\bar{\gamma}_k$ are distinct from each other, and are both roots of $z^m \Psi(z)$. Let ρ_k represent the real roots of $z^m \Psi(z)$ with modulus $|\rho_k| = 1$.

Accordingly, the polynomial $z^m \Psi(z)$ may be expressed as

$$z^m \Psi(z) = r_m \prod_{k=1}^{h} (z-\alpha_k)(z-\bar{\alpha}_k)(z-\alpha_k^{-1})(z-\bar{\alpha}_k^{-1}) \prod_{k=1}^{j} (z-\beta_k)(z-\beta_k^{-1})$$

$$\prod_{k=1}^{\ell} (z-\gamma_k)(a-\bar{\gamma}_k) \prod_{k=1}^{n} (z-\rho_k) \qquad (2.85)$$

where any root of order p is repeated p times.

Let us now turn our attention to equation (2.812),

$$\Psi(\omega) = \sum_{s=o}^{m} a_s e^{-i\omega s} \sum_{t=o}^{m} a_t e^{i\omega t} \qquad (2.812)$$

which expresses the relationship of the power transfer function $\Psi(\omega)$ with the coefficients $a_o, a_1, \ldots a_m$ of the prediction operator which yields $\Psi(\omega)$. Under the transformation $z = e^{-i\lambda}$, $\lambda = \omega + i\,\sigma$, (see Figure 3), we have

$$\Psi(z) = \sum_{s=o}^{m} a_s z^s \sum_{t=o}^{m} a_t z^{-t}. \qquad (2.86)$$

We thus have

$$z^m \Psi(z) = \sum_{\tau=-m}^{m} r_\tau z^{\tau+m} = \sum_{s=o}^{m} a_s z^s \sum_{t=o}^{m} a_t z^{m-t} \qquad (2.861)$$

In section 2.6 we defined A(z), called the transfer function in the z plane, to be

$$A(z) = \sum_{s=o}^{m} a_s z^s \qquad (2.652)$$

so equation (2.86) becomes

$$\Psi(z) = A(z)\, A(z^{-1}) \qquad (2.862)$$

and equation (2.861) becomes

$$z^m \, \Psi(z) = [A(z)] \, [z^m A(z^{-1})] \qquad (2.863)$$

which is

$$r_m z^{2m} + r_{m-1} z^{2m-1} + \ldots + r_1 z^{m+1} + r_o z^m + r_1 z^{m-1} + \ldots + r_{m-1} z + r_m$$

$$= (a_o + a_1 z + \ldots + a_{m-1} z^{m-1} + a_m z^m)(a_o z^m + a_1 z^{m-1} + \ldots + a_{m-1} z + a_m).$$

(2.863)

In order to factor $z^m \, \Psi(z)$ into the two real polynomials $A(z)$ and $z^m A(z^{-1})$, we see that one of the real polynomials $(z-\alpha_k)(z-\bar{\alpha}_k)$ or $(z-\alpha_k^{-1})(z-\bar{\alpha}_k^{-1})$ must be a factor in the polynomial $A(z)$. Since β_k is real, then either $(z-\beta_k)$ or $(z-\beta_k^{-1})$ is a factor in $A(z)$. On the other hand, since the factors $(z-\gamma_k)$ and $(z-\bar{\gamma}_k)$ are complex, both of them must be contained in $A(z)$. Likewise $(z-\rho_k)$ must appear in $A(z)$. Thus it is necessary that the roots γ_k and ρ_k, which have modulus 1, appear an even number of times, that is $\ell = 2\ell'$ and $n = 2n'$. This condition that roots of modulus one appear an even number of times is satisfied since $\Psi(\omega) \geq 0$. Thus we have

$$A(z) = a_m \prod_{k=1}^{h} (z-\alpha_k)(z-\bar{\alpha}_k) \prod_{k=1}^{j} (z-\beta_k) \prod_{k=1}^{\ell'} (z-\gamma_k)(z-\bar{\gamma}_k) \prod_{k=1}^{n'} (z-\rho_k) \tag{2.87}$$

and

$$z^m A(z^{-1}) = \prod_{k=1}^{h} (z-\alpha_k^{-1}) (z-\bar{\alpha}_k^{-1}) \prod_{k=1}^{j} (z-\beta_k^{-1}) \tag{2.871}$$

$$\prod_{k=1}^{\ell'} (z-\gamma_k^{-1})(z-\bar{\gamma}_k^{-1}) \prod_{k=1}^{n'} (z-\rho_k^{-1})$$

since $\bar{\gamma}_k = \gamma_k^{-1}$, $\gamma_k = \bar{\gamma}_k^{-1}$, and $\rho_k = \rho_k^{-1}$.

Thus if the zeros of $A(z)$ are z_k, then the zeros of $z^m A(z^{-1})$ are z_k^{-1}. In order for the linear operator with coefficients $a_o, a_1, \ldots a_m$ to be strictly stable, then all the roots of

$$A(z) = \sum_{s=o}^{m} a_s z^s \tag{2.652}$$

must have modulus $|z_k| > 1$. Thus if the transfer function $\mathbb{T}(\omega)$ yields roots γ_k and ρ_k which do have modulus $|\gamma_k| = |\rho_k|$ equal to one, then there is no strictly stable linear operator which yields this transfer function , although

there is a linear operator on the borderline of stability, such as the cosine operator (2.67).

Let us suppose that $\Psi(\omega)$ yields no roots γ_k and ρ_k with modulus equal to one. Then we have

$$A(z) = a_m \prod_{k=1}^{h} (z-\alpha_k)(z-\bar{\alpha}_k) \prod_{k=1}^{j} (z-\beta_k) = \sum_{s=0}^{m} a_s z^s \qquad (2.88)$$

and

$$z^m A(z^{-1}) = \prod_{k=1}^{h} (z-\alpha_k^{-1})(z-\bar{\alpha}_k^{-1}) \prod_{k=1}^{j} (z-\beta_k^{-1}) = \sum_{s=0}^{m} a_s z^{m-s} \qquad (2.881)$$

where $|\alpha_k| \neq 1$ for $k = 1,2,\ldots h$ and $|\beta_k| \neq 1$ for $k = 1,2,\ldots j$. If the zeros of $A(z)$ are z_k for $k = 1,2,\ldots m$, then the zeros of $z^m A(z^{-1})$ are z_k^{-1} for $k = 1,2,\ldots m$, and since $|z_k| \neq 1$, it follows that half of the 2m roots of

$$z^m \Psi(z) = A(z) z^m A(z^{-1}) \qquad (2.863)$$

have modulus greater than one, and the other half of the 2m roots has modulus less than one.

Thus if we choose those m roots of $z^m \Psi(z)$ which have modulus greater than one, and call these roots the α_k

$(k = 1,2,\ldots h)$, the $\bar{\alpha}_k$ $(k = 1,2,\ldots h)$, and the β_k $(k = 1,2,\ldots h)$ which appear in equation (2.872), then A(z) will represent the transfer function in the z plane, of a stable linear operator. That is, since A(z) has roots, all of which have modulus greater than one, the transfer function

$$A(\omega) = \sum_{s=o}^{m} a_s e^{-i\omega s} \qquad (2.421)$$

will be free from zeros in the lower half plane of $\lambda = \omega + i\sigma$, and be of the minimum phase shift type. The prediction type operator, with coefficients

$$a_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\omega) e^{i\omega t}\, dt \qquad (2.864)$$

is then stable.

On the other hand, if we did not choose the roots in the above fashion, there being at most 2^m different ways of choosing the roots, then A(z) would have roots, some of which have modulus greater than one, and some of which have modulus less than one. Consequently the transfer function $A(\omega)$ would not be of the minimum phase shift type and its linear operator would not be stable.

Let us summarize the computational procedure required to determine the stable operator coefficients $a_o, a_1, \ldots a_m$ from the power transfer function

$$\Psi(\omega) = \sum_{\tau=-m}^{m} r_\tau e^{-i\omega\tau} \geq 0, \; r_o > 0. \tag{2.82}$$

Form the polynomial

$$z^m \Psi(z) = \sum_{\tau=-m}^{m} r_\tau z^{\tau+m} \tag{2.831}$$

and solve for its roots $z_1, z_2, \ldots z_{2m}$. Let $z_1', z_2', \ldots z_m'$ be those $z_k (k = 1,2,\ldots 2m)$ of modulus greater than one and also those z_k of modulus one counted half as many times. (In order for there to be a strictly stable operator, there may be no z_k of modulus one). Then we form the polynomial

$$A(z) = (z-z_1')(z-z_2')\ldots(z-z_m') = \sum_{s=o}^{m} a_s z^s , \tag{2.89}$$

and the operator coefficients are given by the a_s. They represent a stable linear operator, the transfer function of which has minimum phase characteristic. Thus we have shown that the power transfer function $\Psi(\omega)$, equation (2.82), may be factored into

$$\mathbb{\Psi}(\omega) = A(\omega)\,\overline{A(\omega)} \tag{2.891}$$

where the transfer function

$$A(\omega) = \sum_{s=o}^{m} a_s e^{-i\omega s} \tag{2.892}$$

will be free from singularities and zeros in the lower half λ-plane.

CHAPTER III

THE NON-STATISTICAL ANALYSIS OF TIME SERIES

3.1 The Functional Approach

In the last chapter those properties of the finite discrete linear operator were developed which are independent of the properties of the time series x_t under consideration. In this chapter we wish to consider the methodology of the non-statistical or deterministic approach to the study of discrete time series x_t. Since this approach leads to perfect functional representations of the empirical data, the various schemes of analysis are called functional schemes. As we shall see, the harmonic analysis of the time series, that is the analysis in terms of $e^{i\omega t}$, plays an important role in these functional schemes.

As an introduction to the concept of the functional approach, let us for the moment consider the continuous time series $x(t)$. If the integral which represents the total energy of $x(t)$, given by

$$\int_{-\infty}^{\infty} x^2(t)\, dt \qquad (3.11)$$

is finite, then the Fourier integral representation of $x(t)$,

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)\, e^{i\omega t}\, d\omega, \qquad (3.12)$$

is a perfect functional representation of $x(t)$. Here the function $X(\omega)$, given by the inverse Fourier transform

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-i\omega t}\, dt, \qquad (3.13)$$

is a function which contains the same information as $x(t)$, but is in the frequency domain rather than in the time domain as $x(t)$.

In this chapter in which we consider only discrete time series, we first indicate the methodogy of the periodic functional scheme. We then consider the aperiodic functional scheme for transient time series, especially in connection with the so-called linear filtering problem, which is the transformation of one transient time series into another transient by linear operations. Finally we see that the functional approach in certain cases requires methods of averaging in order to approximate certain functions for computational purposes. Since in many cases methods of averaging become more physically meaningful from the standpoint of probability theory, in the

last three chapters we shall consider the analysis of time series from the statistical point of view.

In particular, in Chapter V, we shall see that a stationary time series (with an absolutely continuous spectral distribution) may be considered to be additively composed of wavelets, all with the same shape, and with random strengths and arrival times. As we shall see, methods of averaging play an important role in the determination of the wavelet shape. Since the shape of such a wavelet represents a deterministic transient time function, it may be treated by the methodology of this chapter.

3.2 The Periodic Functional Scheme

The various periodic functional schemes assume that the time series under consideration are composed of strictly periodic components. Since such schemes are treated in detail in the literature and since they have limited application in seismology, we shall only briefly indicate the methodology used. For more detailed discussions, the reader is referred to Schuster (1898, 1900) and Whittaker and Robinson (1926).

Let an observational time series be given by

$$x_1, x_2, \ldots x_T \tag{3.21}$$

for the time range $1 \leq t \leq T$, where the mean value of x_t is zero. We shall consider the case in which it is assumed that the observational data are strictly periodic, say with period T. Then the infinite time series is given by the sequence

$$\ldots x_T, x_1, x_2, \ldots x_T, x_1, x_2, \ldots x_T, x_1, x_2, \ldots x_T, x_1, \ldots \tag{3.22}$$

Representing this time series by $x_t (-\infty < t < \infty)$, the difference equation

$$x_t - x_{t-T} = 0 \tag{3.23}$$

holds for any t. The solution of this difference equation (where for simplicity we let T be even) is

$$\begin{aligned} x_t &= \sum_{n=1}^{T/2} \left(A_n \cos \frac{2\pi}{T} nt - B_n \sin \frac{2\pi}{T} nt\right) \\ &= \sum_{n=1}^{T/2} C_n \cos \left(\frac{2\pi}{T} nt + \theta_n\right) \end{aligned} \tag{3.24}$$

where

$$C_n^2 = A_n^2 + B_n^2 , \quad \tan \theta_n = \frac{B_n}{A_n} . \tag{3.25}$$

The Fourier analysis of the time series x_t will yield the A_n and B_n, so that equation (3.24) will be a perfect functional representation of the original data.

3.3 The Aperiodic Functional Scheme

The aperiodic functional scheme deals with that class of discrete time series x_t whose total energy

$$\sum_{t=-\infty}^{\infty} x_t^2 \tag{3.31}$$

is finite. Examples of such time series are transient time functions, such as the seismic wavelets of Ricker (1940, 1941, 1943, 1944, 1945, 1949, 1953a, 1953b).

We see that all observational time series of finite time duration , say from t = 0 to t = T, fall into this class, if we define the time series x_t to be zero outside the basic time interval, that is,

$$x_t = 0 \text{ for } t < 0 \text{ and } t > T. \tag{3.311}$$

Hence the time series x_t defined for all time has finite total energy given by

$$\sum_{t=-\infty}^{\infty} x_t^2 = \sum_{t=o}^{T} x_t^2 \tag{3.312}$$

For the remaining parts of this Chapter we shall deal with finite time series which are defined to be zero outside of their basic time interval. Although the non-statistical methods of this chapter can be applied to any time series of finite duration, we do not wish to imply that they should be applied to the analyses of all time series of finite time duration. Instead the methodology to be used, statistical or non-statistical, should depend upon the type of problem to be solved, and should be chosen with consideration to all prior knowledge and experience about the problem.

In particular, the methodology of this chapter is applicable to wavelets which damp toward zero sufficiently rapidly so that they may be approximated by zero outside of a finite time interval, and especially applicable to wavelets which are stable in the sense of Section 2.6, where we let the wavelet x_t be the linear operator a_t.

As we shall see in the following chapters, a stationary time series (with an absolutely continuous spectral distribution function) may be considered to be additively composed of wavelets, all of the same shape, but with random strengths and arrival times. Let us outline heuristically how one, in effect, determines a certain linear operation for such a time series. First the random elements of the time series are

destroyed by averaging, (i.e. the computation of the auto-correlation function), and a unique stable wavelet shape (i.e. the $\Psi_1(t)$ of Levinson (1947) which is our b_t in the following chapters) is preserved. Then the particular linear operation may be determined by non-statistical methods on this deterministic wavelet shape. Since the time series is additively composed of such wavelets, this linear operation applies equally as well to the time series itself.

The aperiodic functional scheme is a functional scheme because it leads to a perfect functional representation of the observational or theoretical data. This functional representation is given by the Fourier integral representation of the function

$$x_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega)\, e^{i\omega t}\, d\omega \qquad (3.32)$$

where $X(\omega)$, called the (complex) phase and amplitude spectrum of x_t, is given by

$$X(\omega) = \sum_{t=-\infty}^{\infty} x_t e^{-i\omega t} = \sum_{t=o}^{T} x_t e^{-i\omega t} \qquad (3.321)$$

since $x_t = 0$ for $t < 0$ and $t > T$. Equation (3.32) gives a perfect functional representation of x_t, and we see that the

function $X(\omega)$ contains the same information as the function x_t. The energy spectrum $\overline{\Phi}_{xx}(\omega)$ of x_t is defined to be

$$\overline{\Phi}_{xx}(\omega) = \overline{X(\omega)}\, X(\omega) = |X(\omega)|^2 \geq 0, \quad -\pi < \omega \leq \pi, \tag{3.33}$$

which is equal to

$$\overline{\Phi}_{xx}(\omega) = \sum_{t=o}^{T} x_t e^{i\omega t} \sum_{s=o}^{T} x_s e^{-i\omega s}$$

$$= \sum_{t=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} x_t x_s e^{-i\omega(s-t)}$$

$$= \sum_{\tau=-\infty}^{\infty} e^{-i\omega\tau} \sum_{t=-\infty}^{\infty} x_t x_{t+\tau}$$

$$= \sum_{\tau=-T}^{T} e^{-i\omega\tau} \sum_{t=o}^{T-\tau} x_t x_{t+\tau} \tag{3.331}$$

where $\tau = s - t$. Let us define the autocorrelation function of x_t to be

$$\phi_{xx}(\tau) = \sum_{t=o}^{T-\tau} x_t x_{t+\tau}, \tag{3.332}$$

where we see

$$\phi_{xx}(-\tau) = \phi_{xx}(\tau)$$

$$\phi_{xx}(\tau) = 0 \text{ for } |\tau| > T. \qquad (3.333)$$

Then equation (3.331) becomes

$$\bar{\Phi}_{xx}(\omega) = \sum_{\tau=-T}^{T} \phi_{xx}(\tau)\, e^{-i\omega\tau} = \phi(0) + 2 \sum_{\tau=1}^{T} \phi(\tau) \cos \omega\tau \qquad (3.334)$$

which expresses the energy spectrum as the Fourier transform of the autocorrelation. On the other hand, the transform of the energy spectrum is the autocorrelation, as seen by

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \bar{\Phi}_{xx}(\omega)\, e^{i\omega t}\, d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{\tau=-T}^{T} \phi_{xx}(\tau) e^{-i\omega\tau}\, e^{i\omega t}\, d\omega = \phi_{xx}(t)$$

(3.34)

because

$$\int_{-\pi}^{\pi} e^{-i\omega\tau}\, e^{i\omega t}\, d\omega = \begin{cases} 0 & \tau \neq t \\ 2\pi & \tau = t \end{cases} \qquad (3.341)$$

for integer values of t and τ. We see that the total energy of x_t is given by

$$\phi_{xx}(0) = \sum_{t=0}^{T} x_t^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{xx}(\omega)\, d\omega \qquad (3.342)$$

Equation (3.33), which may be written

$$\Phi_{xx}(\lambda) = X(\lambda)\, \overline{X(\lambda)}, \quad \lambda = \omega + i\, \sigma \qquad (3.35)$$

generally does not represent a factorization of the spectrum which satisfies the Wold-Kolmogorov conditions. (See Section 5.1) In other words, the factor $X(\lambda)$ generally will not be free from singularities and zeros in the lower half λ-plane. This condition will hold only if the finite time series x_t satisfies the same stability conditions which we gave for a finite linear operator a_t in Section 2.6. Since $\Phi_{xx}(\omega)$ in equation (3.334) is expressed in a finite non-negative trigonometric series, the method of Section 2.8 may be used to factor $\Phi_{xx}(\omega)$ into the product

$$\Phi_{xx}(\omega) = B(\omega)\, \overline{B(\omega)}$$

where $B(\lambda)$ is free of singularities and zeros in the lower half λ-plane. The function $B(\omega)$ may then be used in preference to $X(\omega)$ in many applications, although we shall not

explicitly make use of $B(\omega)$ in the remainder of this chapter.

Suppose that we have another time series y_t of finite time duration, say from $t = 0$ to $t = T$. As in the case of x_t, we let

$$y_t = 0, \text{ for } t < 0, \; t > T. \tag{3.36}$$

Similar relations hold for y_t as for the function x_t. As a matter of notation, the complex phase amplitude spectrum is denoted by $Y(\omega)$, the energy spectrum by $\overline{\Phi}_{yy}(\omega)$, and the auto-correlation by $\phi_{yy}(\tau)$.

The cross energy spectrum of x_t and y_t is

$$\overline{\Phi}_{xy}(\omega) = \overline{X(\omega)}\, Y(\omega) \tag{3.37}$$

which is equal to

$$\begin{aligned} \overline{\Phi}_{xy}(\omega) &= \sum_{t=o}^{T} x_t e^{i\omega t} \sum_{s=o}^{T} y_s e^{-i\omega s} \\ &= \sum_{\tau=-T}^{T} e^{-i\omega\tau} \sum_{t=o}^{T} x_t y_{t+\tau} \end{aligned} \tag{3.371}$$

where $\tau = s-t$. The cross-correlation function of x_t and y_t is defined to be

$$\phi_{xy}(\tau) = \sum_{t=o}^{T-\tau} x_t y_{t+\tau} \qquad (3.38)$$

where

$$\phi_{xy}(-\tau) = \phi_{yx}(\tau)$$

$$\phi_{xy}(\tau) = 0 \text{ for } |\tau| > T \qquad (3.381)$$

Equation (3.371) becomes

$$\overline{\Phi}_{xy}(\omega) = \sum_{\tau=-T}^{T} \phi_{xy}(\tau)\, e^{-i\omega\tau} \qquad (3.382)$$

We have

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{\Phi}_{xy}(\omega)\, e^{i\omega t}\, d\omega = \phi_{xy}(t). \qquad (3.39)$$

3.4 The Filter Problem for Transients. Finite Linear Operators

The filter problem is concerned with the determination of the linear operator which transforms the transient time series x_t $(t = 0,1,2,\ldots T)$ into the transient time series s_t $(t = 0,1,2,\ldots T)$. The time series x_t may be called the input, the signal plus noise, or the perturbed signal. The time series s_t may be called the output, the signal, or the

desired information. Under the aperiodic functional scheme, the observational values of x_t ($t = 0,1,2,\ldots T$) and s_t ($t = 0,1,2,\ldots T$) are assumed to be known, and not all equal to zero.

In this section we shall consider the case in which the linear operator is required to have a finite number of coefficients, and more particularly, the same number of coefficients as the number of terms in the time series x_t. Thus the desired linear operation is represented by

$$s_t = \sum_{s=o}^{T} a_s x_{t-s}, \quad t = 0,1,2,\ldots T \tag{3.41}$$

This equation represents the system of simultaneous linear equations, given by

$$\begin{aligned} s_o &= a_o\, x_o \\ s_1 &= a_o\, x_1 + a_1\, x_o \\ s_2 &= a_o\, x_2 + a_1\, x_1 + a_2\, x_o \\ s_T &= a_o\, x_T + a_1\, x_{T-1} + a_2\, x_{T-2} + \ldots + a_T x_o, \end{aligned} \tag{3.42}$$

in which the operator coefficients a_o, a_1, a_2,... a_T are the unknowns. Without loss of generality, we may assume $x_o \neq 0$,

and consequently the determinant of this system, which is equal to x_o^T, does not vanish. As a result, the system of equations has a unique solution which yields the values a_o, a_1, a_2, ... a_T for the operator coefficients.

Under the aperiodic functional scheme we set $x_t = 0$ for t less than zero and for t greater than T. As a result if we assume the equation

$$s_t = \sum_{s=o}^{T} a_s x_{t-s} \tag{3.43}$$

holds for all t $(-\infty < t < \infty)$, then we see that this equation specifies the values of s_t outside of the range $0 \leq t \leq T$, in which range the values of s_t were given in the original statement of the problem. In particular, this specification is

$$\begin{aligned} s_t &= 0, \quad t < 0, \\ s_t &= \sum_{s=o}^{T} a_s x_{t-s}, \quad t = T+1, T+2, \ldots 2T \\ s_t &= 0, \quad t > 2T. \end{aligned} \tag{3.44}$$

Consequently, for these values of s_t the operator equation (3.43) is valid for all integer values of t. Thus by multiplying each side of equation (3.43) by $e^{-i\omega t}$ and summing over

t, we have

$$\sum_{t=-\infty}^{\infty} s_t e^{-i\omega t} = \sum_{t=-\infty}^{\infty} e^{-i\omega t} \sum_{s=o}^{T} a_s x_{t-s}$$

$$= \sum_{s=o}^{T} a_s e^{-i\omega s} \sum_{t=-\infty}^{\infty} x_{t-s}\, e^{-i\omega(t-s)} \qquad (3.45)$$

which is

$$\sum_{t=o}^{2T} s_t e^{-i\omega t} = \sum_{s=o}^{T} a_s e^{-i\omega s} \sum_{n=o}^{T} x_n e^{-i\omega n} \qquad (3.451)$$

Letting

$$S(\omega) = \sum_{t=o}^{2T} s_t e^{-i\omega t} \qquad (3.452)$$

$$A(\omega) = \sum_{s=o}^{T} a_s e^{-i\omega s} \qquad (3.453)$$

and

$$X(\omega) = \sum_{n=o}^{T} x_n e^{-i\omega n} \qquad (3.454)$$

equation (3.45) becomes

$$S(\omega) = A(\omega) \quad X(\omega) \tag{3.46}$$

which is

$$A(\omega) = \frac{S(\omega)}{X(\omega)} = \sum_{s=o}^{m} a_s e^{-i\omega s} \tag{3.461}$$

The linear operator a_s ($s = 0,1,\ldots m$), determined from $X(\omega)$ and $S(\omega)$ by this equation, is not necessarily stable.

In Figure 4, a symmetric smoothing type linear operator is given which contracts a symmetric Ricker wavelet (Ricker, 1953b) into one of lesser breadth. The respective Fourier transforms are shown to the right of these time functions.

3.5 The Filter Problem for Transients Infinite Linear Operators

In this section we wish to consider the filtering problem, which is the transformation of the time series x_t ($t = 0,1,2,\ldots T$) into the time series s_t ($t = 0,1,2,\ldots T$) under the aperiodic hypothesis that

$$s_t = x_t = 0, \text{ for } t < 0, \; t > T. \tag{3.51}$$

That is, the values of s_t and x_t are specified for all values of t. Thus we wish to find the linear operator, with coefficients h_s, which we shall not restrict in number, such that

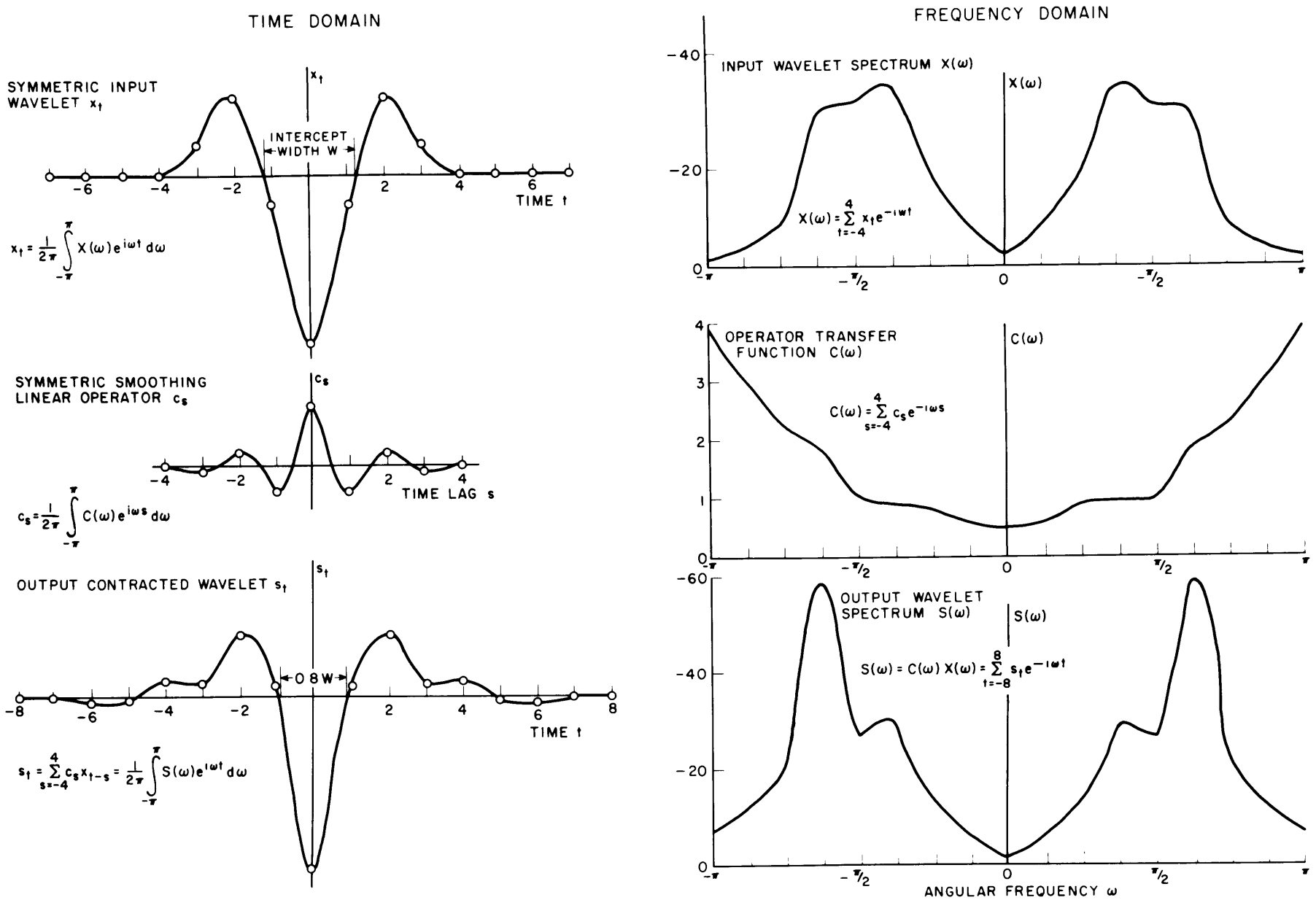


FIGURE 4 RICKER WAVELET CONTRACTION BY LINEAR OPERATORS

$$s_t = \sum_{s=-\infty}^{\infty} h_s x_{t-s}, \quad -\infty < t < \infty. \tag{3.52}$$

Since this equation represents an infinite system of simultaneous equations, that is one equation for each value of t, this system in general will require an infinite number of unknowns h_s for a solution. That is, the linear operator is allowed to be a smoothing operator of infinite extent.

The formal solution for this linear operator may be found in terms of its transfer function

$$H(\omega) = \sum_{s=-\infty}^{\infty} h_s e^{-i\omega s} \tag{3.53}$$

Equation (3.53) becomes

$$\sum_{t=-\infty}^{\infty} s_t e^{-i\omega t} = \sum_{t=-\infty}^{\infty} e^{-i\omega t} \sum_{s=-\infty}^{\infty} h_s x_{t-s} \tag{3.54}$$

$$= \sum_{s=-\infty}^{\infty} h_s e^{-i\omega s} \sum_{t=-\infty}^{\infty} x_{t-s}\, e^{-i\omega(t-s)}$$

which is

$$\sum_{t=o}^{T} s_t e^{-i\omega t} = \sum_{s=-\infty}^{\infty} h_s e^{-i\omega s} \sum_{n=o}^{T} x_n e^{-i\omega n} \tag{3.541}$$

Letting

$$X(\omega) = \sum_{n=-\infty}^{\infty} x_n e^{-i\omega n} = \sum_{n=o}^{T} x_n e^{-i\omega n} \qquad (3.542)$$

and

$$S(\omega) = \sum_{t=-\infty}^{\infty} s_t e^{-i\omega t} = \sum_{t=o}^{T} s_t e^{-i\omega t} \qquad (3.543)$$

equation (3.541) becomes

$$S(\omega) = H(\omega) \quad X(\omega). \qquad (3.55)$$

Thus the formal solution, given by the transfer function of the desired linear operator, is

$$\sum_{s=-\infty}^{\infty} h_s e^{-i\omega s} = H(\omega) = \frac{S(\omega)}{X(\omega)} = \frac{S\bar{X}}{X\bar{X}} \qquad (3.551)$$

The linear operator with coefficients h_s is not necessarily stable. The real part of $H(\omega)$ is

$$Re[H(\omega)] = \frac{1}{2} \quad [H+\bar{H}] = \frac{1}{2} \left[\frac{S}{X} + \frac{\bar{S}}{\bar{X}} \right] = \frac{\bar{S}X + S\bar{X}}{2X\bar{X}} \qquad (3.552)$$

and the imaginary part

$$\mathrm{Im}[H(\omega)] = -\frac{1}{2}[H-\bar{H}] = -\frac{1}{2}\left[\frac{S}{X} - \frac{\bar{S}}{\bar{X}}\right] = i\,\frac{\bar{S}X - S\bar{X}}{2X\bar{X}} \tag{3.553}$$

This result corresponds to that given by Smith (1954) obtained by requiring that the sum of squared errors

$$I = \sum_{t=o}^{T} \left(s_t - \sum_{s=-\infty}^{\infty} h_s x_{t-s}\right)^2$$

$$= 2\pi \int_{-\pi}^{\pi} |S(\omega) - H(\omega)\, X(\omega)|^2 \, d\omega \geq 0 \tag{3.56}$$

be a minimum. We see that the linear operator thus found, with transfer function $H = S/X$ yields the minimum value

$$I_{min} = 2\pi \int_{-\pi}^{\pi} \left|S(\omega) - \frac{S(\omega)}{X(\omega)}\, X(\omega)\right|^2 d\omega$$

$$= 2\pi \int_{-\pi}^{\pi} |S(\omega) - S(\omega)|^2 \, d\omega = 0 \tag{3.561}$$

which is zero.

In other words, the linear operator h_s precisely transforms the input x_t into the signal s_t. Consequently the addition of more information, say in the form of a second input time series y_t, cannot improve this transformation in the sense of reducing the error, which is already zero for all values of t. That is, let us consider the transformation

$$s_t = \sum_{s=-\infty}^{\infty} a_s x_{t-s} + \sum_{s=-\infty}^{\infty} b_s y_{t-s} \qquad (3.57)$$

where s_t, x_t, y_t are specified to be equal to zero for $t < 0$ and $t > T$. The formal solution of this equation is given by the $A(\omega)$ and $B(\omega)$ for which the equation

$$S(\omega) = A(\omega)\, X(\omega) + B(\omega)\, Y(\omega), \qquad (3.571)$$

holds. This equation holds for $A(\omega)$ and $B(\omega)$ given by

$$A(\omega) = \gamma \frac{S(\omega)}{X(\omega)} \qquad (3.572)$$

$$B(\omega) = (1-\gamma) \frac{S(\omega)}{Y(\omega)}$$

where γ is any number. The solution given by equation (3.551) is the case for which $\gamma = 1$. Since γ is arbitrary there is no unique solution for the a_s and b_s in equation (3.57).

This condition is reflected in the fact that the coherency matrix, given by,

$$\begin{bmatrix} \bar{\Phi}_{xx}(\omega) & \bar{\Phi}_{xy}(\omega) \\ \bar{\Phi}_{yx}(\omega) & \bar{\Phi}_{yy}(\omega) \end{bmatrix} = \begin{bmatrix} \bar{X}\bar{X} & \bar{X}\bar{Y} \\ \bar{Y}\bar{X} & \bar{Y}\bar{Y} \end{bmatrix} \tag{3.58}$$

is singular, that is, its determinant is equal to zero. That is, since we have complete knowledge of s_t, and indeed may represent s_t in functional form in terms of x_t alone, as given by equation (3.52), we can introduce no new information about s_t into its representation in the form of an additional time series y_t. Further discussion of this type of problem is given in Section 5.7.

3.6 Averaging and the Probabilistic Point of View

For practical purposes one may only utilize linear operators with a finite number of coefficients. As a result any linear operator with an infinite number of coefficients must be approximated by one with a finite number of coefficients. Thus, for example, we must approximate the infinite linear operator h_s $(-\infty < s < \infty)$ of the preceding section by a finite linear operator, say h'_s $(-m \leq s \leq m)$, or, in other words, approximate the transfer function

$$H(\omega) = \sum_{s=-\infty}^{\infty} h_s e^{-i\omega s} = \frac{S(\omega)}{X(\omega)} \qquad (3.551)$$

by the transfer function

$$H'(\omega) = \sum_{s=-m}^{m} h'_s e^{-i\omega s}. \qquad (3.61)$$

Such an approximation procedure requires that a certain amount of information contained in the infinite linear operator h_s be lost in order to obtain the approximate finite linear operator h'_s, and consequently we shall need to utilize some type of averaging process to carry out this approximation. One such averaging process is to require that the sum of squared errors

$$I = \sum_{t=-\infty}^{\infty} (s_t - \sum_{s=-m}^{m} h'_s e^{-i\omega s})^2 = 2\pi \int_{-\pi}^{\pi} |S(\omega) - H'(\omega)X(\omega)|^2 d\omega \qquad (3.62)$$

be a minimum. Smith (1954) writes this equation in the form

$$I = 2\pi \int_{-\pi}^{\pi} |X(\omega)|^2 \left|\frac{S(\omega)}{X(\omega)} - H'(\omega)\right|^2 d\omega = 2\pi \int_{-\pi}^{\pi} X\bar{X}(H-H')(\bar{H}-\bar{H}') d\omega \qquad (3.63)$$

since $H = S/X$.

Thus, in many cases, even if one utilizes a purely functional approach it is necessary to develop methods of averaging in order to obtain certain desired goals. These methods of averaging may be carried out with respect to certain desirable criteria, where such criteria may be established and justified from a purely mathematical or functional point of view.

Although one may work with various averaging procedures with no recourse to the theory of probability, in many applications it is not until probability theory is introduced that certain procedures become meaningful in a physical sense. The theory of stationary time series, as conceived by Yule (1921, 1926, 1927), and established in full generality by Khintchine (1934), makes use of averaging procedures which were arrived at from the probability point of view. This theory has found many applications in pure and applied science. In fact, Wiener (1942) emphasizes the probabilistic or statistical point of view for engineering problems and applies the theory of stationary time series toward their solution.

Although seismic records are not stationary in the sense of Khintchine (1934), nevertheless one may treat sections of these records as being approximately stationary and consequently apply statistical methods to their analyses (Wadsworth, et al,

1953). Further discussion of this problem is given in Chapter VI. For numerical computational purposes discrete time series must be utilized, and so in the next two chapters we shall present theory of discrete stationary time series. Much of this theory was first established as a general theory by the Swedish statistician and economist, Herman Wold (1938), the work of Khintchine (1934) being confined to continuous stationary time series.

CHAPTER IV

THEORY OF DISCRETE STATIONARY TIME SERIES

4.1 Random Processes

In a recent paper by Wadsworth, et al, (1953), concepts from the theory of stationary time series were presented. In this chapter, and in the next, we wish to extend this presentation in regard to discrete stationary time series, and in particular to develop the concept of the predictive decomposition of stationary time series. For more comprehensive presentations of the theory of discrete stationary time series, the reader is referred to Wold (1938), Doob (1953), and Wold (1953).

A discrete time series is a sequence of equidistant observations x_t which are associated with the discrete time parameter t. Without loss of generality we may take the spacing between each successive observation to be one unit of time, and thus we may represent the time series as

$$\cdots, x_{t-2}, x_{t-1}, x_t, x_{t+1}, x_{t+2}, \cdots \qquad (4.11)$$

where t takes on all integer values from minus infinity to

infinity ($-\infty < t < \infty$). Thus all angular frequencies ω may be required to lie between $-\pi$ and π.

Any observational time series x_t ($-\infty < t < \infty$) may be considered as a realization of a so-called random process, or stochastic process, which is a mathematical abstraction defined with respect to a probability field. In many phenomena, the number of observational time series supplied by a random process is limited. It is often the case, especially in economic applications, that only one time series is generated by a random process. Such a case, nevertheless, is in full accord with the frequency interpretation of probability.

4.2 Stationary Time Series

A time series is said to be stationary if the probabilities involved in the stochastic process are not tied down to a specific origin in time; that is, the probability of any event associated with the time t is equal to the probability of the corresponding event associated with the time $t + \tau$, where t and τ are any integer values.

For any stochastic process, one may form averages with respect to the statistical population or "ensemble" of realizations x_t for a fixed value of time t. Such averages are called ensemble averages or space averages, and we shall

denote such an averaging process by the expectation symbol E. In particular the mean value $m = E[x_t]$ and the variance $\sigma^2 = E[(x_t - m)^2]$ of a stationary stochastic process are independent of time t. Likewise, the (unnormalized) autocorrelation coefficients

$$\phi(\tau) = E[x_t \, x_{t+\tau}] \tag{4.21}$$

are independent of t, and constitute an even function of the time lag τ, that is

$$\phi(\tau) = \phi(-\tau). \tag{4.22}$$

Also we have $|\phi(\tau)| \quad \phi(0).$ (4.221)

The normalized autocorrelation function is defined to be

$$\phi(\tau) = \frac{E[(x_t-m)(x_{t+\tau}-m)]}{E[x_t-m)]^2} \tag{4.222}$$

so that

$$\phi(0) = 1, \; |\phi(\tau)| \leq 1. \tag{4.223}$$

In what follows we shall assume that the mean value m is equal to zero which we may do without loss of generality. Also we shall utilize the unnormalized autocorrelation (4.21).

There is another type of average known as a time average or phase average in which the averaging process is carried out with respect to all values of time t for a fixed realization $x_t(-\infty < t < \infty)$ of the stochastic process. If a stationary process has the property that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{\tau=1}^{T} \phi(\tau) = 0, \tag{4.23}$$

the process is called an ergodic process, and the ensemble averages and time averages are equal with probability one. As a result the autocorrelation of an ergodic process may be expressed as the time average

$$\phi(\tau) = \lim_{T\to\infty} \frac{1}{2T+1} \sum_{t=-T}^{T} x_{t+\tau}\, x_t. \tag{4.24}$$

4.3 The Autocorrelation

The autocorrelation function $\phi(\tau)$ is a non-negative definite function, that is,

$$\phi(\tau) = \phi(-\tau),$$

$$\sum_{j=1}^{N} \sum_{k=1}^{N} \phi(j-k)\, a_j a_k \geq 0, \quad N = 1,2,\ldots \tag{4.31}$$

for every real set of $a_1, a_2, \ldots a_N$. Thus, the autocorrelation matrix ($N = 1,2,\ldots$)

$$\begin{bmatrix} \phi(0) & \phi(1) & \ldots & \phi(N) \\ \phi(-1) & \phi(0) & \ldots & \phi(N-1) \\ & \ldots & \ldots & \\ \phi(-N) & \phi(-N+1) & .. & \phi(0) \end{bmatrix} \tag{4.32}$$

is symmetric, has its elements equal along its diagonal and along any super or sub diagonal, has non-negative eigenroots $\lambda_j \geq 0$ ($j = 1,2,\ldots N$), has a non-negative determinant, and has a non-negative definite quadratic form given by equation (3.31). The non-negative definiteness of the autocorrelation follows from the inequality (Wold, 1938)

$$\sum_{j=1}^{N} \sum_{k=1}^{N} \phi(j-k)\, a_j a_k = \sum_{j=1}^{N} \sum_{k=1}^{N} \phi(|j-k|)\, a_j a_k$$

$$= \sum_{j=1}^{N} \sum_{k=1}^{N} E[a_j a_k x_j x_k]$$

$$= E[(\sum_{j=1}^{N} a_j x_j)^2] \geq 0 \tag{4.33}$$

4.4 The Spectrum

The property that the autocorrelation function $\phi(\tau)$ is non-negative definite is equivalent to its representation by the Fourier transform

$$\phi(\tau) = \frac{1}{\pi} \int_0^{\pi} \cos \omega\tau \, d\, \Lambda(\omega) \qquad (4.41)$$

where $\Lambda(\omega)$, called the integrated spectrum or the spectral distribution function, in a real monotone non-decreasing function of $\omega (0 \leq \omega \leq \pi)$ with $\Lambda(0) = 0$ and $\Lambda(\pi) = \pi$. This theorem, usually known as the Wiener-Khintchine theorem, was used by Wiener (1930) and was first used in this setting by Khintchine (1934) in his development of the theory of continuous stationary time series. The theorem for discrete stationary time series stated here was first given by Wold (1938).

The inversion formula (Wold, 1938) expresses the integrated spectrum in terms of the autocorrelation, that is,

$$\Lambda(\omega) = \omega + 2 \sum_{\tau=1}^{\infty} \frac{\phi(\tau)}{\tau} \sin \omega\tau, \quad 0 \leq \omega \leq \pi \qquad (4.42)$$

Moreover, Wold (1938) shows that if $\sum_{\tau=0}^{\infty} |\phi(\tau)|$ is convergent, then $\Lambda(\omega)$ will be absolutely continuous, with the

continuous derivative

$$\Lambda'(\omega) = \frac{d\,\Lambda(\omega)}{d\omega} = \overline{\Phi}(\omega) = 1 + 2 \sum_{\tau=1}^{\infty} \phi(\tau) \cos \omega\tau$$

$$= \sum_{\tau=-\infty}^{\infty} \phi(\tau) \cos \omega\tau \qquad (4.43)$$

In the remaining part of this thesis, we shall confine ourselves to stochastic processes for which the spectral distribution function $\Lambda(\omega)$ is absolutely continuous and for which $\Sigma|\phi(\tau)| < \infty$, unless it is otherwise stated. Wiener (1942) restricts himself to those processes where the spectral distribution function $\Lambda(\omega)$ is absolutely continuous, whereas Wold (1938) and Kolmogorov (1939,1941) treat such processes as special cases.

The derivative $\overline{\Phi}(\omega) = \Lambda'(\omega)$ is called the spectral density function, the power spectrum, or simply the spectrum. Since $\overline{\Phi}(\omega)$ is the slope of a real monotone non-decreasing function $\Lambda(\omega)$, we have

$$\overline{\Phi}(\omega) \geq 0, \qquad (4.44)$$

and $\overline{\Phi}(\omega)$ may be considered to represent the power density in the time series x_t $(-\infty < t < \infty)$. In order to have equal

power at ω and $-\omega$, let us define $\overline{\Phi}(\omega)$ to be an even function of ω, that is,

$$\overline{\Phi}(-\omega) = \overline{\Phi}(\omega), \quad -\pi \leq \omega \leq \pi \tag{4.45}$$

Equation (3.41) thus becomes

$$\phi(\tau) = \frac{1}{\pi} \int_0^{\pi} \cos \omega\tau \ \overline{\Phi}(\omega)\, d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega\tau} \overline{\Phi}(\omega)\, d\omega \tag{4.46}$$

and, in particular, we have

$$\phi(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{\Phi}(\omega)\, d\omega \tag{4.461}$$

To show that the autocorrelations $\phi(\tau)$ as given by equation (4.46) is a non-negative definite function, we let

$$H_N(\omega) = \sum_{k=1}^{N} a_k e^{-i\omega k} \tag{4.47}$$

where a_k $(k = 1,2,\ldots N)$ is any arbitrary set of real numbers. Then the quadradic form of the autocorrelation matrix is

$$\sum_{j=1}^{N} \sum_{k=1}^{N} \phi(j-k)\, a_j a_k = \int_{-\pi}^{\pi} \sum_{j=1}^{N} \sum_{k=1}^{N} a_j a_k e^{i\omega(j-k)}\, \overline{\Phi}(\omega)\, d\omega \tag{4.48}$$

which is equal to

$$\int_{-\pi}^{\pi} H_N(\omega)\ \overline{H_N(\omega)}\ \Phi(\omega)\, d\omega = \int_{-\pi}^{\pi} |H_N(\omega)|^2\ \Phi(\omega)\, d\omega \geq 0,\ N=1,2,\ldots \tag{4.49}$$

Thus equation (4.31) is verified for $\phi(\tau)$ given by equation (4.46)

The important role played by the harmonic analysis of a stationary time series is brought out by the Spectral Representation Theorem due to Harald Cramer (1942) and others. The theorem allows the time series x_t to be represented by a stochastic integral which involves the harmonic components of the time series. For a statement of and a discussion of this theorem, the reader is referred to Doob (1953, p 481), who also gives the historical background of this theorem (Doob, 1953, p 637).

4.5 Processes with White Light Spectra

A process is said to have a white light spectrum if its power spectrum has a constant value, that is,

$$\Phi(\omega) = \text{constant},\ -\pi \leq \omega \leq \pi. \tag{4.51}$$

In this section we wish to consider two types of processes which have white light spectra, namely, the purely random

process and the mutually uncorrelated process. For a further discussion of these processes, see Doob (1953) and Wold (1953). As a matter of terminology, one should distinguish between the "purely random process" and a "random process". The purely random process is defined in this section. On the other hand, a random or stochastic process designates any process which generates one or more observational time series, and such processes range from purely random processes to non-random or deterministic processes.

A realization from a purely random process is the time series $\xi_t (-\infty < t < \infty)$ where each ξ_t is an independent random variate from a single probability distribution function. Therefore the joint distribution functions of the ξ_t's are simply products of this probability distribution function. Consequently, we have

$$E[\xi_t \xi_s] = E[\xi_t] E[\xi_s] \quad s \neq t \tag{4.52}$$

Such a process is stationary and ergodic. In order to normalize the process so that it will have zero mean and unit variance, we let

$$E[\xi_t] = 0, \; E[\xi_t^2] = 1. \tag{4.53}$$

The autocorrelation function is then given by

$$\phi(0) = E[\xi_t^2] = 1$$

$$\phi(\tau) = E[\xi_t \xi_{t+\tau}] = E[\xi_t]\, E[\xi_{t+\tau}] = 0, \quad \tau = \pm 1, \pm 2, \pm 3, \ldots \tag{4.531}$$

An alternative assumption as to the nature of the ξ_t leads to the definition of the so-called mutually uncorrelated process. That is, if instead of assuming the ξ_t and $\xi_{t+\tau}$ are independent, we assume for a mutually uncorrelated process that they are uncorrelated in pairs, that is

$$E[\xi_t \xi_s] = E[\xi_t]\, E[\xi_s] \quad s \neq t \tag{4.54}$$

which is the same as equation (4.52). That is, independent random variables are uncorrelated, but the converse is not necessarily true. Again we shall normalize the ξ_t as in equation (4.53), in which case the uncorrelated random variables ξ_t are called orthogonal random variables. The orthogonality is illustrated by

$$E[\xi_t^2] = 1,$$

$$E[\xi_t\, \xi_s] = 0. \quad t \neq s \tag{4.55}$$

In what follows, we shall assume all uncorrelated random variables are normalized in this manner (equation (4.53)) so, alternatively, we may call them orthogonal random variables. Since equation (4.531) holds for a (normalized) uncorrelated process we see that this process has the same autocorrelation coefficients as the purely rendom process. Also, an uncorrelated process is stationary and ergodic.

The spectral distribution function, equation (4.42), of a purely random process or of an uncorrelated process is given by

$$\Lambda(\omega) = \omega, \quad 0 \leq \omega \leq \pi \tag{4.56}$$

and the power spectrum, by equation (4.43), is a constant given by

$$\Phi(\omega) = 1, \quad -\pi \leq \omega \leq \pi \tag{4.57}$$

These processes therefore have white light spectra, and the ξ_t may be called "white" noise (Wiener, 1930).

In summary, then, the purely random process and the uncorrelated process both have white light spectra and have autocorrelation coefficients which vanish except for lag zero.

4.6 Processes of Moving Summation

An important theorem states that a time series with an absolute continuous spectral distribution function is generated by a process of moving summation, and conversely. In this section we shall define what is meant by a process of moving summation, and then indicate in an heuristic way why this theorem holds. For a rigorous proof, the reader is referred to Doob (1953).

For the fixed realization

$$\cdots \xi_{t-1},\ \xi_t,\ \xi_{t+1},\ \xi_{t+2}, \cdots \tag{4.61}$$

of a purely random process or of a mutually uncorrelated process, the corresponding fixed realization of a process of moving summation is

$$\cdots x_{t-1},\ x_t,\ x_{t+1},\ x_{t+2}, \cdots \tag{4.62}$$

where

$$x_n = \sum_{k=-\infty}^{\infty} c_k \xi_{n-k},\quad n = t,\ t\pm 1,\ t\pm 2, \ldots \tag{4.621}$$

Since two random variables are uncorrelated if they are independent, whereas the converse is not always true, in what follows we shall impose only the weaker restriction on the

ξ_t and thus only assume they are mutually uncorrelated (ie. orthogonal) in the definition of the process of moving summation. (See Section 4.5). In particular, we assume

$$E[\xi_t] = 0,\ E[\xi_t^2] = 1,\ E[\xi_t \xi_s] = E[\xi_t]\,E[\xi_s] = 0 \text{ for } t \neq s. \tag{4.63}$$

The mean of the x_t process is

$$E[x_t] = \sum_{k=-\infty}^{\infty} c_k E[\xi_{t-k}] = 0 \tag{4.64}$$

and the variance is

$$E[x_t^2] = \sum_{k=-\infty}^{\infty} c_k^2 \quad E[\xi_{t-k}^2] = \sum_{k=-\infty}^{\infty} c_k^2. \tag{4.641}$$

The autocorrelation coefficients are

$$\phi(\tau) = E[x_t x_{t+\tau}] = \sum_{t=-\infty}^{\infty} c_{t-\tau} c_t = \sum_{t=-\infty}^{\infty} c_t c_{t+\tau}. \tag{4.642}$$

We now wish to indicate why a process of moving summation has an absolutely continuous spectral distribution function.

Let x_t be a process of moving summation given by equation (4.621). Define $F(\omega)$ by

$$F(\omega) = \sum_{k=-\infty}^{\infty} c_k e^{-i\omega k} \tag{4.65}$$

which is the transfer function of the infinite smoothing operator c_k. Then, by equation (4.642) we have

$$\phi(\tau) = \sum_{j=-\infty}^{\infty} c_j c_{j-\tau} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{j=-\infty}^{\infty} c_{j-\tau} e^{i\omega j} \sum_{k=-\infty}^{\infty} c_k e^{-i\omega k} \tag{4.651}$$

since

$$\int_{-\pi}^{\pi} e^{i\omega j} e^{-i\omega k} d\omega = \begin{cases} 2\pi, & j=k \\ 0 & j\neq k \end{cases} \tag{4.652}$$

for integer j and k, by letting $j-\tau = \ell$, we have

$$\phi(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega\tau} \sum_{\ell=-\infty}^{\infty} c_\ell e^{i\omega\ell} \sum_{k=-\infty}^{\infty} c_k e^{-i\omega k} d\omega. \tag{4.653}$$

Then, using equation (4.65), we have

$$\phi(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\omega)\ \overline{F(\omega)}\ e^{i\omega\tau}\, d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} |F(\omega)|^2\ e^{i\omega\tau}\, d\omega \tag{4.654}$$

which is in the form of equation (4.46) with the power spectrum

$$\overline{\Phi}(\omega) = |F(\omega)|^2. \tag{4.655}$$

Thus the spectral distribution function

$$\Lambda(\omega) = \int_{0}^{\omega} \overline{\Phi}(u)\, du \tag{4.656}$$

is absolutely continuous.

Conversely, any process with absolutely continuous spectral distribution is a process of moving summation, and in this paragraph we wish to indicate some reasons for this theorem. Because of equation (4.44) which states that the spectrum $\overline{\Phi}(\omega)$ is non-negative for every value of ω, we may set

$$\overline{\Phi}(\omega) = |F(\omega)|^2 = F(\omega)\ \overline{F(\omega)} \tag{4.66}$$

where $F(\omega)$ is the Fourier series of any square root of $\overline{\Phi}(\omega)$.

Let us represent this Fourier series by

$$F(\omega) = \sum_{k=-\infty}^{\infty} c_k e^{-i\omega k}. \tag{4.661}$$

Using the coefficients c_k we may define the process of moving summation

$$x_t = \sum_{k=-\infty}^{\infty} c_k \xi_{t-k}. \tag{4.662}$$

The autocorrelation of the process (4.662) will be given by equation (4.642). The Fourier transform of this autocorrelation function gives the power spectrum of the process (4.662) which is the same as the original power spectrum in equation (4.66). Thus the process x_t given by equation (4.662) is a process of moving summation which has the given power spectrum (4.66).

Hence, for the process of moving summation represented by

$$x_t = \sum_{k=-\infty}^{\infty} c_k \xi_{t-k}, \tag{4.621}$$

the c_k represents a linear operator, and the ξ_{t-k} represents white "noise". The transfer function is given by $F(\omega)$ in equation (4.65), and the power transfer function is then the

power spectrum of the process, that is

$$\Phi(\omega) = |F(\omega)|^2. \qquad (4.655)$$

Thus the time series x_t is the output of a linear system, with power transfer function $\Phi(\omega)$, into which white "noise" is passed. Since the c_k may be an infinite smoothing operator, this system need not necessarily be realizable or stable. In Section 5.1 we shall see that a unique realizable and stable prediction operator may be found, which leads to the predictive decomposition of stationary time series given in Section 5.2.

Finally, let us note that processes of moving summation are ergodic, and for a proof the reader is referred to Doob (1953) and Wold (1953).

CHAPTER V

THE PREDICTIVE DECOMPOSITION OF STATIONARY TIME SERIES

5.1 The Factorization of the Spectrum

In the preceeding chapter we saw that the power spectrum of a time series x_t may be regarded as a power transfer function of a linear system into which white noise ζ_t is passed in order to obtain the time series x_t as output. The gain characteristic of this linear system is $\sqrt{\Phi(\omega)}$. The problem of the factorization of the spectrum is the problem of determining the phase characteristic so that the system is physically realizable and stable, with minimum phase characteristic for the gain $\sqrt{\Phi(\omega)}$. Thus the transfer function of the desired physically realizable minimum phase network may be given by

$$B(\omega) = |B(\omega)|\ e^{i\Theta(\omega)} = \sqrt{\Phi(\omega)}\ \ e^{i\Theta(\omega)} \qquad (5.11)$$

where $\sqrt{\Phi(\omega)}$ is the gain, and $\Theta(\omega)$ represents the desired minimum phase characteristic.

Kolmogorov (1939) gave the general solution of this factorization problem. A rigorous exposition of his results may be found in Doob (1953), and in this section we wish to give

an heuristic exposition.

Let us first turn our attention to the properties of a realizable, stable linear system with minimum phase-shift characteristic. As we have seen in Section 2.7, the conditions that the transfer function be physically realizable and minimum phase is that it may be expressed as

$$B(\omega) = \sum_{s=o}^{\infty} b_s e^{-i\omega s} \tag{5.12}$$

where

$$b_s = 0 \text{ for } s < 0 \tag{5.121}$$

$$\sum_{s=o}^{\infty} b_s^2 < \infty, \tag{5.122}$$

and

$$B(\lambda) = \sum_{s=o}^{\infty} b_s e^{-i\lambda s}, \quad \lambda = \omega + i\,\sigma \tag{5.123}$$

has no singularities or zeros in the lower half λ plane ($\sigma < 0$). Under the transformation $z = e^{-i\lambda}$ (see Figure 3), this last condition becomes that

$$B(z) = \sum_{s=o}^{\infty} b_s z^s \qquad (5.124)$$

have no singularities or zeros for $|z| \leq 1$. Under these conditions, log B(z) will be analytic for $|z| \leq 1$, and consequently has the power series representation

$$\log B(z) = \sum_{t=o}^{\infty} \beta_t z^t \text{ for } |z| \leq 1, \qquad (5.125)$$

and, as $|z|$ approaches 1, this series converges to

$$\log B(\omega) = \sum_{t=o}^{\infty} \beta_t e^{-i\omega t} = \beta_o + \sum_{t=1}^{\infty} \beta_t \cos \omega t - i \sum_{t=1}^{\infty} \beta_t \sin \omega t. \qquad (5.126)$$

Let us now turn our attention to the power spectrum $\Phi(\omega)$. The spectrum $\Phi(\omega)$ is a real function of ω, such that

$$\Phi(\omega) = \Phi(-\omega), \ \Phi(\omega) \geq 0, \ -\pi < \omega \leq \pi. \qquad (5.13)$$

Moreover, the following conditions on $\Phi(\omega)$ must be satisfied:

$$\Phi(\omega) = 0 \qquad (5.131)$$

at most on a set of Lebesque measure zero,

$$\int_{-\pi}^{\pi} \Phi(\omega)\, d\omega < \infty \tag{5.132}$$

and

$$\int_{-\pi}^{\pi} \log \Phi(\omega)\, d\omega > -\infty. \tag{5.133}$$

Under these conditions, $\log \sqrt{\Phi(\omega)}$, which is an even real function of ω, may be expressed in the real Fourier cosine series

$$\log \sqrt{\Phi(\omega)} = \sum_{t=-\infty}^{\infty} \alpha_t \cos \omega t \tag{5.134}$$

where the Fourier coefficients α_t are given by

$$\alpha_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos \omega t \log \sqrt{\Phi(\omega)}\, d\omega$$

$$= \frac{1}{2\pi} \int_{0}^{\pi} \cos \omega t \log \Phi(\omega)\, d\omega. \tag{5.135}$$

From this equation we see that α_t is an even real function of t, that is

$$\alpha_t = \alpha_{-t}. \tag{5.136}$$

Consequently the Fourier expansion (5.134) of $\log \sqrt{\Phi(\omega)}$ becomes

$$\log \sqrt{\Phi(\omega)} = \alpha_o + 2 \sum_{t=1}^{\infty} \alpha_t \cos \omega t. \qquad (5.137)$$

Equation (5.11), which gives the transfer function $B(\omega)$ for the desired minimum phase network, is

$$B(\omega) = |B(\omega)| \, e^{i\Theta(\omega)} = \sqrt{\Phi(\omega)} \quad e^{i\Theta(\omega)} \qquad (5.11)$$

where $\sqrt{\Phi(\omega)}$ is the gain, and $\Theta(\omega)$ represents the minimum phase characteristic. By taking the logarithm of each side of this equation, we have

$$\log B(\omega) = \log \sqrt{\Phi(\omega)} + i\,\Theta(\omega) \qquad (5.14)$$

which, by equation (5.137) is

$$\log B(\omega) = \alpha_o + 2 \sum_{t=1}^{\infty} \alpha_t \cos \omega t + i\,\Theta(\omega). \qquad (5.141)$$

Now equation (5.126) gives an expression for $\log B(\omega)$ which was derived from the knowledge that $\log B(z)$ be analytic

for $|z| \leq 1$, whereas equation (5.141) gives an expression for log B(ω) derived from the knowledge that the gain $|B(\omega)|$ be equal to $\sqrt{\Phi(\omega)}$. Setting these two equations equal to each other, we have

$$\log B(\omega) = \beta_o + \sum_{t=1}^{\infty} \beta_t \cos \omega t - i \sum_{t=1}^{\infty} \beta_t \sin \omega t \qquad (5.126)$$

$$= \alpha_o + \sum_{t=1}^{\infty} 2\alpha_t \cos \omega t + i\theta(\omega). \qquad (5.141)$$

We therefore have

$$\text{Re}\,[\log B(\omega)] = \log \sqrt{\Phi(\omega)}$$

$$= \beta_o + \sum_{t=1}^{\infty} \beta_t \cos \omega t$$

$$= \alpha_o + \sum_{t=1}^{\infty} 2\alpha_t \cos \omega t \qquad (5.15)$$

so that

$$\beta_o = \alpha_o, \quad \beta_t = 2\alpha_t \text{ for } t = 1,2,3,\ldots \qquad (5.151)$$

where α_t $(t = 0,1,2,\ldots)$ is given by equation (5.135). We also have

$$\text{Im}\,[\log B(\omega)] = \Theta(\omega) = -\sum_{t=1}^{\infty} \beta_t \sin \omega t \qquad (5.16)$$

which, by equation (5.151) is

$$\Theta(\omega) = -2 \sum_{t=1}^{\infty} \alpha_t \sin \omega t. \qquad (5.161)$$

This equation expresses the minimum phase characteristic $\Theta(\omega)$ in terms of the α_t ($t = 1,2,\ldots$), which are computed from knowledge of the power spectrum $\Phi(\omega)$ by means of equation (5.135).

As a result the operator coefficients b_s may be determined in the following manner. Equations (5.12) and (5.11) give

$$B(\omega) = \sum_{s=o}^{\infty} b_s e^{-i\omega s} = \sqrt{\Phi(\omega)}\; e^{i\Theta(\omega)} \qquad (5.17)$$

which, because of equations (5.15) and (5.161), yields

$$\begin{aligned} \log B(\omega) &= \text{Re}[\log B(\omega)] + i\, \text{Im}[\log B(\omega)] \\ &= \log \sqrt{\Phi(\omega)} + i\, \Theta(\omega) \\ &= \alpha_o + \sum_{t=1}^{\infty} 2\, \alpha_t \cos \omega t - i \sum_{t=1}^{\infty} 2\, \alpha_t \sin \omega t \\ &= \alpha_o + 2 \sum_{t=1}^{\infty} \alpha_t e^{-i\omega t}\,. \end{aligned} \qquad (5.171)$$

Since

$$B(\omega) = e^{\log B(\omega)} \tag{5.172}$$

and using equation (5.12) we therefore have

$$B(\omega) = \sum_{s=o}^{\infty} b_s e^{-i\omega s} = e^{\alpha_o + 2\sum_{t=1}^{\infty} \alpha_t e^{-i\omega t}} . \tag{5.173}$$

Letting $\lambda = \omega + i\sigma$ and making the substitution $z = e^{-i\lambda}$, we have

$$B(z) = \sum_{s=o}^{\infty} b_s z^s = e^{\alpha_o + 2\sum_{t=1}^{\infty} \alpha_t z^t} , \quad |z| \leq 1. \tag{5.174}$$

By means of this equation, we may solve for the linear operator b_s in terms of the α_t. In particular, we have from equations (5.174) and (5.135) that

$$e^{\alpha_o} = b_o = e^{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \Phi(\omega)\, d\omega} > 0. \tag{5.175}$$

Therefore Kolmogorov (1939) shows that the power spectrum $\Phi(\omega)$ may be factored in the following manner:

$$\Phi(\omega) = |B(\omega)|^2 = \left| \sum_{t=o}^{\infty} b_t e^{-i\omega t} \right|^2 \qquad (5.181)$$

where the linear operator b_t may be determined from

$$b_t = 0, \text{ for } \quad t < 0, \qquad (5.121)$$

$$b_o = e^{\frac{1}{2\pi} \int_o^{\pi} \log \Phi(\omega)\, d\omega} > 0, \qquad (5.175)$$

and

$$\sum_{s=1}^{\infty} b_s z^s = e^{2 \sum_{t=1}^{\infty} \alpha_t z^t} . \qquad (5.182)$$

The transfer function of the linear operator b_t, given by

$$B(\omega) = \sum_{t=o}^{\infty} b_t e^{-i\omega t} , \qquad (5.12)$$

has gain

$$|B(\omega)| = \sqrt{\Phi(\omega)} \qquad (5.183)$$

and minimum phase $\Theta(\omega)$ given by equation (5.161). The transfer function $B(\omega)$ is the factor of the spectrum. Since

$$b_t = 0 \text{ for } t < 0, \tag{5.121}$$

$$b_o > 0 \tag{5.175}$$

$$\sum_{t=o}^{\infty} b_t^2 < \infty \tag{5.122}$$

$$\sum_{t=o}^{\infty} |b_t| < \infty \tag{5.184}$$

and

$$\sum_{t=o}^{\infty} b_t z^t \neq 0, \text{ for } |z| \leq 1, \tag{5.124}$$

the linear operator b_o, b_1, b_2...is physically realizable and stable.

In order to conform with Wold's notation, Kolmogorov (1939) normalizes the b_t so that $b_o = 1$. That is, he gives

$$\begin{aligned} B(\omega) &= e^{\frac{1}{2\pi}\int_o^{\pi} \log \Phi(\omega)} \sum_{s=o}^{\infty} b_s e^{-i\omega s}, \; b_o = 1 \\ &= e^{\frac{1}{2\pi}\int_o^{\pi} \log \Phi(\omega)} (1 + b_1 e^{-i\omega} + b_2 e^{-i\omega 2} + \ldots) \end{aligned} \tag{5.19}$$

where

$$\log \overline{\Phi}(\omega) = 2\alpha_o + 4\alpha_1 \cos \omega + 4\alpha_2 \cos 2\omega + \ldots \qquad (5.191)$$

and

$$\exp [2(\alpha_1 z + \alpha_2 z^2 + \ldots.)] = 1 + b_1 z + b_2 z^2 + \ldots \qquad (5.192)$$

Kolmogorov (1939) concludes that the linear operator b_t has here the same significance as in Theorem 7 of the book by Wold (1938). In the next section we shall examine this theorem which is called the Predictive Decomposition Theorem for stationary time series, and in Sections 5.3 and 5.4 we shall see that this theorem is the fundamental theorem used in the solution of the prediction and filtering problems.

For further discussion on the results of this section see Theorems 4.1 and 4.3 of Chapter XII in Doob (1953).

The concept of the factorization of the power spectrum is due to Cramér and Wold (Wold, 1938, p 123). As we shall see in Section 5.5 (together with Section 2.8), Wold (1938) factored the spectra of the autoregressive process and the process of finite moving averages, both of which have rational spectra, in order to yield the realizable and stable linear

operator b_t. Kolmogorov's method given in this section is essentially a generalization of Wold's method to more general spectra, that is, spectra which are not necessarily rational functions. Accordingly we shall refer to $B(\omega)$ given by equation (5.181) as the Wold-Kolmogorov factor of the power spectrum.

Although Wold (1938) did not have Kolmogorov's more general method of determining the operator b_t (Kolmogorov, 1939) Wold did anticipate correctly the existence of such an operator, and accordingly gave the results of the following two sections.

5.2 The Predictive Decomposition Theorem

In the preceding section we have seen that the power spectrum $\Phi(\omega)$ may be factored in the following manner:

$$\Phi(\omega) = B(\omega)\,\overline{B(\omega)} = |B(\omega)|^2 = \Big|\sum_{s=o}^{\infty} b_s e^{-i\omega s}\Big|^2 \tag{5.181}$$

where

$$b_s = 0, \quad s < 0 \tag{5.121}$$

$$b_o > 0 \tag{5.175}$$

$$\sum_{s=o}^{\infty} b_s^2 < \infty \tag{5.122}$$

$$\sum_{s=o}^{\infty} |b_s| < \infty \tag{5.184}$$

and where

$$B(\lambda) = \sum_{s=o}^{\infty} b_s e^{-i\lambda s} \quad , \quad \lambda = \omega + i\,\sigma \qquad (5.123)$$

has no singularities or zeros in the lower half λ plane ($\sigma < 0$). In other words, the linear operator represented by b_o, b_1, b_2,... is realizable and stable, and its transfer function $B(\omega)$ has minimum phase characteristic.

In Section 4.6 we saw that a time series with an absolutely continuous spectral distribution function may be represented by a process of moving summation. We see that equation (5.181) may be used in place of equations (4.66) of Section 4.6. In other words we may replace the linear operator c_t of Section 4.6 by the realizable and stable prediction operator b_t of Section 5.1. Thus the process of moving summation is given by

$$x_t = \sum_{s=o}^{\infty} b_s \xi_{t-s} = b_o \xi_t + b_1 \xi_{t-1} + b_2 \xi_{t-2} + \cdots \qquad (5.21)$$

which replaces equation (4.662) of Section 4.6. In this equation ξ_t ($-\infty < t < \infty$) represents a realization from a mutually uncorrelated process, the x_t is the time series with power spectrum $\overline{\Phi}(\omega)$, and b_o, b_1, b_2,... is the realizable and stable operator determined as in Section 5.1. More particularly, the variables ξ_t ($-\infty < t < \infty$) have zero mean $E(\xi_t) = 0$, unit variance $E(\xi_t^2) = 1$, and are mutually uncorrelated $E(\xi_t \xi_s)$

$= 0$ for $t \neq s$, and consequently have a white light spectrum, (see Section 4.5).

That all stationary processes with absolutely continuous spectral distribution functions may be represented in the form (5.21) is a special case of the more general Predictive Decomposition Theorem of Herman Wold (1938), his Theorem 7. That is, the more general form of this theorem (Wold, 1938) is not restricted to processes with absolutely continuous spectral distribution functions. Statements of and further discussions of this theorem may be found in Wold (1938, Theorem 7), Doob (1953, Theorem 4.2 of Chapter XII, p 576) and Wold (1953, Theorem 1 of Chapter 12.6, p 200).

Let us now state the Predictive Decomposition Theorem for a stationary process with an absolutely continuous spectral distribution: Given a stationary process x_t $(-\infty < t < \infty)$ with discrete time parameter t, suppose x_t $(-\infty < t < \infty)$ has an absolutely continuous spectral distribution. Then x_t $(-\infty < t < \infty)$ allows the decomposition

$$x_t = b_o\xi_t + b_1\xi_{t-1} + b_2\xi_{t-2} + \cdots \qquad (5.21)$$

where the components $(-\infty < t < \infty)$ have the following properties

A. Each of the variables ξ_t is linear in x_t, x_{t-1}, x_{t-2},...

B. The variables ξ_t have zero mean, $E[\xi_t] = 0$; unit variance, $E[\xi_t^2] = 1$; and are mutually uncorrelated, $E[\xi_s \xi_t] = 0$ for $s \neq t$.

C. We have $b_o > 0$, and $b_o^2 + b_1^2 + b_2^2 + \ldots \infty$.

D. The representation (5.21) is unique in the sense that $x_t(-\infty < t < \infty)$ allows no other decomposition of the type (5.21) with components which have the properties A, B, and C.

The predictive Decomposition Theorem, as expressed by equation (5.21), renders the time series x_t in terms of a stable prediction operator $b_o, b_1, b_2, \ldots$ operating on the present and past values $\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots$ of a realization of a mutually uncorrelated process. That is, the value of x_t is expressed in terms of the present value ξ_t and past values $\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots$ but no future values $\xi_{t+1}, \xi_{t+2}, \ldots$ In other words, equation (5.21) represents a "predictive" decomposition of the time series, x_t.

Whereas Wold (1938) explicitly found the $b_o, b_1, b_2, \ldots$ only for the auto-regressive process and the process of finite moving averages which have rational spectra, Kolmogorov (1939) gave the method of Section 5.1 which explicitly yields the $b_o, b_1, b_2, \ldots$ for processes with arbitrary power spectra $\Phi(\omega)$.

In addition Kolmogorov (1941) shows that the decomposition (5.21) is unique (Property D of our statement of the Decomposition Theorem). That is, there is only one sequence of constants $b_0, b_1, b_2, \ldots$ and only one sequence of random variables ξ_t satisfying the conditions of the Theorem.

Let us now consider the predictive Decomposition Theorem in the language of the engineer (Bode and Shannon, 1950). The non-random or deterministic elements of a stochastic process may be represented by a physically realizable and stable elctric or mechanical network or filter. This filter has minimum phase characteristic. The time function $b_t (t = 0,1,2,3,\ldots)$ is equal to the output obtained from the filter in response to a unit impulse impressed upon the filter at time $t = 0$. That is, the linear operator b_t is the impulsive response of the filter, and we shall call it the response function of the stochastic process.

The random or non-deterministic elements of the stochastic process are represented by the ξ_t $(-\infty < t < \infty)$, which may be considered to be the mutually uncorrelated impulses of wide-band resistance noise or "white" noise. The time series x_t $(-\infty < t < \infty)$ is the response of the filter to the white noise input $\xi_s (-\infty < s \leq t)$. That is $\xi_s (-\infty < s \leq t)$ may be regarded as an impulse of strength ξ_s, which will produce a

response $\xi_s b_{t-s}$ at the subsequent time t. By adding the contributions of all the impulses $\xi_s (-\infty < s < t)$, we obtain the total response, which is the time series x_t:

$$x_t = \sum_{s=t}^{-\infty} \xi_s b_{t-s}. \tag{5.22}$$

Letting t-s = n, we have

$$x_t = \sum_{n=o}^{\infty} b_n \xi_{t-n} = b_o \xi_t + b_1 \xi_{t-1} + b_2 \xi_{t-2} + \ldots \tag{5.21}$$

which is the predictive decomposition (5.21).

Since the impulsive response of the filter is given by $b_t (t \geq 0)$ with $b_t = 0$ for $t < 0$, its transfer function is the Fourier transform

$$B(\omega) = \sum_{s=o}^{\infty} b_s e^{-i\omega s} \tag{5.23}$$

which has gain

$$|B(\omega)| = \sqrt{\Phi(\omega)} \tag{5.231}$$

and has minimum phase characteristic, given by equation (5.161). The power spectrum of the ξ_t, which is equal to 1 $(-\pi < \omega \leq \pi)$, multiplied by the power transfer function of the filter, which

is $|B(\omega)|^2$, yields the power spectrum $\Phi(\omega)$ of the time series x_t. The transfer function $B(\omega)$ is the Wold-Kolmogorov factor of the power spectrum $\Phi(\omega)$.

We see that the Predictive Decomposition Theorem states that any stationary time series (with an absolutely continuous spectral distribution) can be considered to be composed of many overlapping pulses, or wavelets, or responses, all with the same shape b_n, where $b_n = 0$ for $n < 0$. The arrival times of these wavelets, and their relative weighting, is given by the impulses ξ_{t-n}. The response function b_n, which is the shape of these wavelets, reflects the dynamics of the process, whereas the mutually uncorrelated impulses ξ_{s-t} reflects the statistical character of the process.

The wavelet shape b_n is physically stable, that is, it is the solution of a stable difference equation. The autocorrelation function of the stochastic process is

$$\phi(\tau) = E[x_t x_{t+\tau}] = E\Big[\sum_{s=0}^{\infty} b_s \xi_{t-s} \sum_{r=0}^{\infty} b_r \xi_{t+\tau-r}\Big]$$

$$= \sum_{s=0}^{\infty} b_s b_{s+\tau}. \qquad (5.232)$$

Thus the autocorrelation function of the wavelet b_s (where $b_s = 0$ for $s < 0$) is the same as the autocorrelation function

of the time series x_t. As a result the energy spectrum of the wavelet (see Section 3.3) is the same function as the power spectrum of the time series x_t, a fact of which we made use in Section 5.1 where we determined the shape of the wavelet b_n from the power spectrum $\Phi(\omega)$ of the time series x_t.

Since the filter is realizable and stable with minimum phase, there exists the inverse filter which is also realizable and stable with minimum phase. Let the response function of this inverse filter be $a_t (t \geq 0)$ with $a_t = 0$ for $t < 0$, so that its transfer function is

$$A(\omega) = B^{-1}(\omega) = \sum_{s=o}^{\infty} a_s e^{-i\omega s} = \frac{1}{B(\omega)} = \frac{1}{\sum_{s=o}^{\infty} b_s e^{-i\omega s}} , \qquad (5.24)$$

so that

$$B^{-1}(\omega)\, B(\omega) = \sum_{s=o}^{\infty} a_s e^{-i\omega s} \sum_{s=o}^{\infty} b_s e^{-i\omega s} = 1. \qquad (5.241)$$

The relationships of the inverse filter, $A(\omega)$ with the response a_t, to the filter $B(\omega)$ with the response b_t, both of which are realizable and stable, are given in equations (2.78) through (2.787) of Section 2.7. In Section 2.7 the response functions a_t and b_t are referred to as linear operator coefficients.

Accordingly, the white noise ξ_t $(-\infty < t < \infty)$ is the total response at t of the inverse filter to the input x_s $(-\infty < s \leq t)$. That is, the $x_s(-\infty < s \leq t)$ may be regarded as an impulse of strength x_s, which will produce a response $x_s a_{t-s}$ at the subsequent time t. Adding all these contributions we have the total response

$$\xi_t = \sum_{s=t}^{-\infty} x_s a_{t-s} \tag{5.25}$$

which, by letting n = t-s, is

$$\xi_t = \sum_{n=o}^{\infty} a_n x_{t-n} = a_o x_t + a_1 x_{t-1} + a_2 x_{t-2} + \cdots \tag{5.26}$$

Since the present value x_t and the past values $x_{t-1}, x_{t-2}, \cdots$ yield the value of ξ_t, we see that knowledge of x_t up to the time t is equivalent to knowledge of ξ_t up to time t. The representation (5.26), which is the predictive decomposition of the impulse ξ_t, will be called the inverse predictive decomposition of the time series x_t.

5.3 Prediction of Stationary Time Series

The value $x_{t+\alpha}$ of the time series at time $t+\alpha$, because of the predictive decomposition (5.21), is given by

$$x_{t+\alpha} = \sum_{s=o}^{\infty} b_s \xi_{t+\alpha-s}$$

$$= (b_o \xi_{t+\alpha} + b_1 \xi_{t+\alpha-1} + \ldots + b_{\alpha-1} \xi_{t+1}) + (b_\alpha \xi_t + b_{\alpha+1} \xi_{t-1} + \ldots). \tag{5.31}$$

Let us now consider time t to be the present time with respect to the filter; that is, time t is the time at which the computations are to be carried out. As a result, all values of the time series $x_s (-\infty < s \leq t)$ at and prior to time t are known at time t, and consequently all values of the white noise ξ_s $(-\infty < s \leq t)$ at and prior to time t may be found by means of the inverse filter, represented by

$$\xi_s = \sum_{n=o}^{\infty} a_n x_{s-n} \text{ for } -\infty < s \leq t. \tag{5.26}$$

Thus the component

$$(b_\alpha \xi_t + b_{\alpha+1} \xi_{t-1} + b_{\alpha+2} \xi_{t-2} + \ldots) \tag{5.311}$$

of the value of $x_{t+\alpha}$ given by equation (5.31) may be computed at time t, since the values $\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots$ are available at time t. On the other hand, the component

$$(b_o \xi_{t+\alpha} + b_1 \xi_{t+\alpha-1} + \ldots + b_{\alpha-1} \xi_{t+1}) \qquad (5.312)$$

of $x_{t+\alpha}$ given by the predictive decomposition (5.31) can not be computed at time t, since the values $\xi_{t+1}, \xi_{t+2}, \ldots \xi_{t+\alpha}$ are not available at time t. In other words, the component (5.311) is the predictable component of $x_{t+\alpha}$ at time t, and the component (5.312) is the unpredictable component of $x_{t+\alpha}$ at time t. That is, the predictable part of $x_{t+\alpha}$ is made up of the response due to the impulses $\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots$ which have occured at and prior to time t, and the unpredictable part is made up of the impulses $\xi_{t+1}, \xi_{t+2}, \ldots, \xi_{t+\alpha}$ which occur between the present time t and the time $t + \alpha$.

Thus Wold (1938) gives the following solution of the prediction problem for stationary time series with absolutely continuous spectral distribution functions. Since this solution follows directly from the Predictive Decomposition Theorem (5.21) which utilizes the operator b_t, it was not until Kolmogorov (1939) gave the method of section 5.1 which explicitly yields the b_t that the general solution of the prediction problem was fully established.

Since the impulses ξ_t are uncorrelated, Wold (1938) by utilizing the Gram-Schmidt process of orthogonalizing random vectors shows that the best linear forecast in the sense of principle of the least squares is yielded by

$$\hat{x}_{t+\alpha} = b_\alpha \xi_t + b_{\alpha+1} \xi_{t-1} + b_{\alpha+2} \xi_{t-2} + \ldots; \quad \alpha = 1,2,\ldots \tag{5.32}$$

The forecast (5.32) is the predictable component (5.311).

For the forecast (5.32) the error is given by the unpredictable component (5.312), and Wold (1938) shows that the mean square error, given by the expectation

$$I_{min} = E[x_{t+\alpha} - \hat{x}_{t+\alpha}]^2 = E[b_o \xi_{t+\alpha} + b_1 \xi_{t+\alpha-1} + \ldots + b_{\alpha-1} \xi_{t+1}]^2, \tag{5.321}$$

is a minimum. Since $E(\xi_t \xi_s) = 0$ for $t \neq s$, this minimum value is

$$I_{min} = E[x_{t+\alpha} - \hat{x}_{t+\alpha}]^2 = (b_o^2 + b_1^2 + \ldots + b_{\alpha-1}^2) E[\xi_t^2]. \tag{5.322}$$

By letting $E[\xi_t^2] = 1$, we have

$$I_{min} = \sum_{n=o}^{\alpha-1} b_n^2, \tag{5.323}$$

thereby showing that the efficiency of the forecast decreases as the prediction distance α increases. Since

$$E[x_t^2] = E[(\sum_{s=o}^{\infty} b_s \xi_{t-s})^2] = E[\xi_t^2] \sum_{t=o}^{\infty} b_t^2 , \qquad (5.324)$$

the prediction error I_{min} tends toward $E[x_t^2]$ as the prediction distance α tends toward infinity and hence for large values of the prediction distance α the trivial forecast of $\hat{x}_{t+\alpha} = 0$ has about the same efficiency as the forecast(5.32).

Kolmogorov (1939) generalizes Wold's result (5.323) by showing that the minimum mean square prediction error for a process with a non-absolutely continuous spectral distribution is given by

$$I_{min} = \sum_{n=o}^{\alpha-1} b_n^2 . \qquad (5.325)$$

Let us use the Wold-Kolmogorov normalization, which is that $b_o = 1$. Then $B(\omega)$ is given by equation (5.19) and I_{min} becomes

$$I_{min} = \exp [\frac{1}{\pi} \int_o^{\pi} \log \Phi(\omega) \, d\omega] \, (1 + b_1^2 + \ldots + b_{\alpha-1}^2). \qquad (5.326)$$

Kolmogorov states that

$$\exp [\frac{1}{\pi} \int_o^{\pi} \log \Phi (\omega) \, d\omega] = 0 \qquad (5.327)$$

and consequently $I_{min} = 0$ if $\Phi(\omega) = 0$ on a set of positive measure and also if the integral

$$\int_0^{\pi} \log \Phi(\omega)\, d\omega \qquad (5.328)$$

diverges, referring to this situation as the singular case of Wold (1938). Thus we see the reasons for the restrictions (5.131) and (5.133) in Section 5.1.

Doob (1953, p 584) points out that in the transformation of discrete time series, with the spectrum $\Phi(\omega)$ to continuous time series, with the spectrum $W(\omega)$, the integrals

$$\int_{-\pi}^{\pi} \log \Phi(\omega)\, d\omega, \quad \int_{-\infty}^{\infty} \frac{\log W(\omega)}{1+\omega^2}\, d\omega \qquad (5.329)$$

are finite and infinite together. The condition that the second integral be finite may be referred to as the Paley-Wiener criterion (Paley and Wiener, 1934), and Wiener (1942) uses this condition in the same connection as Kolmogorov (1939).

Let us now summarize the solution of the prediction problem for stationary time series with absolutely continuous spectral distribution functions as given by Wold (1938).

In order to obtain the predicted values $\hat{x}_{t+\alpha}$ from the values $x_t, x_{t-1}, x_{t-2}, \ldots$ we first apply the linear operator

$$\xi_s = \sum_{n=0}^{\infty} a_\eta x_{s-n} \text{ for } -\infty < s \leqslant t \qquad (5.26)$$

to yield the values ..., ξ_{t-2}, ξ_{t-1}, ξ_t, and then apply the operator

$$\hat{x}_{t+\alpha} = \sum_{s=0}^{\infty} b_{\alpha+s} \xi_{t-s} \tag{5.32}$$

to yield the predicted value $\hat{x}_{t+\alpha}$. The operations on the past values x_t, x_{t-1}, x_{t-2},..., represented by equations (5.26) and (5.32) may be combined into

$$\hat{x}_{t+\alpha} = \sum_{s=0}^{\infty} b_{\alpha+s} \xi_{t-s} = \sum_{s=0}^{\infty} b_{\alpha+s} \sum_{n=0}^{\infty} a_n x_{t-s-n} \tag{5.33}$$

which becomes, by letting $r = s+n$, and recalling that $a_t = 0$ for $t < 0$,

$$\hat{x}_{t+\alpha} = \sum_{s=0}^{\infty} b_{\alpha+s} \sum_{r=0}^{\infty} a_{r-s} x_{t-r} = \sum_{r=0}^{\infty} \left(\sum_{s=0}^{\infty} b_{\alpha=s} a_{r-s} \right) x_{t-r}. \tag{5.331}$$

Let us define

$$k_r(\alpha) = \sum_{s=0}^{\infty} b_{\alpha+s} a_{r-s} \tag{5.332}$$

so that equation (5.331) becomes

$$\hat{x}_{t+\alpha} = \sum_{r=o}^{\infty} k_r(\alpha)\, x_{t-r}. \qquad (5.333)$$

This equation thereby expresses the predicted value $\hat{x}_{t+\alpha}$ in terms of the present value x_t and past values x_{t-1}, x_{t-2}, x_{t-3}, ..., and has the same form as the pure prediction operator (2.21) Chapter II except that we now allow the number of operator coefficients k_s to becomes infinite. Also the dependence of the operator coefficients k_s on the value α of the prediction distance, is indicated. We note that $k_s(\alpha) = 0$ for s less than zero, so that the operator is realizable. Also $k_s(\alpha)$ is stable since the operators a_t and b_t are stable.

Further, Wold (1938) shows that

$$\hat{x}_{t+\alpha} = - a_1\hat{x}_{t+\alpha-1} - a_2\hat{x}_{t+\alpha-2} - \cdots - a_{\alpha-1}\hat{x}_{t+1} - a_\alpha x_t - a_{\alpha+1}x_{t-1} - a_{\alpha+2}x_{t-2} - \cdots \qquad (5.34)$$

and he also shows that the operator coefficients $k_r(\alpha)$ satisfy

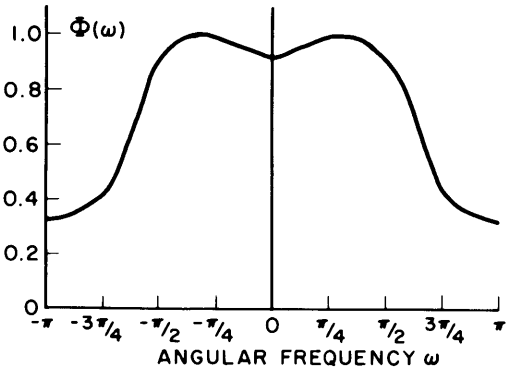
$$k_o(\alpha) + a_1k_o(\alpha-1) + a_2k_o(\alpha-2) + \cdots + a_{\alpha-1}k_o(1) + a_\alpha = 0$$

$$k_1(\alpha) + a_1k_1(\alpha-1) + a_2k_1(\alpha-2) + \cdots + a_{\alpha-1}k_1(1) + a_{\alpha+1} = 0 \qquad (5.35)$$

Figure 5 illustrates the least squares linear operator for unit prediction distance. That is, the operator coefficients are determined by the condition that the mean square prediction error be a minimum. The operator is in the form of equation (2.28) From the expression for $\Phi(\omega)$ in Figure 5, we see that time series is an autoregressive time series (see Section 5.5-B) and thus the solution of the simultaneous equations given in Figure 5 may be found by the method of the factorization of the autoregressive spectrum given in Section 2.8.

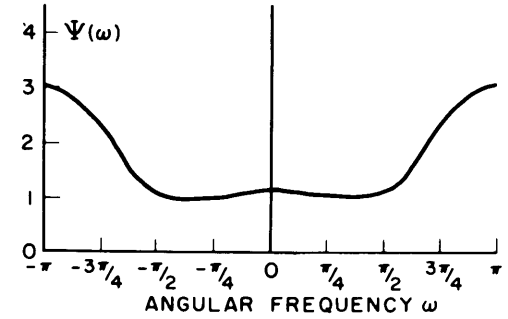
In the remaining part of this section, we wish to compare the solution of the prediction problem for discrete time series with absolutely continuous spectral distribution functions as given jointly by Wold (1938) and Kolmogorov (1939), with the solution given by Wiener (1948) in his book Cybernetics. Since Wiener works chiefly with continuous time series, and then translates his final results to the discrete case, we shall translate his entire solution to the discrete case. In the main, this translation consists of rewriting stochastic integrals of the Stieltjes-Lebesque type by the corresponding discrete summations (i.e. rewriting stochastic integral equations as stochastic difference equations) and converting all equations to the same notation and conventions which we have used up to now. Stochastic integrals are discussed by Doob (1953), who points

GIVEN THE POWER SPECTRUM $\Phi(\omega)$ OF THE TIME SERIES x_t $(-\infty < t < \infty)$ WITH $E[x_t] = 0$, WE WISH TO FIND THE LINEAR OPERATOR WITH OPERATOR COEFFICIENTS a_0, a_1, a_2, a_3, a_4 SUCH THAT THE PREDICTION ERROR VARIANCE $E\left[\xi_t^2\right]$ IS A MINIMUM WHERE THE PREDICTION ERROR IS GIVEN BY $\xi_t = a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + a_3 x_{t-3} + a_4 x_{t-4}$



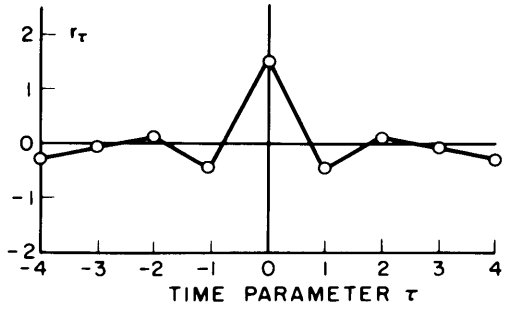
POWER SPECTRUM $\Phi(\omega)$

$$\Phi(\omega) = \frac{1}{r_0 + 2\sum_{\tau=1}^{4} r_\tau \cos \omega\tau}$$



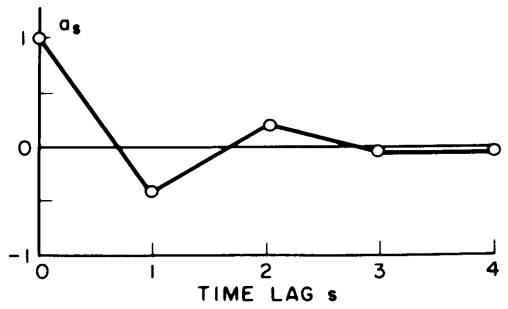
POWER TRANSFER FUNCTION $\Psi(\omega) = \frac{1}{\Phi(\omega)}$

$$\Psi(\omega) = |A(\omega)|^2 = r_0 + 2\sum_{\tau=1}^{4} r_\tau \cos \omega\tau = \sum_{\tau=-4}^{4} r_\tau e^{-i\omega\tau}$$



FOURIER TRANSFORM OF POWER TRANSFER FUNCTION

$$r_\tau = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Psi(\omega) e^{i\omega\tau}\, d\omega$$



STABLE OPERATOR COEFFICIENTS a_s

DETERMINED FROM SOLUTION OF THE SIMULTANEOUS EQUATIONS

$r_0 = a_0^2 + a_1^2 + a_2^2 + a_3^2 + a_4^2$

$r_1 = a_0a_1 + a_1a_2 + a_2a_3 + a_3a_4$

$r_2 = a_0a_2 + a_1a_3 + a_2a_4$

$r_3 = a_0a_3 + a_1a_4$

$r_4 = a_0a_4$

SUBJECT TO THE CONDITION THAT THE TRANSFER FUNCTION $A(\omega) = \sum_{t=0}^{4} a_t e^{-i\omega t}$ HAVE NO SINGULARITIES OR ZEROS IN THE LOWER HALF λ PLANE, WHERE $\lambda = \omega + i\sigma$

FIGURE 5 THE LEAST SQUARES LINEAR OPERATOR

out that they were introduced by Wiener (1923). In regard to the conventions used, Wiener (1942, 1948) in his various solutions of the prediction and filtering problems at times interchanges with respect to our convention past with future on the various time scales, that is, left with right in the basic ergodic transformation, and also interchanges the interior and exterior of the unit circle. For a further discussion of this point, see Hammerle (1951). So that the reader may follow our mathematical argument in relation to that of Wiener (1948), we shall give the number of his corresponding equation in parentheses on the left-hand side of our equation. Also corresponding equation numbers of Wiener (1942) are written in brackets on the left hand side of our equation.

Wiener considers the ensemble of time series

$$(3.34),(3.940) \quad \sum_{-\infty}^{\infty} c_{t+\tau}\ \xi_\tau = x_t \qquad (4.621)$$

generated by a process of moving summation (Section 4.6) so that the process is ergodic and has an absolutely continuous spectral distribution function. The autocorrelation function is

$$(3.35) \quad \phi(\tau) = \sum_{s=-\infty}^{\infty} c_s c_{s+\tau} = \sum_{s=-\infty}^{\infty} c_{s+t}\ c_{s+t+\tau}. \qquad (4.642)$$

The infinite smoothing operator coefficients c_s $(-\infty < s < \infty)$ are real, and the only significant quantity connected with them is $\phi(\tau)$. Thus one wishes to replace them by a physically realizable and stable operator b_t where $b_t = 0$ for $t < 0$, for which

[2.038]
$$\sum_{s=0}^{\infty} b_s b_{s+\tau} = \phi(\tau) \qquad (5.232)$$

which is the only independent statistical parameter of the time series. We have called b_t the response function of the time series, that is, it is the shape of the wavelets which may be considered to comprise the time series.

Wiener's method of determining b_t is the same as that of Kolmogorov (1939), which we gave in Section 5.1. Thus Wiener puts

(3.925)
[1.167]
$$\phi(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi(\omega)\, e^{i\tau\omega}\, d\omega \qquad (4.46)$$

or

(3.926)
[2.605]
$$\Phi(\omega) = \sum_{-\infty}^{\infty} \phi(\tau)\, e^{-i\tau\omega}. \qquad (4.43)$$

He lets

(3.927)
$$\frac{1}{2} \log \Phi(\omega) = \sum_{-\infty}^{\infty} \alpha_t \cos t\omega \qquad (5.134)$$

and

$$G(\omega) = \frac{1}{2} g(\omega) = \frac{1}{2} \log B(\omega) = \frac{\alpha_o}{2} + \sum_1^\infty \alpha_t e^{-i\omega t} \tag{3.928}$$

(5.171)

which yields

$$e^{g(\omega)} = B(\omega) \tag{3.929}$$

[2.620] (5.172)

which is

$$\exp(\alpha_o + 2 \sum_1^\infty \alpha_t e^{-i\omega t}) = B(\omega) \tag{5.173}$$

where

$$B(\omega) = B(z=e^{-i\omega}) = \sum_{s=o}^\infty b_s e^{-i\omega s} \tag{5.12}$$

[2.625]

is the boundary value on the unit circle $|z| = 1$ of a function $B(z)$ without zeros or singularities inside the unit circle, where ω is the angle. (See Figure 3). Thus, we have

$$|B(\omega)|^2 = \overline{\Phi}(\omega). \tag{3.930}$$

(5.181)

The method of Wiener (1948) to determine the operator which

furnishes the optimum prediction for processes with absolutely continuous spectral distribution functions is the same as that of Wold (1938), which we gave in the first part of this Section. Thus Wiener also starts with the predictive decomposition of the time series

$$(3.75) \qquad x_t = \sum_{\tau=-t}^{\infty} b_{t+\tau}\xi_{-\tau} = b_o\xi_t + b_1\xi_{t-1} + b_2\xi_{t-2} + \cdots \qquad (5.21)$$

We note that in his equations, Wiener lets time -t for the ξ time series correspond to time +t for the x time series. That is, whereas Wold lets the past of the x time series correspond to the past of the ξ time series, Wiener for the ξ time series interchanges left with right, and accordingly lets the past of the x time series correspond to the future of the ξ time series. As we have stated, all the equations which we give are converted to Wold's convention, i.e. the convention which we have been using.

Wiener gives the inverse predictive decomposition

(3.77)

[Second equation below 2.038]

$$\xi_t = \sum_{\sigma=-t}^{\infty} a_{t+\sigma}x_{-\sigma} = a_o x_t + a_1 x_{t-1} + a_2 x_{t-2} + \cdots \qquad (5.25)$$

which has the formal property

(3.78) $$x_t = \sum_{\tau=-t}^{\infty} b_{t+\tau} \sum_{\sigma=\tau}^{\infty} a_{-\tau+\sigma} x_{-\sigma} .$$ (2.762), (2.77)

The transfer functions are $B(\omega)$ and $A(\omega)$ which have the representation

(3.80a)
(3.933) $$b_s = \frac{1}{2\pi} \int_{-\pi}^{\pi} B(\omega)\, e^{i\omega s}\, d\omega$$ (5.36)

and

(3.80b) $$a_s = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\omega)\, e^{i\omega s}\, d\omega$$ (5.361)

where

(3.82), (3.935) $$\frac{1}{B(\omega)} = A(\omega).$$ (5.24)

Thus the past and present of ξ_t determine the past and present of x_t and vice versa; the Predictive Decomposition Theorem.

For the prediction distance $\alpha > 0$, we have

(3.83) [First equation below 2.038]
$$x_{t+\alpha} = \sum_{\tau=-t-\alpha}^{\infty} b_{t+\alpha+\tau}\, \xi_{-\tau}$$
$$= \sum_{\tau=-t-\alpha}^{-t-1} b_{t+\alpha+\tau}\, \xi_{-\tau} + \sum_{\tau=-t}^{\infty} b_{t+\alpha+\tau}\, \xi_{-\tau}$$

$$= (b_o \xi_{t+\alpha} + b_1 \xi_{t+\alpha-1} + \ldots + b_{\alpha-1} \xi_{t+1}) + (b_\alpha \xi_t + b_{\alpha+1} \xi_{t-1} + \ldots).$$

(5.31)

Wiener states that the first term of the last expression depends on a range of ξ_s of which knowledge of x_σ for $\sigma < t$ tells us nothing, that is, it is the unpredictable component. The mean square value of this unpredictable component is

(3.84)
[2.645]
$$I_{min} = \sum_{\tau=-t-A}^{-t-1} b_{t+\alpha+\tau}^2 = \sum_{\tau=o}^{\alpha-1} b_\tau^2. \qquad (5.323)$$

The best prediction in the sense of least squares is the last term, the predictable component, $x_{t+\alpha}$, given by

(3.85)
$$\hat{x}_{t+\alpha} = \sum_{\tau=-t}^{\infty} b_{t+\alpha+\tau} \xi_{-\tau} = \sum_{\tau=-t}^{\infty} b_{t+\alpha+\tau} \sum_{\sigma=\tau} a_{-\tau+\sigma} x_{-\sigma}$$

(5.33)

The first term, the unpredictable component, is not necessarily independent of the second term, the predictable component, as Wiener (1948, p99) states, but we can only state (Wold, 1938) that the unpredictable component is uncorrelated with the predictable component. That is, the ξ_t in general are not independent, as Wiener (1942, p62) states, but in general are uncorrelated. Thus,

we infer from Wiener that the unpredictable component will have zero expectation for any fixed values of ξ_t, ξ_{t-1}, ξ_{t-2},..., whereas we can in general assert that the unpredictable component will be of zero expectation only when formed with respect to the joint distribution of ξ_t, ξ_{t-1}, ξ_{t-2},... .

The predictable component (5.311) may be written

$$\hat{x}_{t+\alpha} = \sum_{n=o}^{\infty} b_{n+\alpha} \xi_{t-n} = \sum_{n=o}^{\infty} b_{n+\alpha} \sum_{s=o}^{\infty} a_s x_{t-n-s} \qquad (5.33)$$

where $n = t+\tau$ and $s = t+\sigma - n = \sigma - \tau$. Let us define the prediction operator coefficients $k_s(\alpha)$ so that

(3.931)
$$\hat{x}_{t+\alpha} = \sum_{s=o}^{\infty} x_{t-s} k_s(\alpha). \qquad (5.333)$$

The transfer function $K_\alpha(\omega)$ may be found by applying the operator (5.333) to $x_t = e^{i\omega t}$, and then dividing by $e^{i\omega t}$, as in Section 2.4. Thus we have

(3.87)
$$\sum_{n=o}^{\infty} b_{n+\alpha} \sum_{s=o}^{\infty} a_s e^{i\omega(t-n-s)} = e^{i\omega t} K_\alpha(\omega) = e^{i\omega t} \sum_{s=o}^{\infty} k_s(\alpha) e^{-i\omega s} \qquad (5.37)$$

Let us put

(3.86
$$B_\alpha(\omega) = \sum_{t=o}^{\infty} b_{t+\alpha}\, e^{-i\omega t} . \qquad (5.371)$$

Then

$$K_\alpha(\omega) = \sum_{n=o}^{\infty} b_{n+\alpha} \sum_{s=o}^{\infty} a_s e^{-i\omega(n+s)}$$

$$= \sum_{n=o}^{\infty} b_{n+\alpha}\, e^{-i\omega n} \sum_{s=o}^{\infty} a_s e^{-i\omega s}$$

$$= B_\alpha(\omega)\, A(\omega). \qquad (5.372)$$

Therefore, because of equations (5.24) and (5.36), we have

(3.88),(3.932)
[2.630]
$$K_\alpha(\omega) = \frac{B_\alpha(\omega)}{B(\omega)}$$

$$= \frac{1}{B(\omega)} \sum_{n=o}^{\infty} b_{n+\alpha} e^{-i\omega n}$$

$$= \frac{1}{B(\omega)} \sum_{t=\alpha}^{\infty} b_t e^{-i\omega(t-\alpha)}$$

$$= \frac{1}{2\pi B(\omega)} \sum_{t=\alpha}^{\infty} e^{-i\omega(t-\alpha)} \int_{-\pi}^{\pi} B(u) e^{iut} du$$

$$= \frac{1}{2\pi B(\omega)} \sum_{s=o}^{\infty} e^{-i\omega s} \int_{-\pi}^{\pi} B(u) e^{iu(s+\alpha)}\, du. \qquad (5.373)$$

This equation may be written

$$(3.934) \quad K_\alpha(\omega) = \sum_{r=o}^{\infty} k_r(\alpha) e^{-i\omega r}$$

$$= \frac{B_\alpha(\omega)}{B(\omega)} = \frac{e^{i\omega\alpha} \sum_{t=\alpha}^{\infty} b_t e^{-i\omega t}}{\sum_{t=o}^{\infty} b_t e^{-i\omega t}}$$

$$= e^{i\omega\alpha} \left(1 - \frac{\sum_{o}^{\alpha-1} b_t e^{-i\omega t}}{\sum_{o}^{\infty} b_t e^{-i\omega t}} \right). \qquad (5.374)$$

Since

$$(3.935) \quad \frac{1}{B(\omega)} = \sum_{s=o}^{\infty} a_s e^{-i\omega s} \qquad (5.24)$$

we have

(3.936)

$$K_\alpha(\omega) = \sum_{r=o}^{\infty} k_r(\alpha)e^{-i\omega r} = e^{i\omega\alpha}(1 - \sum_{t=o}^{\alpha-1} b_t e^{-i\omega t} \sum_{s=o}^{\infty} a_s e^{-i\omega s}). \qquad (5.375)$$

For prediction distance $\alpha = 1$, we have

(3.937)
$$\sum_{o}^{\infty} k_r(1)\, e^{-i\omega r} = e^{i\omega}(1-b_o \sum_{s=o}^{\infty} a_s e^{-i\omega s}), \quad \alpha=1 \qquad (5.376)$$

so that

(3.938)
$$k_r(1) = -a_{r+1}\, b_o = -a_{r+1},\ b_o=1. \qquad (5.377)$$

Thus the prediction for one step ahead is

(3.939)
$$\hat{x}_{t+1} = - \sum_{r=o}^{\infty} a_{r+1}\, x_{t-r} \qquad (5.378)$$

and by step-by-step prediction, the prediction for any prediction distance α may be found. This statement is equivalent to equation (5.34) given by Wold (1938).

The prediction operator in the time domain is

[2.635]

$$k_t(\alpha) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega t} K_\alpha(\omega)\, d\omega. \qquad (5.38)$$

Utilizing equation (5.362) for $K_\alpha(\omega)$, and letting $r = s+n$, we have

$$k_t(\alpha) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega t} \sum_{n=0}^{\infty} a_n e^{-i\omega n} \sum_{s=0}^{\infty} b_{\alpha+s}\, e^{-i\omega s}\, d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega t} \sum_{s=0}^{\infty} b_{\alpha+s}\, e^{-i\omega s} \sum_{r=0}^{\infty} a_{r-s} e^{-i\omega(r-s)}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega t} \sum_{r=0}^{\infty} e^{-i\omega r} \left(\sum_{s=0}^{\infty} b_{\alpha+a}\, a_{r-s} \right) = \sum_{s=0}^{\infty} b_{\alpha+s} a_{t-s}$$

(5.381)

which is the Wold (1938) solution for $k_t(\alpha)$ given by equation (5.332).

As a final exercise let us take the Fourier transform of the equation (5.332), which gives the filter characteristics

$$K_\alpha(\omega) = \sum_{r=0}^{\infty} k_r(\alpha) e^{-i\omega r} = \sum_{r=0}^{\infty} e^{-i\omega r} \sum_{s=0}^{\infty} b_{\alpha+s} a_{r-s} \qquad (5.382)$$

which is

$$K_\alpha(\omega) = \sum_{s=o}^{\infty} b_{\alpha+s} e^{-i\omega s} \sum_{r=o}^{\infty} a_{r-s} e^{-i\omega(r-s)}. \qquad (5.383)$$

Recalling that $r = s+n$ and that $a_t = 0$ for $t < 0$, we have

$$K_\alpha(\omega) = \sum_{s=o}^{\infty} b_{\alpha+s} e^{-i\omega s} \sum_{n=o}^{\infty} a_n e^{-i\omega n}. \qquad (5.384)$$

We recall from Section 5.1 that the Kolmogorov (1939) factor of the power spectrum is

$$B(\omega) = \sum_{s=o}^{\infty} b_s e^{-i\omega s} = A^{-1}(\omega) \qquad (5.12), (5.24)$$

so therefore

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} B(\omega)\, e^{i\omega t}\, d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{s=o}^{\infty} b_s e^{-i\omega s}\, e^{i\omega t}\, d\omega = b_t \qquad (5.385)$$

since

$$\int_{-\pi}^{\pi} e^{-i\omega s}\, e^{i\omega t} = \begin{cases} 0 & s \neq t \\ 2\pi & s = t \end{cases} \qquad (5.386)$$

for integer s and t. Accordingly, by letting $\alpha + s = t$ in equation (5.385) we have

$$b_{\alpha+s} = \frac{1}{2\pi} \int_{-\pi}^{\pi} B(\omega)\, e^{i\omega(\alpha+s)}\, d\omega. \qquad (5.387)$$

Let us define

$$B_\alpha(\omega) = \sum_{s=o}^{\infty} b_{\alpha+s}\, e^{-i\omega s} \qquad (5.371)$$

which, by equation (5.387), is

$$B_\alpha(\omega) = \frac{1}{2\pi} \sum_{s=o}^{\infty} e^{-i\omega s} \int_{-\pi}^{\pi} B(\omega)\, e^{i\omega(\alpha+s)}\, d\omega. \qquad (5.388)$$

Also, equation (5.24) shows that

$$A(\omega) = \sum_{n=o}^{\infty} a_n e^{-i\omega n} = \frac{1}{B(\omega)} = B^{-1}(\omega) \qquad (5.24)$$

where $B(\omega)$ is the Kolmogorov (1939) factor of the power spectrum. Therefore the transfer function $K_\alpha(\omega)$, equation (5.384), becomes

$$K_\alpha(\omega) = \left(\sum_{n=o}^{\infty} a_n e^{-i\omega n}\right)\left(\sum_{s=o}^{\infty} b_{\alpha+s} e^{-i\omega s}\right)$$

$$= A(\omega)\, B_\alpha(\omega) = \frac{B_\alpha(\omega)}{B(\omega)} \qquad (5.389)$$

which by using equation (5.388) becomes our equation (5.373)

$$K_\alpha(\omega) = \frac{1}{2\pi B(\omega)} \sum_{s=0}^{\infty} e^{-i\omega s} \int_{-\pi}^{\pi} B(u)\, e^{iu(\alpha+s)}\, du \qquad (5.373)$$

which is equation (2.630) of Wiener (1942), (where the last should be $\alpha + \nu$), and is equation (3.932) of Wiener (1948).

Equation (5.389) for the transfer function for the optimum prediction operator may be interpreted (Bode and Shannon, 1950) as first passing the time series $x_s(-\infty < s \leq t)$ through the inverse filter $A(\omega)$, which yields the impulses $\xi_s(-\infty < s \leq t)$, and then passing the impulses through the transfer function $B_\alpha(\omega)$ which yields $\hat{x}_{t+\alpha}$.

In conclusion, then, Wold (1938, pp 102-103) actually obtains a coefficient function $k_t(\alpha)$ (which in Wold's notation is $f_{\alpha,t}$), to be used directly in making an optimum prediction for a discrete stationary time series with an absolutely continuous spectral distribution function. Further, Wold develops an expression for the mean square error of prediction, given by equation (5.322). Wiener (1942, 1948) also obtains the coefficient function $k_t(\alpha)$ which in the notation of Wiener (1942) is K(t) and the expression for the mean square error of prediction. In addition, Wiener (1942, 1948) gives the solution of the prediction problem for the continuous stationary time series together with its analytic aspects and emphasizes the

applicability of these methods to electrical engineering problems.

5.4 The Filtering Problem

The solution of the filtering problem for stationary time series, continuous and discrete, with absolutely continuous spectral distribution functions is given by Wiener (1942, 1948). As in the case of the prediction problem the solution of the filtering problem is a direct consequence of the Predictive Decomposition Theorem. Again we shall translate the solution of Wiener (1948) for continuous time series to discrete time series, with his equation numbers on the left. The predictive decomposition of the time series x_t consisting of message plus noise is

(3.89)

$$x_t = m_t + n_t = \sum_{\tau=0}^{\infty} b_\tau \xi_{t-\tau}. \qquad (5.41)$$

Since the message has an absolutely continuous spectral distribution function it may be represented by the process of moving summation (see Section 4.6) given by

(3.90)

$$m_t = \sum_{\tau=-\infty}^{\infty} q_\tau \xi_{t-\tau} + \sum_{\tau=-\infty}^{\infty} r_\tau \gamma_{t-\tau} \qquad (5.411)$$

where the random variables ξ_t and γ_t are mutually uncorrelated, that is,

$$E[\xi_t^2] = 1, \quad E[\xi_t \xi_{t+k}] = 0 \text{ for } k \neq 0,$$

$$E[\xi_t \gamma_s] = 0, \quad E[\gamma_t^2] = 0, \quad E[\gamma_t \gamma_{t+k}] = 0 \text{ for } k \neq 0. \tag{5.412}$$

The predictable part of the message $m(t+\alpha)$, where α is the lead, is

(3.901)
$$\hat{m}_{t+\alpha} = \sum_{\tau=0}^{\infty} q_{\tau+\alpha} \xi_{t-\tau} = \sum_{\tau=\alpha}^{\infty} q_\tau \xi_{t-\tau-\alpha} \tag{5.413}$$

and the (minimum) mean square error of prediction is

(3.902)
$$I_{min} = \sum_{-\infty}^{\alpha} q_\tau^2 + \sum_{-\infty}^{\infty} r_\tau^2. \tag{5.414}$$

From equations (5.41) and (5.411) we see that noise n_t is given by

$$n_t = \sum_{\tau=0}^{\infty} b_\tau \xi_{t-\tau} - \sum_{-\infty}^{\infty} q_\tau \xi_{t-\tau} - \sum_{\tau=-\infty}^{\infty} r_\tau \gamma_{t-\tau}. \tag{5.415}$$

Thus the autocorrelation $\phi_{22}(t)$ of the noise is given by (3.903)

$$\phi_{22}(t) = E[n_{t+\tau} n_\tau]$$

$$= \sum_{\tau=0}^{\infty} (b_{|t|+\tau} - q_{|t|+\tau})(b_\tau - q_\tau)$$

$$+ \sum_{\tau=-|t|}^{-1} (b_{|t|+\tau} - q_{|t|+\tau})(-q_\tau)$$

$$+ \sum_{-\infty}^{-|t|-1} q_{|t|+\tau}\, q_\tau + \sum_{-\infty}^{\infty} r_{|t|+\tau}\, r_\tau$$

$$= \sum_{\tau=0}^{\infty} b_{|t|+\tau} b_\tau - \sum_{\tau=0}^{\infty} q_{|t|+\tau}\, b_\tau - \sum_{\tau=-|t|}^{\infty} q_\tau b_{|t|+\tau}$$

$$+ \sum_{\tau=-\infty}^{\infty} q_{|t|+\tau}\, q_\tau + \sum_{\tau=-\infty}^{\infty} r_{|t|+\tau}\, r_\tau . \qquad (5.42)$$

The autocorrelation $\phi_{11}(t)$ of the message is

(3.904)

$$\phi_{11}(t) = E[m_{|t|+\tau}\, m_\tau]$$

$$= \sum_{\tau=-\infty}^{\infty} q_{|t|+\tau}\, q_\tau + \sum_{\tau=-\infty}^{\infty} r_{|t|+\tau}\, r_\tau . \qquad (5.421)$$

The cross-correlation $\phi_{12}(t)$ of the message and noise is

(3.905) $$\phi_{12}(t) = E[m_{t+\tau}\, n_\tau]$$

$$= E[m_{t+\tau}(m_\tau + n_\tau) - m_{t+\tau}\, m_\tau]$$

$$= E[m_{t+\tau}(m_\tau + n_\tau)] - \phi_{11}(\tau)$$

$$= E[\sum_{\sigma=-1}^{\infty} b_{\sigma+t}\, \xi_{\tau-\sigma} \sum_{\sigma=-t}^{\infty} q_\sigma\, \xi_{\tau-\sigma}] - \phi_{11}(\tau)$$

$$= \sum_{\tau=-t}^{\infty} b_{t+\tau}\, q_\tau - \phi_{11}(\tau). \qquad (5.422)$$

Let us define

(3.907) $$B(\omega) = \sum_{s=o}^{\infty} b_s e^{-i\omega s}$$

$$Q(\omega) = \sum_{s=-\infty}^{\infty} q_s\, e^{-i\omega s}$$

$$R(\omega) = \sum_{s=-\infty}^{\infty} r_s e^{-i\omega s}\ . \qquad (5.423)$$

The Fourier transforms of the correlation functions $\phi_{22}(t)$, $\phi_{11}(t)$, and $\phi_{12}(t)$ are the power spectrum $\Phi_{22}(\omega)$ of the noise, the power spectrum $\Phi_{11}(\omega)$ of the message, and the cross-spectrum $\Phi_{12}(\omega)$ of the message and noise respectively, and are given by

(3.906) $\Phi_{22}(\omega) = B\bar{B} + Q\bar{Q} - Q\bar{B} - B\bar{Q} + R\bar{R}$

$$\Phi_{11}(\omega) = Q\bar{Q} + R\bar{R}$$

$$\Phi_{12}(\omega) = B\bar{Q} - Q\bar{Q} - R\bar{R} = B\bar{Q} - \Phi_{11}(\omega). \qquad (5.43)$$

Therefore, we have

(3.908) $\Phi_{11}(\omega) + \Phi_{12}(\omega) + \overline{\Phi_{12}(\omega)} + \Phi_{22}(\omega) = B\bar{B} = |B(\omega)|^2.$ (5.431)

In order to compute $B(\omega)$, we must have the sum of spectra given by the left hand side of this equation. Let us call this sum $\Phi(\omega)$, that is

$$\Phi(\omega) = \Phi_{11}(\omega) + 2\mathrm{Re}\,[\,\Phi_{12}(\omega)] + \Phi_{22}(\omega). \qquad (5.432)$$

We see that $\Phi(\omega)$ is the power spectrum of the time series x_t, equation (5.41), and thus $\Phi(\omega)$ may be computed directly from this time series. We factor $\Phi(\omega)$ into $B(\omega)\ \overline{B(\omega)}$ according to the method of Section 5.1. In addition, from equations (5.43) we have

$$B\bar{Q} = \Phi_{12}(\omega) + \Phi_{11}(\omega) \qquad (5.433)$$

so that

$$(3.909) \qquad Q\overline{B} = \overline{\Phi_{12}(\omega)} + \Phi_{11}(\omega) = \Phi_{11}(\omega) + \Phi_{21}(\omega) \qquad (5.434)$$

because

$$\Phi_{21}(\omega) = \overline{\Phi_{12}(\omega)}. \qquad (5.435)$$

Thus we have

$$(3.910) \qquad Q(\omega) = \frac{\Phi_{11}(\omega) + \Phi_{21}(\omega)}{\overline{B(\omega)}} \qquad (5.436)$$

and

$$q_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} Q(\omega)\, d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_{11}(\omega)+\Phi_{21}(\omega)}{\overline{B(\omega)}}. \qquad (5.437)$$

We let the inverse predictive decomposition of the time series $x_t = m_t + n_t$ be

$$\xi_t = \sum_{s=o}^{\infty} a_s x_{t-s} = \sum_{s=o}^{\infty} a_s \left(m_{t-s} + n_{t-s} \right) \qquad (5.44)$$

which gives the prediction errors ξ_t (for unit prediction distance) of the time series x_t. Thus equation (5.413), which gives the

predictable part of $m(t+\alpha)$, consists of reaveraging these prediction errors by the realizable linear operation

$$\hat{m}_{t+\alpha} = \sum_{\tau=0}^{\infty} q_{\tau+\alpha}\, \xi_{t-\tau} = \sum_{\tau=0}^{\infty} q_{\tau+\alpha} \sum_{s=0}^{\infty} a_s (m_{t-\tau-s} + n_{t-\tau-s}). \tag{5.441}$$

The transfer function $H_\alpha(\omega)$ of this linear operator is the totality of transfer ratios obtained by letting $m_t + n_t = e^{i\omega t}$ in equation (5.441) and dividing by $e^{i\omega t}$, that is

(3.913),

(3.941)

$$H_\alpha(\omega) = \sum_{\tau=0}^{\infty} q_{\tau+\alpha} \sum_{s=0}^{\infty} a_s e^{-i\omega(\tau+s)}$$

$$= \sum_{\tau=0}^{\infty} q_{\tau+\alpha}\, e^{-i\omega\tau} \sum_{s=0}^{\infty} a_s e^{-i\omega s}$$

$$= \frac{1}{B(\omega)} \sum_{n=\alpha}^{\infty} q_n e^{-i\omega(n-\alpha)}$$

$$= \frac{1}{2\pi B(\omega)} \sum_{t=\alpha}^{\infty} e^{-i\omega(t-\alpha)} \int_{-\pi}^{\pi} \frac{\overline{\Phi}_{11}(\omega) + \overline{\Phi}_{21}(\omega)}{\overline{B(\omega)}}\, e^{i\omega t} d\omega. \tag{5.45}$$

As Wiener (1948) points out, the equation for continuous time series (Wiener's equation 3.913) which corresponds to our equation (5.45) for discrete time series is the transfer function of what electricians know as a "wave filter". The quantity α is the lead of the filter, and may be positive or negative. When it is negative, $-\alpha$ is known as the lag. Wiener also points out that apparatus corresponding to this equation may be always constructed with as much accuracy as we like, and he refers to papers by Dr. Y. W. Lee.

The mean square filtering error (5.414) is

$$I_{min} = \sum_{-\infty}^{\alpha} q_\tau^2 + \sum_{-\infty}^{\infty} r_\tau^2 \qquad (5.414)$$

where the first term on the right depends on the lag $-\alpha$, whereas the second term does not. For infinite lag, that is, $\alpha = -\infty$, the error becomes

$$(3.914) \qquad I_{min}\,(\alpha=-\infty) = \sum_{\tau=-\infty}^{\infty} r_\tau^2 = \phi_{11}(0) - \sum_{\tau=-\infty}^{\infty} q_\tau^2 \qquad (5.46)$$

because of equation (5.421), with $t = 0$. Equation (5.46) becomes

(3.914)

$$I_{min}\ (\alpha=-\infty) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\begin{vmatrix} \Phi_{11}(\omega) & \Phi_{12}(\omega) \\ \Phi_{21} & \Phi_{22}(\omega) \end{vmatrix}}{\Phi_{11}(\omega) + \Phi_{12}(\omega) + \Phi_{21}(\omega) + \Phi_{22}(\omega)} d\omega$$

(5.461)

which is

$$I_{min}(\alpha=-\infty)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\text{Determinant of Coherency Matrix of Message and of Noise}}{\text{Power Spectrum of Message plus Noise}} d\omega.$$

(5.462)

Thus if message and noise were completely coherent, then $I_{min}\ (\alpha=-\infty) = 0$. The part of I_{min} depending on lag is

(3.915)

$$\sum_{\tau=-\infty}^{\alpha} q_\tau^2 = \sum_{\tau=-\infty}^{\alpha} \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_{11}(\omega) + \Phi_{21}(\omega)}{B(\omega)} e^{i\omega t} d\omega \right|^2 .$$

(5.47)

In conclusion, then, we see that the general solution of the filtering problem for the time series $x_t = m_t + n_t$ consist of re-averaging the prediction errors ξ_t (for unit prediction distance)

of the time series x_t by means of the operator

$$\hat{m}_{t+\alpha} = \sum_{\tau=0}^{\infty} q_{\tau+\alpha} \; \xi_{t-\tau}. \tag{5.441}$$

Here the operator coefficients q_t are determined from $\overline{\Phi}_{11}(\omega)$ which is the power spectrum of the message m_t, from $\overline{\Phi}_{21}(\omega)$ which is the cross-spectrum of the noise n_t and the message m_t, and from $\overline{B(\omega)}$ which is the Wold-Kolmogorov factor of the power spectrum $\overline{\Phi}(\omega)$ of the time series $x_t = m_t + n_t$.

5.5 Time Series with Rational Power Spectra

Of particular interest to the working statistician are those time series with power spectra $\overline{\Phi}(\omega)$ which are rational functions in $z = e^{-i\omega}$ (see Figure 3).

A. Process of Finite Moving Averages

Let the stable response function b_t be of finite extent, that is

$$b_t = 0 \text{ for } t < 0 \text{ and } t > M,$$

$$b_0 \neq 0, \; b_M \neq 0. \tag{5.51}$$

Then the predictive decomposition of the time series x_t, with this response function, is

$$x_t = b_0 \, \xi_t + b_1 \, \xi_{t-1} + \cdots + b_M \, \xi_{t-M} \tag{5.511}$$

which represents a stationary process with an absolutely continuous spectral distribution.

Strictly speaking, the process x_t given by this decomposition is defined to be a process of finite moving averages if white noise ξ_t $(-\infty < t < \infty)$ is a purely random process, that is, if all of the components ξ_t are independent random variables with the same distribution function. Nevertheless, we shall also include in this definition, those processes (5.511) for which the $\xi_t(-\infty < t < \infty)$ represents a mutually uncorrelated process. See Section 4.5. We shall let ξ_t have mean value zero and unit variance. Thus, the $\xi_t(-\infty \quad t \quad \infty)$ is a process such that

$$E(\xi_t) = 0$$

$$E(\xi_t^2) = 1$$

$$E(\xi_t \xi_s) = 0 \quad s \neq t \,. \tag{5.512}$$

The autocorrelation is given by

$$\phi(\tau) = E[x_t \; x_{t+\tau}] = \sum_{t=o}^{M-\tau} b_t b_{t+\tau} \tag{5.513}$$

so that

$$\phi(\tau) = 0 \text{ for } |\tau| > M. \tag{5.514}$$

The power spectrum of the process of finite moving averages is then

$$\overline{\Phi}(\omega) = \phi(0) + 2 \sum_{\tau=1}^{M} \phi(\tau) \cos \omega\tau = \sum_{\tau=-M}^{M} \phi(\tau)\, e^{-i\omega\tau} \qquad (5.515)$$

which is a rational function in $z = e^{-i\omega}$. Let $B(\omega)$ be the transfer function

$$B(\omega) = \sum_{s=0}^{M} b_s e^{-i\omega s} \qquad (5.516)$$

so we see that

$$\overline{\Phi}(\omega) = B(\omega)\, \overline{B(\omega)} = |B(\omega)|^2 = \sum_{\tau=-M}^{M} \phi(\tau) e^{-i\omega\tau} \geq 0 \qquad (5.517)$$

is non-negative.

In general, an arbitrary set of coefficients

$$\phi(0), \phi(1), \phi(2), \ldots, \phi(M) \qquad (5.52)$$

will not be such that the rational function

$$\phi(0) + 2 \sum_{\tau=1}^{M} \phi(\tau) \cos \omega\tau = \sum_{\tau=-M}^{M} \phi(\tau)\, e^{-i\omega\tau} \qquad (5.521)$$

is non-negative, and hence it is not an acceptable function

to represent the power spectrum of a process of moving averages.

Thus Wold (1938) gives the following theorem, his Theorem 12.

Theorem 12. Let Q(u) be the polynomial of order M obtained from

$$\phi(0) + \sum_{\tau=1}^{M} \phi(\tau) \left[z^{-\tau} + z^{\tau}\right]$$

by the substitutuion $z^{-1} + z = u$. A necessary and sufficient condition that $\phi(0), \phi(1), \phi(2), \ldots \phi(M)$ be the correlogram of a process of moving averages (5.511) is that the polynomial should have no real root of odd multiplicity in the interval $-2 < u < 2$.

Thus for a sequence (5.52) which does fulfill the conditions of Wold's Theorem 12, the function

$$\phi(0) + 2 \sum_{\tau=1}^{M} \phi(\tau) \cos \omega\tau \geq 0 \qquad (5.522)$$

is non-negative and hence may represent the power spectrum $\underline{\Phi}(\omega)$ of a process of moving averages. In order to determine the response function $b_t (t = 0,1,2,\ldots M)$ of the process with this power spectrum (5.522), it is necessary to factor this power spectrum into

$$\Phi(\omega) = \phi(0) + 2 \sum_{\tau=1}^{M} \phi(\tau) \cos \omega\tau = B(\omega)\, \overline{B(\omega)} \qquad (5.523)$$

where

$$B(\omega) = \sum_{s=0}^{M} b_s e^{-i\omega s} \qquad (5.524)$$

is free from singularities and zeros in the lower half λ plane. Then the stable operator b_t is given by

$$b_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} B(\omega)\, e^{i\omega t}\, d\omega, \quad t = 0,1,2,\ldots, M. \qquad (5.525)$$

In order to carry out the factorization (5.523), Wold (1938) gave the method which we presented in Section 2.8. That is, we let $\Phi(\omega)$, given by equation (5.523) be the $\Psi(\omega)$ of Section 2.8, given by equation (2.82). Then we determine the stable finite linear operator b_t from $\Phi(\omega)$ in the same manner as given in Section 2.8. The inverse predictive decomposition is given by

$$\xi_t = a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \ldots = \sum_{s=0}^{\infty} a_s x_{t-s} \qquad (5.526)$$

where the inverse linear operator a_t (see Section 2.7) is realizable, stable, and infinite in extent.

The process of finite moving averages was introduced by Yule (1921, 1926) and Slutsky (1927), and consequently was the first stochastic process studied which was neither a purely random or an uncorrelated process (Section 4.5) nor a deterministic strictly periodic process (Section 3.2).

The solution of the prediction problem for such processes for the special case in which

$$b_t = \frac{1}{\sqrt{M+1}} \tag{5.527}$$

is given by Kosulajeff (1941).

B. The Autoregressive Process

The autoregressive process is a stochastic process for which the response function b_t $(t = 0,1,2,\ldots)$ is of infinite extent, but the inverse response function a_t $(t = 0,1,2,\ldots m)$ is of finite extent. Thus the inverse predictive decomposition of an autoregressive time series x_t of the m-th order is

$$\xi_t = a_o x_t + a_1 x_{t-1} + \ldots + a_m x_{t-m} = \sum_{s=o}^{m} a_s x_{t-s} \tag{5.53}$$

and the predictive decomposition is

$$x_t = b_o \xi_t + b_1 \xi_{t-1} + b_2 \xi_{t-2} + \ldots = \sum_{s=o}^{\infty} b_s \xi_{t-s} \tag{5.531}$$

where the operators a_t and b_t are stable and inverse to each other. See Section 2.7. Such a process is stationary with an absolutely continuous spectral distribution.

Strictly speaking, in the definition of the autoregressive process, the impulses ξ_t which are independent with the same distribution function, but we shall also admit ξ_t which are mutually uncorrelated. We shall let ξ_t be a process such that

$$E(\xi_t) = 0$$

$$E(\xi_t^2) = 1$$

$$E(\xi_t \xi_s) = 0 \quad s \neq t. \tag{5.532}$$

The autoregressive process, or the process of disturbed harmonics, was introduced by Yule (1927), and was a major step forward toward the establishment of the general theory of stochastic processes by Kolmogorov (1933) and Khintchine (1933). Because equation (5.53) for the prediction error ξ_t (for unit prediction distance) has only a finite number of operator coefficients, a_0, a_1, a_2,...a_m, the prediction operator for any prediction distance will require only a finite number of coefficients, as seen by equation (5.34).

The autocorrelation function is given by

$$\phi(\tau) = E\,[x_t x_{t+\tau}] = \sum_{\tau=0}^{\infty} b_t b_{t+\tau} \tag{5.533}$$

and the power spectrum by

$$\Phi(\omega) = \sum_{\tau=-\infty}^{\infty} \phi(\tau) e^{-i\omega\tau} = |B(\omega)|^2 = \frac{1}{|A(\omega)|^2} \geq 0 \tag{5.534}$$

where

$$B(\omega) = \sum_{s=0}^{\infty} b_s e^{-i\omega s} \tag{5.535}$$

and

$$A(\omega) = B^{-1}(\omega) = \sum_{s=0}^{\infty} a_s e^{-i\omega s}. \tag{5.536}$$

Since

$$\Phi(\omega) = \frac{1}{|A(\omega)|^2} = \frac{1}{\left|\sum_{s=0}^{m} a_s e^{-i\omega s}\right|^2} = \frac{1}{\sum_{\tau=-m}^{m} r_\tau e^{-i\omega\tau}} \tag{5.537}$$

where

$$r_\tau = \sum_{s=0}^{M} a_s a_{s+\tau} \tag{5.538}$$

we see that the power spectrum is a rational function in $e^{-i\omega}$.

In general the rational function in $e^{-i\omega}$

$$\left[\sum_{\tau=-m}^{m} r_\tau e^{-i\omega\tau} \right]^{-1} \tag{5.54}$$

will not be a non-negative integrable function for $-\pi < \omega \leq \pi$. The function (5.54) will be non-negative if the $r_o, r_1, r_2 \cdots r_m$ satisfy the conditions of Wold's Theorem 12, which we stated in Part A of this Section in connection with the process of finite moving averages. The function (5.54) will be integrable for $-\pi \leq \omega \leq \pi$ if the polynomial

$$z^m \sum_{\tau=-m}^{m} r_\tau z^\tau \tag{5.541}$$

has no roots of modulus one. If these conditions are satisfied, then we may let the rational function (5.54) represent the power spectrum $\Phi(\omega)$ of an autoregressive process. Then the inverse power spectrum

$$\Psi(\omega) = \frac{1}{\Phi(\omega)} = \sum_{\tau=-m}^{m} r_\tau e^{-i\omega\tau} = r_o + 2 \sum_{\tau=1}^{m} r_\tau \cos \omega\tau. \tag{5.542}$$

may be factored according to Wold's method given in Section 2.8 so that

$$\mathbb{\Psi}(\omega) = A(\omega)\,\overline{A(\omega)} \tag{5.543}$$

where

$$A(\omega) = \sum_{s=o}^{m} a_s e^{-i\omega s} \tag{5.545}$$

has no zeros or singularities in the lower half λ-plane. Then the spectrum is given by

$$\mathbb{\Phi}(\omega) = \frac{1}{A(\omega)}\,\frac{1}{\overline{A(\omega)}} = B(\omega)\,\overline{B(\omega)} \tag{5.546}$$

where the factor

$$\frac{1}{A(\omega)} = B(\omega) \tag{5.547}$$

is also free of singularities and zeros in the lower half λ-plane. In Figures 6 and 7, the various time and frequency functions of a particular second order autoregressive process are given. Two fundamental sets of difference equations (Wold, 1938; Kendall, 1946) exist for the autoregressive process

$$x_t = \sum_{r=o}^{\infty} b_r \xi_{t-r}. \tag{5.55}$$

Let us multiply this equation by the equation for the inverse decomposition at t+n:

$$\sum_{s=o}^{m} a_s x_{t+n-s} = \xi_{t+n}, \quad n > 0 \tag{5.551}$$

which yields

$$x_t \sum_{s=o}^{m} a_s x_{t+n-s} = \xi_{t+n} \sum_{r=o}^{\infty} b_r \xi_{t-r}, \; n > 0 \tag{5.552}$$

and, by taking the expectation of each side, we have

$$\sum_{s=o}^{m} a_s E[x_t x_{t+n-s}] = \sum_{r=o}^{\infty} b_r E[\xi_{t+n} \xi_{t-r}], \quad n > 0 \tag{5.553}$$

which yields the set of difference equations

$$\sum_{s=o}^{m} a_s \phi(n-s) = 0, \quad n > 0. \tag{5.554}$$

Let us multiply the process (5.55) by the inverse decomposition at t-n:

$$\sum_{s=o}^{m} a_s x_{t-n-s} = \xi_{t-n}, \; n \geq 0 \tag{5.555}$$

which yields

$$x_t \sum_{s=o}^{m} a_s x_{t-n-s} = \xi_{t-n} \sum_{r=o}^{\infty} b_r \xi_{t-r}, \; n \geq 0. \tag{5.556}$$

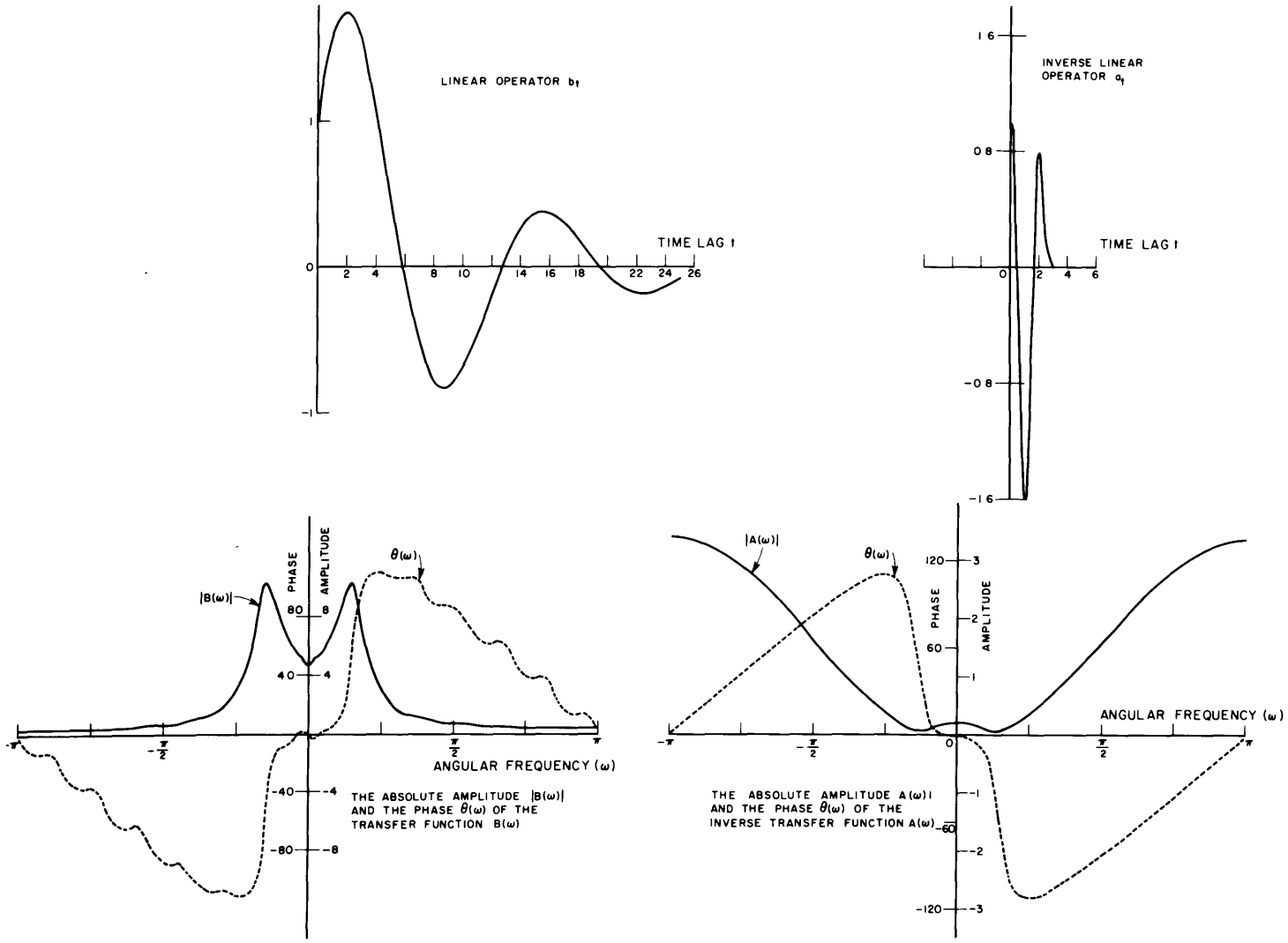
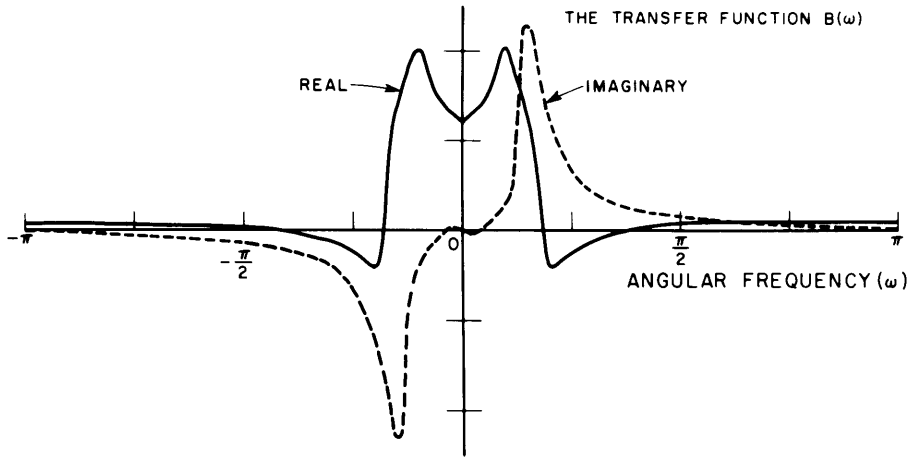
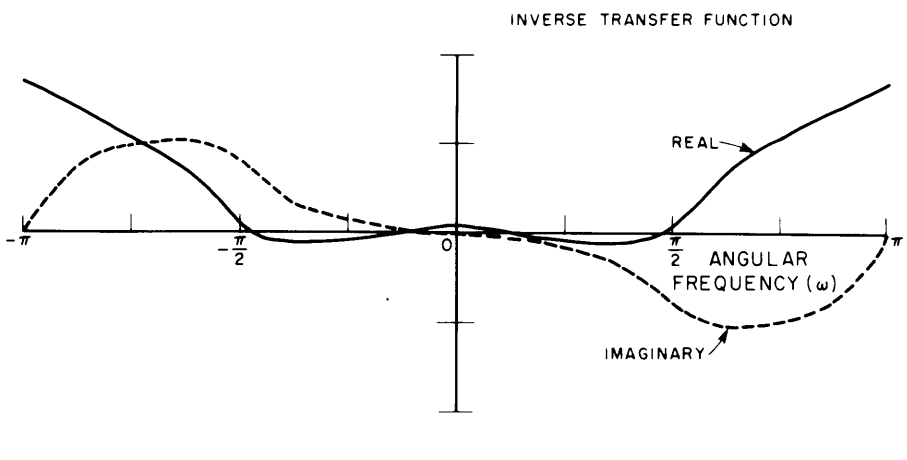
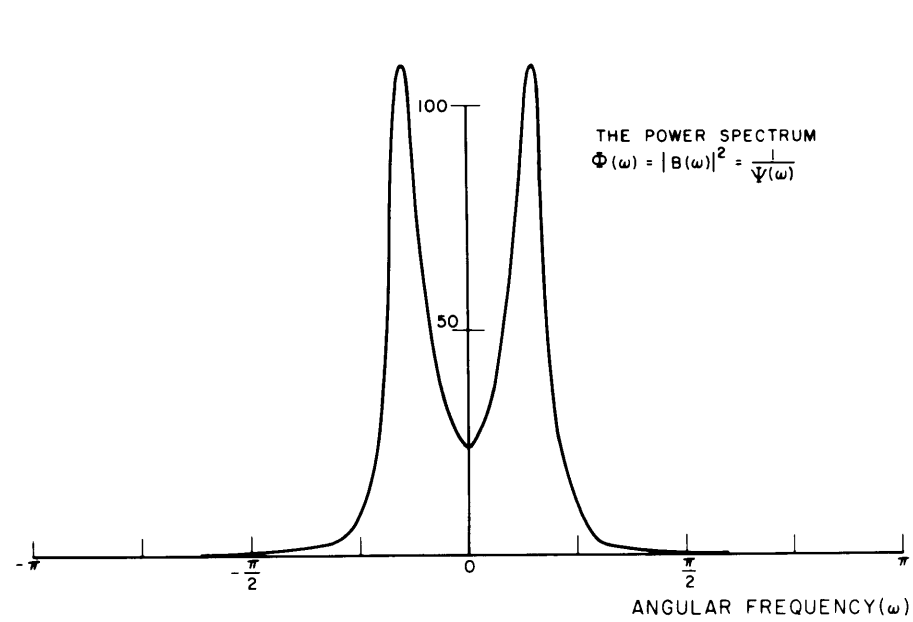


FIGURE 6 THE AUTOREGRESSIVE PROCESS $a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} = \xi_t$







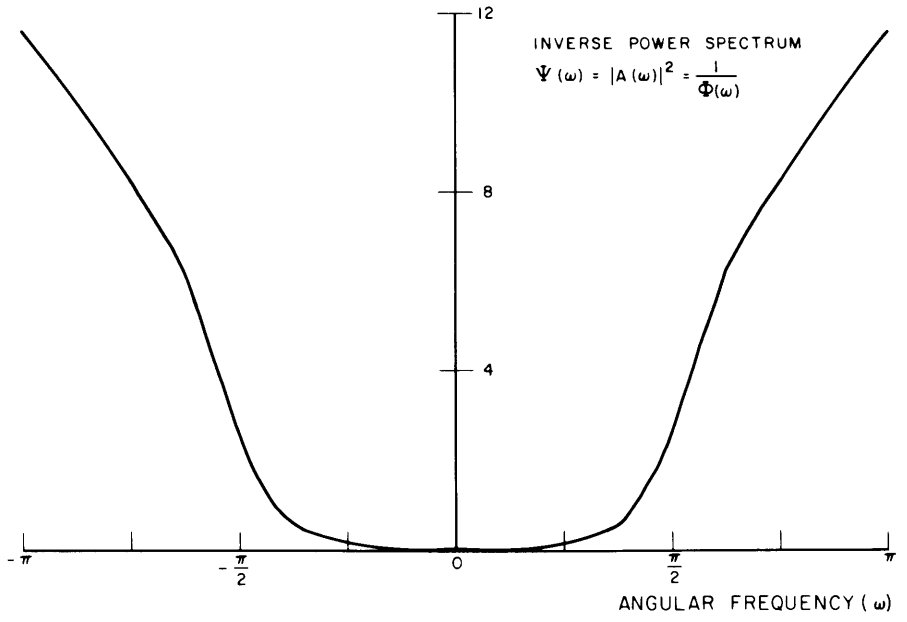


FIGURE 7 THE AUTOREGRESSIVE PROCESS $a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} = \xi_t$

Taking the expectation of each side, we obtain the other set of difference equations

$$\sum_{s=o}^{m} a_s \phi(n+s) = b_n, \quad n \geq 0. \tag{5.557}$$

The first m-1 linear equations of the set (5.554) are

$$\begin{aligned}
&\phi(1) + a_1\phi(0) + a_2\phi(1) + \ldots + a_m\phi(m-1) = 0 \\
&\phi(2) + a_1\phi(1) + a_2\phi(0) + \ldots + a_m\phi(m-2) = 0 \\
&\phi(3) + a_1\phi(2) + a_2\phi(1) + \ldots + a_m\phi(m-3) = 0 \\
&\ldots \\
&\phi(m-1) + a_1\phi(m-2) + a_2\phi(m-3) + \ldots + a_m\phi(1) = 0 \\
&\phi(m) + a_1\phi(m-1) + a_2\phi(m-2) + \ldots + a_m\phi(0) = 0
\end{aligned} \tag{5.56}$$

which correspond to the normal equations of the Gauss method of least squares (equations (40) of Wadsworth, et al (1953)). Thus, if we know the values of $\phi(0), \phi(1), \ldots, \phi(m)$, that is, the first m lags of the autocorrelation function, we may compute $a_o = 1, a_1, a_2, \ldots a_m$. It is this property which makes the autoregressive process a fundamental model in statistical work. That is, the statistician need only estimate the first m lags of the autocorrelation to specify an m-th order autoregressive process. Under the autoregressive hypothesis, the higher lags of the autocorrelation function may be found by successive recursions of equation (5.554).

C. The Markov Process

An autoregressive process of the first order as defined in Section 5.5-B is a Markov process. For a more general definition of the Markov process, see Feller (1950) and Doob (1953). Applications of the autoregressive process and Markov process to economic analysis are given by the author (1952).

In order to obtain a Markov process, we let m=1 in equation (5.53). Hence the impulse ξ_t has the representation

$$\xi_t = x_t + a_1 x_{t-1}, \quad a_o = 1, \quad |a_1| < 1. \tag{5.57}$$

The response function b_t (t=0,1,2,...) may be found by repeated iterations of the difference equation

$$0 = b_t + a_1 b_{t-1}, \quad t = 1,2,\ldots \quad b_o = 1 \tag{5.571}$$

which yields

$$b_t = - a_1 b_{t-1} = (-a_1)^t. \tag{5.572}$$

Thus the predictive decomposition of the Markov process is

$$x_t = \sum_{s=o}^{\infty} b_s \xi_{t-s} = \sum_{s=o}^{\infty} (-a_1)^s \xi_{t-s}. \tag{5.573}$$

The autocorrelation function is

$$\phi(\tau) = E[x_t x_{t+\tau}] = \sum_{t=o}^{\infty} b_t b_{t+\tau} = \sum_{t=o}^{\infty} (-a_1)^{2t+\tau}, \quad \tau > 0 \qquad (5.574)$$

which is

$$\phi(\tau) = (-a_1)^{|\tau|} \sum_{t=o}^{\infty} (-a_1)^{2t} = \frac{(-a_1)^{|\tau|}}{(1-a_1^2)}. \qquad (5.575)$$

Thus we see that the autocorrelation of such a Markov process is an exponential which is the case of pure persistance. Equation (5.575) is the discrete time series analogue of equation (16) in Wadsworth, et al (1953).

The spectrum $\overline{\Phi}(\omega)$ is given by equation (5.534) with m = 1; that is,

$$\overline{\Phi}(\omega) = \frac{1}{|1+a_1 e^{-i\omega}|^2}, \quad a_o = 1, \ |a_1| < 1. \qquad (5.576)$$

The prediction operator $k_t(\alpha)$, given by equation (5.332), becomes

$$k_o(\alpha) = b_\alpha = (-a_1)^\alpha$$

$$k_t(\alpha) = 0, \text{ for } t = 1,2,3,... \qquad (5.58)$$

so that the optimum prediction is given by

$$x_{t+\alpha} = (-a_1)^{\alpha} x_t. \tag{5.581}$$

The filter characteristics of the prediction operator are

$$K_{\alpha}(\omega) = \sum_{t=o}^{\infty} k_t(\alpha) e^{-i\omega t} = (-a_1)^{\alpha} \tag{5.582}$$

The mean square error of prediction given by equation (5.323), becomes

$$I_{min} = 1 + b_1^2 + b_2^2 + \ldots + b_{\alpha-1}^2$$

$$= 1 + a_1^2 + a_1^4 + \ldots + a_1^{2\alpha-2}. \tag{5.583}$$

D. Hybrid Processes

Let x_t be a process with power spectra, which is a rational function in $e^{-i\omega}$,

$$\Phi(\omega) = \frac{|B(\omega)|^2}{|A(\omega)|^2} \tag{5.59}$$

where

$$A(\omega) = \sum_{s=o}^{m} a_s e^{-i\omega s} \tag{5.591}$$

and

$$B(\omega) = \sum_{s=o}^{M} a_s e^{-i\omega s}. \qquad (5.592)$$

Here the polynomials A(z) and B(z), $z = e^{-i\omega}$, are required to have no common factors, the roots of A(z) to have modulus greater than one, and the roots of B(z) to have modulus greater than or equal to one (See Figure 3). Then Doob (1949, 1953) shows that the process x_t, with spectrum $\Phi(\omega)$, is an hybrid between an autoregressive process and a finite moving average process. That is, we have

$$\sum_{s=o}^{m} a_s x_{t-s} = \gamma_t \qquad (5.593)$$

where γ_t is the moving average

$$\gamma_t = \sum_{s=o}^{M} b_s \xi_{t-s} \qquad (5.594)$$

with

$$E(\xi_t^2) = 1$$

$$E(\xi_t \xi_s) = 0 \text{ for } s \neq t. \qquad (5.595)$$

5.6 Multiple Time Series

In this section, and in the next, we wish to consider multiple discrete stationary ergodic time series with absolutely continuous spectral distribution functions. We shall have to modify our previous notation to some extent in order to accommodate the bulk of notation required. Let us consider the set of stationary processes $x_j(t)$, $(j=1,2,\ldots n)$ which we take to be real functions. From now on, where two or more secondary symbols appear, subscripts will denote the particular time series under consideration, whereas the time parameter will appear in the parentheses following the symbol for the function.

We define the correlation functions (Cramer, 1940) to be

$$\phi_{jk}(\tau) = E[x_j(t+\tau)x_k(t)] = \lim_{T\to\infty} \frac{1}{2T+1} \sum_{t=-T}^{T} x_j(t+\tau)\, x_k(t). \tag{5.61}$$

For $j = k$ this equation gives the autocorrelation function of $x_j(t)$, whereas for $j \neq k$ it gives the cross-correlation function of $x_j(t)$ and $x_k(t)$. We have

$$\phi_{jj}(\tau) = \phi_{jj}(-\tau), \quad j = 1,2,\ldots n \tag{5.611}$$

which states that the autocorrelation function is an even function of τ. For the cross-correlation functions, we have

$$\phi_{jk}(\tau) = \phi_{kj}(-\tau), \quad k \neq j, \qquad j,k = 1,2,\ldots n. \tag{5.612}$$

Since the time series $x_j(t)$ are real functions of time t, the correlation functions are real functions of τ. From their definition (5.61), the Schwarz inequality gives

$$|\phi_{jk}(\tau)| \leq \sqrt{\phi_{jj}(0)\,\phi_{kk}(0)} \tag{5.613}$$

which provides a basis for normalizing the correlation functions.

Cramer (1940) shows that these correlation functions may be expressed as Fourier-Stieltjes integrals of the form:

$$\phi_{jk}(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega\tau} \, \Phi_{jk}(\omega)\, d\omega, \text{ for } j,k = 1,2,\ldots n. \tag{5.62}$$

The inverse transforms may be written as

$$\Phi_{jk}(\omega) = \sum_{\tau=-\infty}^{\infty} e^{-i\omega\tau} \, \phi_{jk}(\tau). \tag{5.63}$$

Here the $\Phi_{jk}(\omega)$ are the spectra of the set of stationary processes. For j = k, we have the power spectrum of $x_j(t)$:

$$\Phi_{jj}(\omega) = \Phi_{jj}(-\omega), \quad j = 1,2,\ldots n \tag{5.631}$$

which is a positive real function of ω. For $j \neq k$, we have the cross-power spectrum $\Phi_{jk}(\omega)$ of $x_j(t)$ and $x_k(t)$. The cross-power spectrum, which is a complex valued function of the real variable ω, satisfies

$$\Phi_{jk}(\omega) = \Phi_{kj}(-\omega) \tag{5.632}$$

and

$$\Phi_{jk}(\omega) = \overline{\Phi_{kj}(\omega)} \tag{5.633}$$

where the bar indicates the complex conjugate. Consequently, we have

$$\Phi_{jk}(\omega) = \overline{\Phi_{jk}(-\omega)}. \tag{5.634}$$

Thus we see that the real part of the cross spectrum

$$\mathrm{Re}\,[\,\Phi_{jk}(\omega)] = \mathrm{Re}\,[\,\Phi_{jk}(-\omega)] \tag{5.635}$$

is an even function of the real variable ω, whereas the imaginary part

$$\mathrm{Im}[\,\bar{\Phi}_{jk}(\omega)] = -\mathrm{Im}\,[\,\bar{\Phi}_{jk}(-\omega)] \tag{5.636}$$

is an odd function of the real variable ω.

By letting $\tau = 0$ in equation (5.62), we see that

$$\phi_{jk}(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \bar{\Phi}_{jk}(\omega)\, d\omega = \frac{1}{\pi}\int_{0}^{\pi} \mathrm{Re}\,[\,\bar{\Phi}_{jk}(\omega)]\, d\omega \tag{5.64}$$

Let us consider the linear combination

$$x(t) = \sum_{1}^{n} a_j x_j(t) \tag{5.65}$$

which defines the time series $x(t)$. Here the weighting coefficients a_j are real. If $x_j(t)$, $j = 1,2,\ldots n$, represents the traces on a seismogram, and if $a_j = 1$, $j = 1,2,\ldots n$, then $x(t)$ is the so-called Reber trace. The autocorrelation of $x(t)$ is

$$\phi(\tau) = \lim_{T\to\infty} \frac{1}{2T+1} \sum_{t=-T}^{T} x(t+\tau)\, x(t)$$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k \lim_{T\to\infty} \frac{1}{2T+1} \sum_{t=-T}^{T} x_j(t+\tau)\, x_k(t) \tag{5.66}$$

which is

$$\phi(\tau) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \quad \phi_{jk}(\tau). \tag{5.661}$$

Since $\phi(\tau)$ is the autocorrelation function of the time series $x(t)$ it is an even function of τ:

$$\phi(\tau) = \phi(-\tau). \tag{5.662}$$

The spectrum of $x(t)$ is given by the hermitian form

$$\overline{\Phi}(\omega) = \sum_{\tau=-\infty}^{\infty} \phi(\tau)\, e^{-i\omega\tau}\, d\tau = \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \quad \overline{\Phi}_{jk}(\omega) \geq 0 \tag{5.67}$$

which is a non-negative function of ω, (Cramér, 1940). The matrix of hermitian form (5.67) may be represented by the hermitian matrix

$$[\, \overline{\Phi}_{jk}(\omega)], \text{ for } j,\ k = 1,2,\ldots,\ n \tag{5.671}$$

which is

$$\begin{bmatrix} \overline{\Phi}_{11}(\omega) & \overline{\Phi}_{12}(\omega) & \cdots & \overline{\Phi}_{1n}(\omega) \\ \overline{\Phi}_{21}(\omega) & \overline{\Phi}_{22}(\omega) & \cdots & \overline{\Phi}_{2n}(\omega) \\ & \cdots & & \\ \overline{\Phi}_{n1}(\omega) & \overline{\Phi}_{n2}(\omega) & \cdots & \overline{\Phi}_{nn}(\omega) \end{bmatrix} \tag{5.672}$$

We shall call this matrix, which determines the spectra of all possible linear combinations of $x_1(t),\ldots,x_n(t)$, the coherency matrix. The elements of our coherency matrix are the derivatives of the elements of Wiener's coherency matrix (Wiener, 1930, 1942).

Further, for the time series $x_1(t)$ and $x_2(t)$ with the coherency matrix

$$\begin{bmatrix} \Phi_{11}(\omega) & \Phi_{12}(\omega) \\ \Phi_{21}(\omega) & \Phi_{22}(\omega) \end{bmatrix} \tag{5.673}$$

the significant invariants of this Hermitian matrix are

$$\mathrm{coh}_{12}(\omega) = \frac{\Phi_{12}(\omega)}{[\,\Phi_{11}(\omega)\,\Phi_{22}(\omega)]^{\frac{1}{2}}} \tag{5.674}$$

which Wiener (1930) calls the coefficient of coherency of $x_1(t)$ and $x_2(t)$ for frequency ω, and

$$\sigma_1(\omega) = \frac{\Phi_{12}(\omega)\sqrt{\Phi_{11}(\omega)}}{\Phi_{22}(\omega)} \tag{5.675}$$

and

$$\sigma_2(\omega) = \frac{\Phi_{21}(\omega)\sqrt{\Phi_{22}(\omega)}}{\Phi_{11}(\omega)} \tag{5.676}$$

which Wiener (1930) calls the coefficients of regression respectively of x_1 on x_2 and of x_2 on x_1. Wiener (1930) points out that the modulus of the coefficient of coherency represents the amount of linear coherency between $x_1(t)$ and $x_2(t)$ and the argument, the phase-lag of this coherency. The coefficients of regression determine in addition the relative scale for equivalent changes of $x_1(t)$ and $x_2(t)$.

Cramér (1940) shows that the determinant

$$\Phi_{11}(\omega)\,\Phi_{22}(\omega) - \Phi_{12}(\omega)\,\Phi_{21}(\omega) \tag{5.677}$$

of the coherency matrix (5.673) is non-negative. Therefore we have

$$\Phi_{12}(\omega)\,\Phi_{21}(\omega) = |\Phi_{12}(\omega)|^2 \leq \Phi_{11}(\omega)\,\Phi_{22}(\omega) \tag{5.678}$$

so that the magnitude of the coefficient of coherency,

$$|\mathrm{coh}_{12}(\omega)| = \frac{|\Phi_{12}(\omega)|}{\sqrt{\Phi_{11}(\omega)\Phi_{22}(\omega)}} \tag{5.679}$$

lies between zero and one.

Inequality (5.678) may be written

$$\sum_{\tau=-\infty}^{\infty} \phi_{12}(\tau)e^{-i\omega\tau} \sum_{s=-\infty}^{\infty} \phi_{21}(s)e^{-i\omega s} \leq \sum_{\tau=-\infty}^{\infty} \phi_{11}(\tau)e^{-i\omega\tau} \sum_{s=-\infty}^{\infty} \phi_{22}(s)e^{-i\omega s}$$

(5.68)

which is

$$\sum_{r=-\infty}^{\infty} \sum_{\tau=-\infty}^{\infty} \phi_{12}(\tau)\phi_{12}(r)e^{-i\omega(\tau-r)} \leq \sum_{r=-\infty}^{\infty} \sum_{\tau=-\infty}^{\infty} \phi_{11}(\tau)\phi_{22}(r)e^{-i\omega(\tau-r)}$$

(5.681)

or

$$\sum_{n=-\infty}^{\infty} e^{-i\omega n} \sum_{r=-\infty}^{\infty} \phi_{12}(n+r)\phi_{12}(r) \leq \sum_{n=-\infty}^{\infty} e^{-i\omega n} \sum_{r=-\infty}^{\infty} \phi_{11}(n+r)\ \phi_{22}(r).$$

(5.682)

This inequality states that for two stationary time series the Fourier transform of the autocorrelation of their cross-correlation cannot exceed the Fourier transform of the cross-correlation of their autocorrelations.

As Wold (1953, Chapter 12.7, p 202) observes, in view of the abundance of possible variations and combinations available

in the extention of the theory of one dimensional stochastic processes to multiple stochastic processes, the main difficulty does not lie in developing formulae of great generality, but rather in picking out those processes for further study that merit interest from the point of view of applications. This observation is particularly pertinent concerning seismic applications, where any statistical approach to the study of multiple seismic traces should originate from considerations of the physical phenomena which generate these traces in time and space.

As an example of multiple time series let us consider the special case in which two stationary time series have the predictive decompositions (see Section 5.2),

$$x_1(t) = \sum_{s=o}^{\infty} b_s \xi_{t-s} \tag{5.683}$$

$$x_2(t) = \sum_{s=o}^{\infty} d_s \gamma_{t-s}$$

where the prediction error ξ_t of $x_1(t)$ represents a mutually uncorrelated process, that is,

$$E[\xi_t^2] = 1, \; E[\xi_t \xi_s] = 0, \; t \neq s,$$

where the prediction error of $x_2(t)$ γ_t represents a mutually uncorrelated process, that is

$$E[\gamma_t^2] = 1, \; E[\gamma_t \gamma_s] = 0, \; t \neq s, \tag{5.684}$$

and where the cross-correlation of the prediction errors, ξ_t and γ_t, is given by

$$E[\xi_t \gamma_{t+\tau}] = \begin{cases} \rho \text{ for } \tau=\alpha \\ 0 \text{ otherwise.} \end{cases} \tag{5.685}$$

The inequality (5.682) becomes

$$\rho^2 \leq 1. \tag{5.686}$$

The autocorrelation $\phi_{11}(\tau)$ of $x_1(t)$ is

$$\phi_{11}(\tau) = E[x_1(t)x_1(t+\tau)] = \sum_{s=o}^{\infty} b_s b_{s+\tau} \tag{5.687}$$

and the spectrum is

$$\Phi_{11}(\omega) = B(\omega)\,\overline{B(\omega)} = |B(\omega)|^2 \tag{5.688}$$

where

$$B(\omega) = \sum_{s=o}^{\infty} b_s e^{-i\omega s} \tag{5.689}$$

Likewise we have

$$\phi_{22}(\tau) = \sum_{s=o}^{\infty} d_s d_{s+\tau} \tag{5.69}$$

and

$$\overline{\Phi}_{22}(\tau) = D(\omega)\ \overline{D(\omega)} = |D(\omega)|^2 \tag{5.691}$$

where

$$D(\omega) = \sum_{s=o}^{\infty} d_s e^{-i\omega s}. \tag{5.692}$$

The cross-correlation of $x_1(t)$ and $x_2(t)$ for this special case is

$$\phi_{12}(\tau) = E[x_1(\omega)x_2(t)] = \rho \sum_{s=o}^{\infty} b_s d_{s+\tau-\alpha} \tag{5.693}$$

and the cross-spectrum is

$$\overline{\Phi}_{12}(\omega) = \rho \sum_{\tau=-\infty}^{\infty} \left(\sum_{s=o}^{\infty} b_s d_{s+\tau-\alpha} \right) e^{-i\omega\tau} \tag{5.694}$$

$$= \rho e^{-i\omega\alpha} \sum_{n=-\infty}^{\infty} \sum_{s=o}^{\infty} b_s d_{s+n}\ e^{-i\omega n}$$

$$= \rho e^{-i\omega\alpha}\ \overline{B(\omega)}\ D(\omega).$$

Thus the coherency matrix, equation (5.673), for this special case is

$$\begin{bmatrix} B\bar{B} & \rho e^{-i\omega\alpha} B\bar{D} \\ \rho e^{i\omega\alpha} B\bar{D} & D\bar{D} \end{bmatrix} \tag{5.695}$$

and the coefficient of coherency is

$$coh_{12}(\omega) = \rho e^{-i\omega\alpha} \frac{\overline{B(\omega)}}{|B(\omega)|} \frac{D(\omega)}{|D(\omega)|} , \tag{5.696}$$

the magnitude of which is

$$|\rho| \leq 1. \tag{5.697}$$

We see that the coherency of the two time series, $x_1(t)$ and $x_2(t)$, depends on the cross-correlation of their respective prediction errors, ξ_t and γ_t.

In general, if the magnitude of the coefficient of coherency is equal to one, we say the two time series are completely coherent; if equal to zero, completely incoherent. For completely coherent time series, the coherency matrix is singular.

The study of multiple time series may lead into the consideration of various structural systems, for example, recursive systems and non-linear systems, (Wold, 1953). For any set of multiple time series with well-defined moments, a theorem of Wold (1948, 1953) secures a recursive representation in which the endogenous variables form a recursive causal chain. Also Wold (1953) discusses the conditions under which a unique predictive decomposition of a system of multiple time series may be obtained.

The concept of coherency is an important one in the study of seismic records. Computations carried out by the MIT Geophysical Analysis Group indicate that seismic traces are more coherent on the average in an interval containing a major reflection than in an adjacent non-reflection interval. This coherency property of reflections assists the visual detection of reflections on a seismogram, and hence may be exploited in the detection of weak reflections by statistical methods. In these computations, the coherency was estimated through the estimation of the various power spectra and cross-spectra. The problem of how to estimate spectra from finite time series is a major problem. In Figure 8, we show examples of correlation functions and spectra computed according to the method of Tukey (1949) and Tukey and Hamming (1949) from MIT Record No. 1 (supplied by the Magnolia Petroleum Co.) over the time

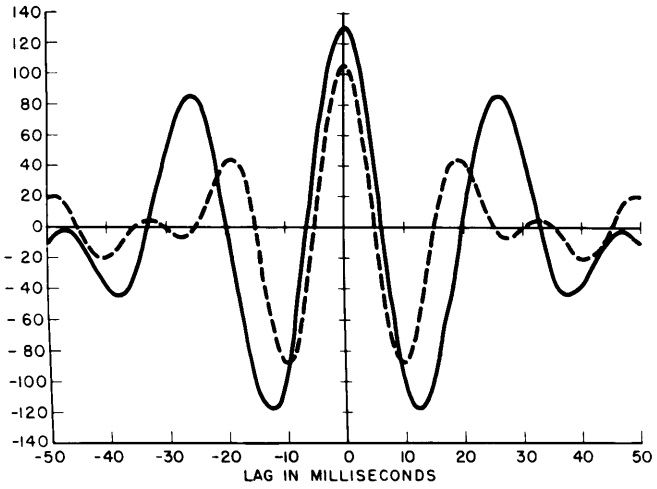


DIAGRAM A. AUTOCORRELATIONS FOR: ——— TRACE N750
– – – TRACE N250

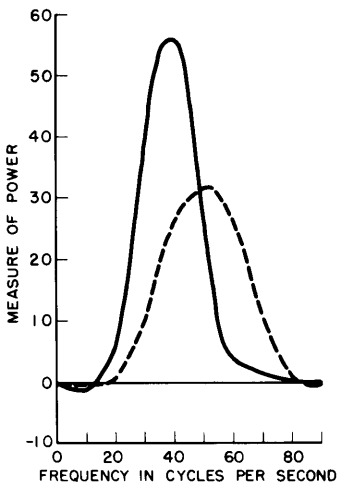


DIAGRAM B: POWER DENSITY SPECTRA FOR: ——— TRACE N 750
– – – TRACE N 250
(COMPUTED FROM AUTO-CORRELATIONS OF DIAGRAM A)

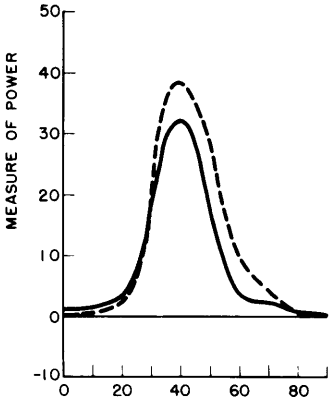


DIAGRAM E: – – – GEOMETRIC MEAN OF SPECTRA SHOWN IN DIAGRAM B
——— ABSOLUTE VALUE OF CROSS SPECTRUM SHOWN IN DIAGRAM D.

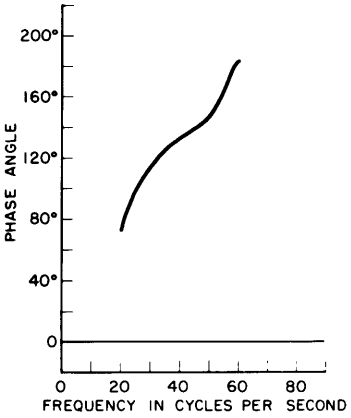


DIAGRAM F: PHASE ANGLE OF COEFFICIENT OF COHERENCY AS A FUNCTION OF FREQUENCY FROM CROSS SPECTRUM SHOWN IN DIAGRAM D

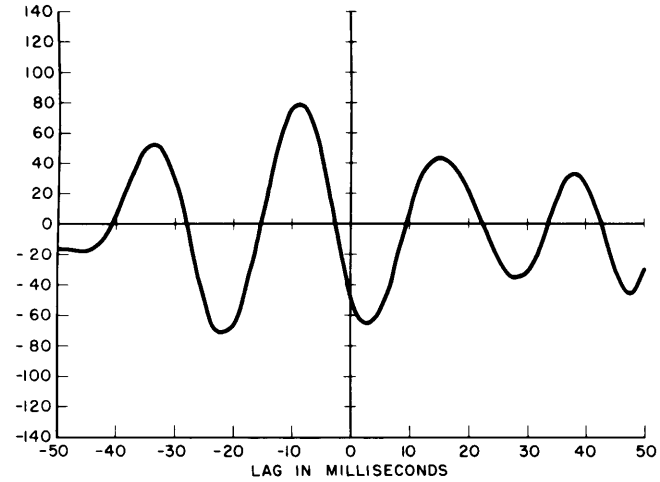


DIAGRAM C: CROSS CORRELATION OF TRACES N750 AND N250

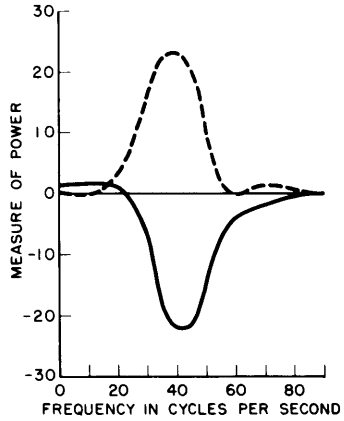


DIAGRAM D: CROSS POWER DENSITY SPECTRUM OF TRACES N750 AND N250:
——— REAL COMPONENT
– – – IMAGINARY COMPONENT
(COMPUTED FROM CROSS-CORRELATION OF DIAGRAM C)

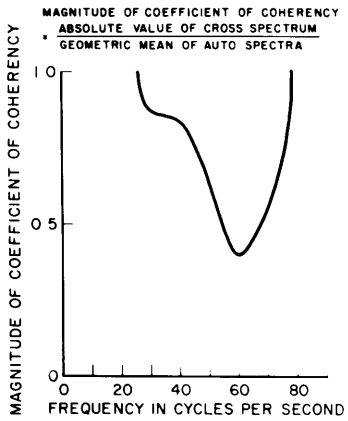


DIAGRAM G: MAGNITUDE OF COEFFICIENT OF COHERENCY AS A FUNCTION OF FREQUENCY FROM SPECTRA SHOWN IN DIAGRAM E

FIGURE 8 CORRELATION FUNCTIONS AND SPECTRA, ON M.I.T. RECORD NO. 1 FROM TIME EQUAL TO 1.05 SECONDS TO 1 225 SECONDS

interval from 1.05 seconds to 1.225 seconds. Further discussion of this Figure and of MIT Record No. 1 is given in Wadsworth, et al, (1953) where also the method of Tukey and Hamming for the estimation of power spectra is presented. Also the reader is referred to MIT GAG Report No. 5 for an extention of the method of Tukey and Hamming to the estimation of cross-spectra.

Under the hypothesis that a finite section of a seismic trace is a section of a stationary time series, which we discuss in the next chapter, a good estimate of the prediction operator for a single trace is one which yields prediction errors which satisfy statistical tests of being mutually uncorrelated. Then another approach to the study of coherency involves the examination of the cross-correlation existing between the sets of prediction errors for the various traces. Here again the estimation problem is a major problem.

Much of the book of Wold (1953) is devoted to the linear least-squares regression method which is the method for the determination of linear operators by the Gauss method of least squares (Wadsworth, et al, 1953). In particular Wold considers least-squares regression methods for time series from the viewpoint of sampling theory, and obtains conditions under which empirical operator coefficients computed from a finite amount of data are unbiased and consistent estimates of the theoretical operator coefficients of the stochastic process.

A simple test to determine whether the empirical operator for a stationary time series is a good estimate of the theoretical operator consists of testing whether the empirical prediction errors are mutually uncorrelated.

5.7 General Technique of Discrete Prediction

In this section we should like to review the "General Technique of Discrete Prediction" of Wiener (1942, Section 4.6). In particular we should like to question his technical point concerning the computation of power spectra for finite time series because, for one reason, it leads to an indeterminate solution in the application of his general technique of discrete prediction to finite time series.

His technical point is that the correlation functions of the finite time series $x_i(t)$, $(-N \leq t \leq N)$, should be computed by the formula

[4.625]

$$\phi_{ij}(\tau) = \frac{1}{2N+1} \sum_{\substack{-N \leq n \leq N \\ -N \leq n+\tau \leq N}} x_i(n+\tau)x_j(n) \quad i,j = 1,2,\ldots n \tag{5.71}$$

which, except for the constant factor $(2N+1)^{-1}$, is the same formula as used for the computation of the correlation functions of finite time series considered to be aperiodic functions.

(See Section 3.3) Wiener (1930) also suggests formula (5.71) for the computation of correlation functions for finite time series. That is, in effect, Wiener (1930, 1942) assumes that the finite time series $x_i(t)$ are aperiodic functions such that

$$x_t(t) = 0 \text{ for } t < -T \text{ and } t > T \tag{5.711}$$

in the computation of the correlation function (5.71).

As we have seen in Section 3.3 the phase and amplitude spectra are given by

$$X_i(\omega) = \sum_{t=-N}^{N} x_i(t)e^{-i\omega t} \tag{5.712}$$

and the spectra by

$$\overline{\Phi}_{ij}(\omega) = \overline{X_i(\omega)}X_j(\omega) = \sum_{\tau=-N}^{N} \phi_{ij}(\tau)e^{-i\omega\tau} \tag{5.713}$$

which are terminating series in positive and negative powers of $e^{-i\omega}$. In equation (5.713) the correlation function $\phi_{ij}(\tau)$ is given by formula (5.71). Thus equation (5.713) represents the spectra computed according to Wiener's technical point.

Wiener states that the most essential property of a single $\phi_{ii}(\tau)$ is that it may be written in the form

(4.61) $$\int_{-\pi}^{\pi} e^{i\omega t}\, d\,\Lambda\,(\omega) \qquad (5.714)$$

where $\Lambda\,(\omega)$ is monotonically increasing; that is, the power spectra $\mathfrak{F}_{ii}(\omega)$ must be non-negative. Letting i = j in equation (5.713), we have the power spectrum given by

$$\mathfrak{F}_{ii}(\omega) = \sum_{\tau=-N}^{N} \phi_{ii}(\tau) \cos \omega\tau = \overline{X_i(\omega)}\, X_i(\omega) = |X_i(\omega)|^2 \geq 0 \qquad (5.715)$$

which is non-negative. Therefore we see that the condition that the power spectrum $\mathfrak{F}_{ii}(\omega)$ will be non-negative is satisfied for those cases in which we compute the spectrum according to the aperiodic hypothesis given in Chapter III, that is, according to Wiener's technical point as expressed by our equations (5.713) and (5.71).

Wiener's second condition is that the quadratic form

(4.62) $$\sum_{i}\sum_{j}\sum_{\tau}\sum_{\lambda} \phi_{ij}(\tau-\lambda)\; a_\tau a_\lambda \qquad (5.716)$$

must be non-negative. This condition is also satisfied for the correlation functions computed according to Wiener's technical point.

Instead of examining this quadratic form in more detail, let us examine the hermitian form

$$\sum_i \sum_j \overline{\Phi}_{ij}(\omega)\, a_i a_j \qquad (5.72)$$

which because of equation (5.67) is non-negative. The coherency matrix, which is the matrix of this hermitian form, is given by equation (5.672) where the $\overline{\Phi}_{ij}(\omega)$ are computed by Wiener's technical point, that is, by equation (5.713), together with equation (5.71). Explicitly, this coherency matrix is

$$[\mathrm{Coh}(\omega)] = [\overline{\Phi}_{ij}(\omega)] = [\sum_{\tau=-N}^{N} \phi_{ij}(\tau)\, e^{-i\omega\tau}] = [\overline{X_i(\omega)} X_j(\omega)] \qquad (5.722)$$

where $i, j = 1, 2, \ldots n$, and where the correlations $\phi_{ij}(\tau)$ are given by equation (5.71).

The hermitian matrix (5.722) for finite time series given by Wiener (1930), and again by Wiener (1942), is singular, that is, its determinant is equal to zero. This fact readily follows from the representation

$$[\mathrm{Coh}\ (\omega)] = [\overline{X}_i\ X_j]. \qquad (5.723)$$

Let us now examine the solution to the prediction problem for multiple discrete time series given by Wiener (1942, Section

4.6). We shall write the equation number of Wiener (1942) in brackets to the left of our corresponding equation. Thus we start with the fundamental set of equations (discrete analogue of Wiener-Hopf integral equation)

[4.635]

$$\phi_{1k}(\alpha+t) = \sum_{j=1}^{n} \sum_{\tau=0}^{\infty} \phi_{jk}(t-\tau)\, q_j(\tau), \quad t \geq 0 \qquad k = 1,2,\ldots n, \tag{5.73}$$

the solution of which yields the linear operator $q_j(\tau)$ where $\tau = 0,1,2,\ldots$ Since this set of equations need not hold for $t < 0$, we shall define the residual for $t < 0$ to be β_{-t}. (The β_{-t} for $t < 0$ used here is Wiener's function b_t for $t > 0$, in his Section 4.6. To avoid confusion in this section we shall follow Wiener's notation quite closely except, as before, we let small letters refer to time functions and capital letters refer to frequency functions.) We have

[4.635]

$$\phi_{1k}(\alpha+t) - \sum_{j=1}^{n} \sum_{\tau=0}^{\infty} \phi_{jk}(t-\tau)\, q_j(\tau) = \begin{cases} 0 & t \geq 0 \\ \beta_{-t} & t < 0 \end{cases} \tag{5.731}$$

where the residual β_{-t} for $t < 0$ may be arbitrarily chosen

so that the equation (5.73) is satisfied for $t \geq 0$. Let us put

[4.64]

$$\sum_{n=-\infty}^{\infty} \phi_{jk}(n)\, e^{-i\omega n} = \Phi_{jk}(\omega) \qquad (5.732)$$

and

[4.64]

$$\sum_{n=0}^{\infty} q_j(n)\, e^{-i\omega n} = Q_j(\omega). \qquad (5.733)$$

Then we have

$$\sum_{t=-\infty}^{\infty} \phi_{lk}(\alpha+t)e^{-i\omega t} - \sum_{j=1}^{n} \sum_{\tau=0}^{\infty} \sum_{t=-\infty}^{\infty} e^{-i\omega t}\, \phi_{jk}(t-\tau)\, q_j(\tau)$$

$$= \sum_{t=-\infty}^{-1} \beta_{-t}\, e^{-i\omega t} \qquad (5.734)$$

which is

$$\sum_{t=-\infty}^{\infty} \phi_{lk}(t)e^{-i\omega(t-\alpha)} - \sum_{j=1}^{n} \sum_{\tau=0}^{\infty} e^{-i\omega\tau}\, q_j(\tau) \sum_{-\infty}^{\infty} \phi_{jk}(t-\tau)e^{-i\omega(t-\tau)}$$

$$= \sum_{t=-\infty}^{-1} \beta_{-t}e^{-i\omega t} \qquad (5.735)$$

or

[4.645]

$$e^{i\omega\alpha} \Phi_{1k}(\omega) - \sum_{j=1}^{n} Q_j(\omega) \Phi_{jk}(\omega) = \sum_{t=1}^{\infty} \beta_t e^{i\omega t} \text{ for } k = 1,2,\ldots n. \tag{5.736}$$

For n = 1, (the prediction problem for a single time series), we have

$$e^{i\omega\alpha} \Phi_{11}(\omega) - \Phi_{11}(\omega) Q_1(\omega) = \sum_{t=1}^{\infty} \beta_t e^{i\omega t}. \tag{5.74}$$

Let us factor the power spectrum so that

$$\Phi_{11}(\omega) = |B(\omega)|^2 = B(\omega) \overline{B(\omega)} \tag{5.741}$$

where

$$B(\omega) = b_o + \sum_{1}^{n} b_s e^{-i\omega s}. \tag{5.742}$$

(The b_o, b_1,...b_n which we use here is our stable response function b_t of section 5.2, which in Wiener's notation is d_o, d_1, ... d_m). That is, Wiener assumes that $\Phi_{11}(\omega)$ is given by a terminating series in positive and negative powers of $e^{i\omega}$ so that it is the spectrum of a process of finite moving averages. The factorization may then be carried out as described

in Section 5.5-A. Here $B(\omega)$, given by equation (5.742), is to have no zeros or singularities below the real axis.

Then we have

$$[e^{i\omega\alpha} - Q_1(\omega)]\, B(\omega) = \sum_{t=1}^{\infty} \beta_t\, e^{i\omega t}\, \frac{1}{B(\omega)}. \quad (5.743)$$

Let us define $e_t (t = 1,2,3,\ldots)$ by

$$\frac{\sum_{t=1}^{\infty} \beta_t e^{i\omega t}}{\sum_{t=0}^{\infty} b_t e^{i\omega t}} = \sum_{t=1}^{\infty} e_t e^{i\omega t}. \quad (5.744)$$

Then equation (5.743) becomes

$$\sum_{s=\alpha}^{\infty} b_s e^{-i\omega(s-\alpha)} - Q_1(\omega)B(\omega) = \sum_{t=1}^{\infty} e_t e^{i\omega t} - \sum_{s=0}^{\alpha-1} b_s e^{-i\omega(s-\alpha)} \quad (5.745)$$

where the left hand side contains only non-positive powers of $e^{i\omega}$, whereas the right hand side contains only positive powers of $e^{i\omega}$. Thus each side is respectively equal to zero, so we have

$$\sum_{s=\alpha}^{\infty} b_s e^{-i\omega(s-\alpha)} - Q_1(\omega)\, B(\omega) = 0, \tag{5.746}$$

and solving for the filter characteristic $Q_1(\omega)$ we have

$$Q_1(\omega) = \frac{1}{B(\omega)} \sum_{s=\alpha}^{\infty} b_s e^{-i\omega(s-\alpha)}. \tag{5.747}$$

Since

$$b_s = \frac{1}{2\pi} \int_{-\pi}^{\pi} B(u)\, e^{ius}\, du, \tag{5.748}$$

we have

[2.630]

$$Q_1(\omega) = \frac{1}{2\pi B(\omega)} \sum_{s=\alpha}^{\infty} e^{-i\omega(s-\alpha)} \int_{-\pi}^{\pi} B(u)\, e^{ius}\, du \tag{5.749}$$

which is equation (5.373).

For n = 2, (the prediction problem for double time series) we have

[4.65]

$$[e^{i\alpha\omega} - Q_1(\omega)]\, \Phi_{11}(\omega) - \Phi_{21}(\omega)\, Q_2(\omega) = \sum_{s=1}^{\infty} \beta_s e^{i\omega s}$$

$$[e^{i\alpha\omega} - Q_1(\omega)]\, \Phi_{12}(\omega) - \Phi_{22}(\omega)\, Q_2(\omega) = \sum_{s=1}^{\infty} c_s e^{i\omega s} \tag{5.75}$$

where the residuals of the Wiener-Hopf equation (5.73) are β_s and c_s ($s = 1,2,3,\ldots$) for $k = 1$ and $k = 2$ respectively. The determinant of these equations is

$$\overline{\Phi}(\omega) = \begin{bmatrix} \overline{\Phi}_{11}(\omega) & \overline{\Phi}_{21}(\omega) \\ \overline{\Phi}_{12}(\omega) & \overline{\Phi}_{22}(\omega) \end{bmatrix} = \overline{\Phi}_{11}(\omega)\,\overline{\Phi}_{22}(\omega) - \overline{\Phi}_{21}(\omega)\,\overline{\Phi}_{12}(\omega) \tag{5.751}$$

which is the determinant of the coherency matrix of the time series $x_1(t)$ and $x_2(t)$. If we compute the coherency matrix according to Wiener's technical point given by our equations (5.71) and (5.713), the matrix will be singular, as we have seen by equation (5.723). Thus the determinant of the matrix of the simultaneous equations (5.75) vanishes, that is,

$$\overline{\Phi}(\omega) = (\overline{X}_1 X_1)(\overline{X}_2 X_2) - (\overline{X}_2 X_1)(\overline{X}_1 X_2) = 0.$$

Consequently, equations(5.75) have no solution, or an indeterminate solution. In other words, Wiener by his technical point, in effect, suggests that the aperiodic functional scheme (Section 3.3) be utilized for multiple finite time series, and hence the same type of difficulty is encountered for multiple linear operators as we described in Section 3.5. Thus in his section 4.6, Wiener (1942) suggests that the

geophysicist utilize a certain technical point for his computations. If the geophysicist does utilize this technical point, and if he does utilize Wiener's method of determining multiple linear operators given in the same section as the technical point, the geophysicist will obtain nonsense results.

Thus if we are to use Wiener's general technique of discrete prediction, which is a statistical technique, let us estimate our correlation functions and spectra according to formulae chosen for statistical reasons, and not for formal mathematical reasons. In particular we refer to the method of Tukey (1949) and Tukey and Hamming (1949) for the estimation of power spectra. That is, under a statistical hypothesis observational time series are not completely coherent, and thus one should utilize formulae which provide estimates of the theoretical coherency which actually exists between them.

Thus for time series which are not completely coherent, the determinant of the coherency matrix does not vanish, and equations (5.75) will have a solution for the coefficients of the multiple linear operator as given by Wiener (1942) in the remainder of his section 4.6.

For finite observational time series, the Gauss method of least squares (Wadsworth, et al, 1953) takes into account the empirical coherency existing between finite time series,

and thus yields a unique solution for the empirical operator coefficients.

In the practical task of computation, for example, on a digital computer, one must utilize discrete approximations of continuous integrals, as illustrated in Figure 9 in which the trapezoidal rule is utilized. Figures 10 and 11 illustrate discrete approximations of sets of sine and cosine waves.

Finally, we note that Michel Loève (1946) has obtained a Predictive Decomposition Theorem for a non-stationary random process which generates time series $x(t)$ of finite time duration $0 \leq t \leq T$. Karhunen (1947) also treats this decomposition problem. Davis (1952) applies this predictive decomposition to the prediction problem for non-stationary time series. Since the time series are non-stationary, ensemble averages are used instead of time averages. Thus the autocorrelation function, which plays a central role in this Predictive Decomposition Theorem, is given by

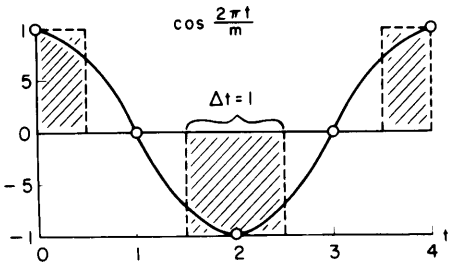
$$\phi(t,s) = E[x(t)\, x(s)] \tag{5.76}$$

which now is a function of the two time instants t and s, and is no longer a function of only their difference $\tau = t-s$ as for a stationary process. Toward the determination of the applicability of this Predictive Decomposition Theorem to

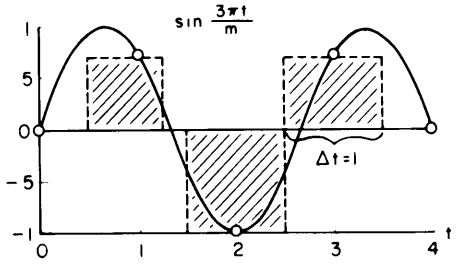
TO APPROXIMATE THE INTEGRAL $\int_0^4 \cos \frac{2\pi t}{m} \sin \frac{3\pi t}{m}\, dt = 1.5279$ WE TRANSFORM

THE CONTINUOUS COSINE AND SINE FUNCTIONS INTO BLOCKS OF WIDTH Δt FOR THE DISCRETE MULTIPLICATION

$$\cos 0 \sin 0 \frac{\Delta t}{2} + \sum_{t=1}^{3} \cos \frac{2\pi t}{m} \sin \frac{3\pi t}{m} \Delta t + \cos 2\pi \sin 3\pi \frac{\Delta t}{2}$$



AND



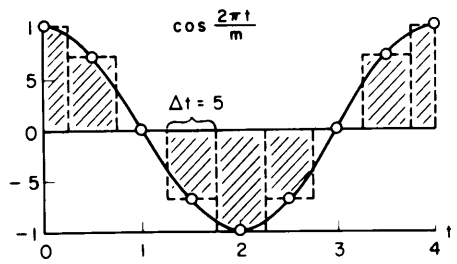
IN MATRIX NOTATION THIS IS

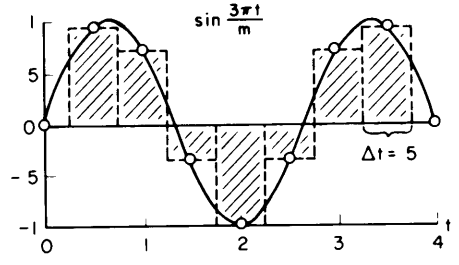
$\left(\frac{c_0}{\sqrt{2}}, c_1, c_2, c_3, \frac{c_4}{\sqrt{2}}\right)$ MULTIPLIED BY $\left(\frac{s_0}{\sqrt{2}}, s_1, s_2, s_3, \frac{s_4}{\sqrt{2}}\right)^T = \frac{c_0 s_0}{\sqrt{2}} + \sum_{j=1}^{3} c_j s_j + \frac{c_4 s_4}{\sqrt{2}}$

WHICH IS NUMERICALLY

$\left(\frac{1}{\sqrt{2}}, 0, -1, 0, \frac{1}{\sqrt{2}}\right)$ MULTIPLIED BY $\left(\frac{0}{\sqrt{2}}, .7071, -1, 7071, \frac{0}{\sqrt{2}}\right)^T = 1$

BY MAKING THE INCREMENT Δt SMALLER THE APPROXIMATION BECOMES BETTER





IN MATRIX NOTATION THIS IS

$\left(\frac{c_0}{2}, \frac{c_1}{\sqrt{2}}, \frac{c_2}{\sqrt{2}}, \frac{c_3}{\sqrt{2}}, \frac{c_4}{\sqrt{2}}, \frac{c_5}{\sqrt{2}}, \frac{c_6}{\sqrt{2}}, \frac{c_7}{\sqrt{2}}, \frac{c_8}{2}\right)$ MULTIPLIED BY $\left(\frac{s_0}{2}, \frac{s_1}{\sqrt{2}}, \frac{s_2}{\sqrt{2}}, \frac{s_3}{\sqrt{2}}, \frac{s_4}{\sqrt{2}}, \frac{s_5}{\sqrt{2}}, \frac{s_6}{\sqrt{2}}, \frac{s_7}{\sqrt{2}}, \frac{s_8}{2}\right)^T = \frac{c_0 s_0}{2} + \frac{1}{\sqrt{2}} \sum_{j=1}^{7} c_j s_j + \frac{c_8 s_8}{2}$

WHICH IS NUMERICALLY

$\left(\frac{1}{2}, \frac{.7071}{\sqrt{2}}, \frac{0}{\sqrt{2}}, \frac{-.7071}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, \frac{-.7071}{\sqrt{2}}, 0, \frac{.7071}{\sqrt{2}}, \frac{1}{2}\right)$ MULTIPLIED BY $\left(\frac{0}{2}, \frac{.9239}{\sqrt{2}}, \frac{.7071}{\sqrt{2}}, \frac{-.3827}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, \frac{-.3827}{\sqrt{2}}, \frac{.7071}{\sqrt{2}}, \frac{.9239}{\sqrt{2}}, \frac{0}{2}\right)^T = 1.4239$

FIGURE 9 DISCRETE APPROXIMATION

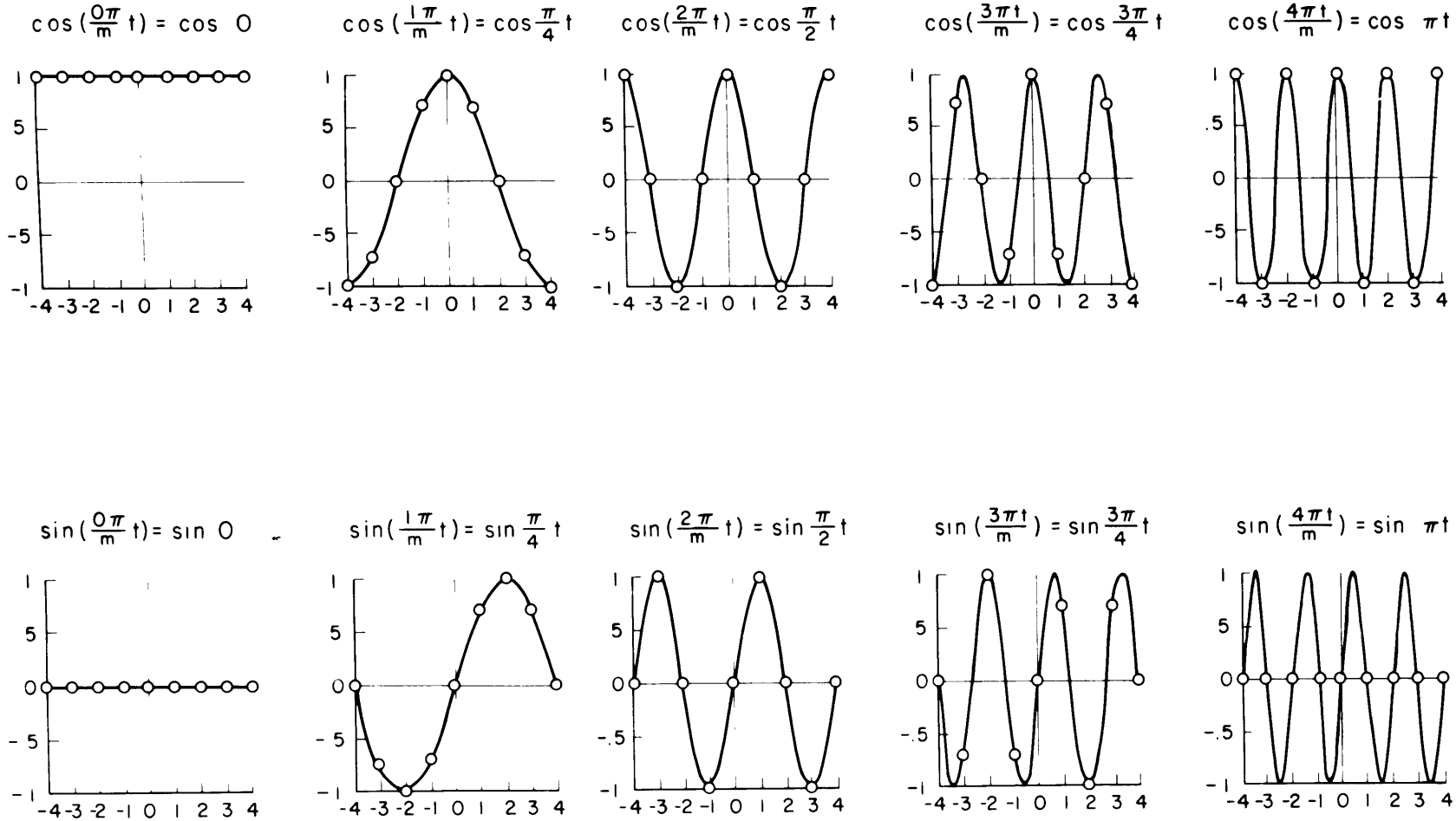


FIGURE 10 COSINE AND SINE WAVES OF ANGULAR FREQUENCIES $W = \frac{n\pi}{m}$ $(n = 0,1, \ldots, m; m = 4)$ OVER TIME INTERVAL $-4 \le t \le 4$.

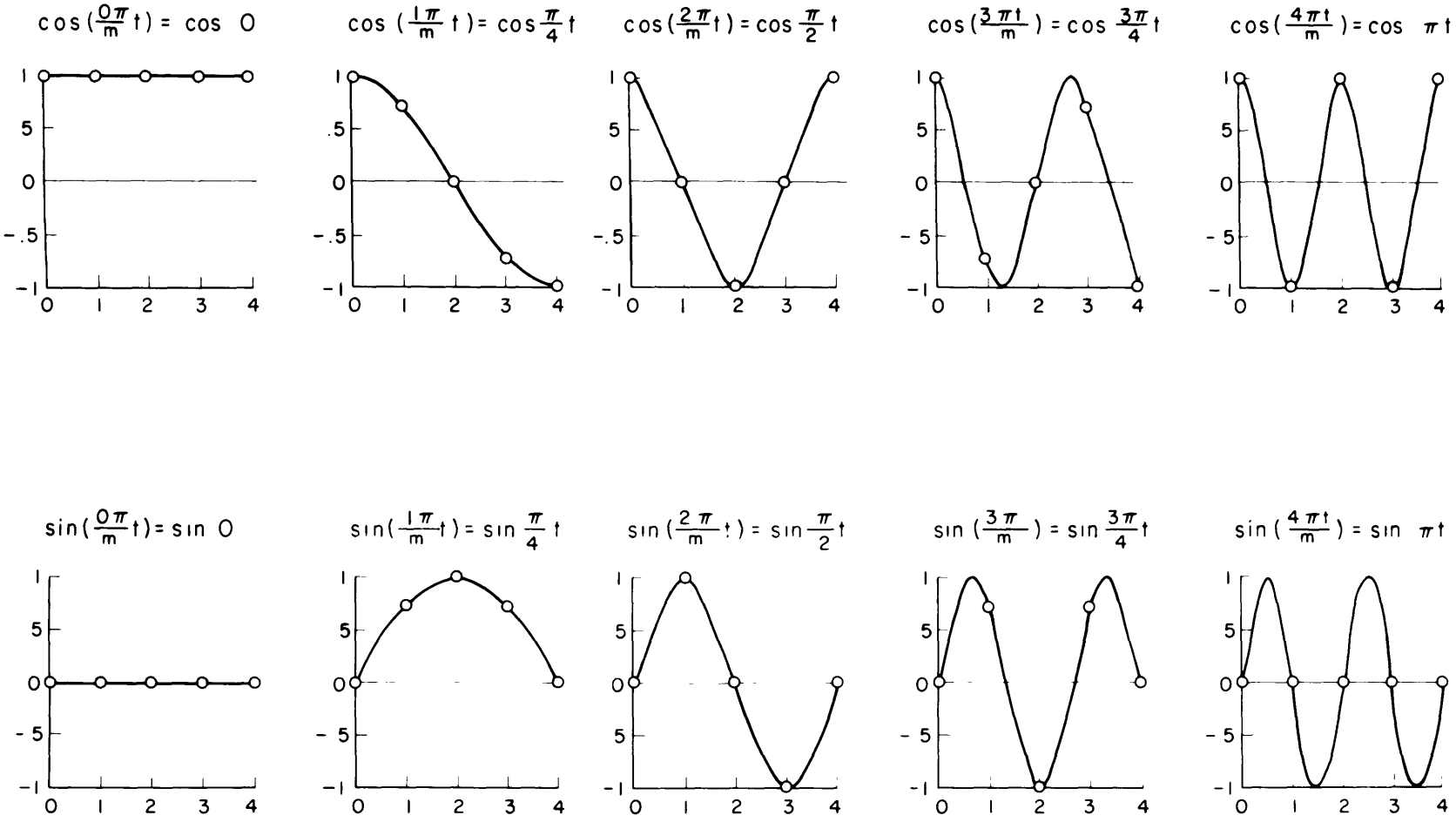


FIGURE 11 COSINE AND SINE WAVES OF ANGULAR FREQUENCIES $W = \frac{n\pi}{m}$ $(n = 0,1, m; m = 4)$ OVER TIME INTERVAL $0 \le t \le 4$

seismic data, an exploratory step would be to carry out computations to determine what degree of statistical regularity exists for estimates of the autocorrelation function (5.76).

CHAPTER VI

APPLICATIONS TO SEISMIC EXPLORATION

6.1 The Response Function

From a physical point of view, the seismic trace is the response of the system consisting of the earth and recording apparatus to the impulsive source, the explosion. This system, although usually very complicated, is susceptible to a deterministic approach toward its analysis. The explosion may be considered to yield an impulse, ξ_t, of relatively short time duration, so that the impulse ξ_t is equal to zero before the explosion and is equal to zero a short time after the explosion. In some instances the impulse function ξ_t may be considered to be sharp impulsive disturbance of very short duration. In other instances, for example, the occurence of bubble pulses in seismic prospecting over bodies of water (Worzel and Ewing, 1948), the shape of the impulse ξ_t may have a more complicated form.

The impulse ξ_t yields the energy of the seismic disturbance. This energy is dissipated in various ways as it spreads out from the source. Some of this energy is transmitted to the geophones and recorded in the form of seismic traces. Such recorded energy may be considered to be the response

function to the impulse ξ_t. The study of this response function, that is, the response of the earth and recording system to the seismic explosion, has led to many important contributions to theoretical and practical seismology, and the reader is referred to GEOPHYSICS. Instead of dealing with the response function as such, which is a time function, one may deal with its Fourier transform which is the frequency and phase components of this function in the form of spectra.

Nevertheless, the complicated nature of seismograms taken in seismic exploration many times precludes the study of the overall response of the earth and recording system as a whole. Also in the final analysis one is interested in the various components of this total response, for example, one wishes to separate components of reflected energy from those of non-reflected energy.

In a sequence of fundamental papers, Norman Ricker (1940, 1941, 1943, 1944, 1945, 1949, 1953a, 1953b) proposes the wavelet theory of seismogram structure. A seismogram, according to Ricker is an elaborate wavelet complex, and the analysis of a seismogram consists in breaking the record down into its components.

Ricker (1940) points out that, according to the theory of the propagation of elastic waves in homogeneous, isotropic

media, a wave form remains unchanged as it is transmitted. Thus a wave due to a sharp impulse, such as explosion, should be propagated without change in form and received at a distance as the same wave form. Consequently in media strictly obeying the elastic equations a seismogram should consist of a succession of knife-sharp disturbances, due to waves which have traveled different paths by refractions and reflections. Ricker goes on to state that if such a sharp and clear-cut series of impulses did constitute a seismogram many of the difficulties in seismic prospecting would disappear. As we know, however, no such simple seismogram is received in the propagation of seismic waves through the earth. Instead he points out that we obtain more complicated seismograms, which are familiar to every geophysicist.

In order to explain this complicated nature of a seismogram, Ricker proposes his wavelet theory of seismogram structure. The reader is referred to Ricker's work in which he demonstrates mathematically and experimentally that a sharp seismic disturbance, or impulse, gives rise to a traveling wavelet, the shape of which is determined by the nature of the absorption spectrum of the earth for elastic waves. The shape of this wavelet, which is a time function, is the response of the earth to the sharp seismic disturbance, or impulse. A seismogram, then, consists of many of these wavelets, with

different strengths and arrival times, due to disturbances which have traveled different paths by refractions and reflections.

6.2 The Statistical Determination of Ricker Wavelets

Thus the seismogram may be visualized as the totality of responses to impulses, each impulse being associated with a disturbance which has traveled a certain path by refractions and reflections. These responses, or response functions, are the seismic wavelets or Ricker wavelets. The analysis of a seismogram consists in breaking down this elaborate wavelet complex into its component wavelets. In particular we desire the arrival times of the theoretical sharp impulses which produce these wavelets or responses.

There are two basic approaches which one may use toward the solution of this problem, the deterministic approach and the probabilistic or statistical approach. In the deterministic approach one utilizes basic physical laws, for example, in order to determine the shape of the wavelet, or the adsorption spectrum of the earth. At all stages in such an investigation, one may compare mathematical results with direct and indirect observation of the physical phenomenon.

In this thesis we are concerned with the statistical approach. Such an approach in no way conflicts with the

deterministic approach, although each approach has certain advantages and disadvantages which do not necessarily coincide. The emphasis we place on the probabilistic approach is due to its being the subject of investigation of this thesis. In practice the two approaches may be utilized in such a manner so as to compliment each other.

Let us apply the probabilistic approach to one specific problem, which, as we shall see, is a problem treated by Wadsworth, et al, (1935). Let us set up a hypothetical situation. Let us assume that a given section of seismic trace is additively composed of Ricker wavelets, where each wavelet has the same shape or form. We shall assume that the shape of the wavelet is mathematically stable, that is, the discrete representation of the wavelet shape is a solution of a stable difference equation. Further, we assume that from knowledge of the arrival time of one wavelet we cannot predict the arrival time of another wavelet; and, we assume that from knowledge of the strength of one wavelet we cannot predict the strength of another wavelet. Finally, let us assume that the seismic trace is an automatic volume control (AVC) recording so that the strengths of these wavelets have a constant standard deviation (or variance) with time.

The specific problem which we wish to consider is the following: given the seismic trace described in the above

paragraph, determine the arrival times and strengths of the Ricker wavelets, and determine the basic wavelet shape. We shall discuss a theoretical solution of this problem, and shall also discuss a practical solution which involves statistical estimation.

Let us translate our assumptions about the seismic trace into mathematical notation for discrete time t. First let the shape of the fundamental Ricker wavelet be given by the discrete stable time function b_t where $b_t = 0$ for t less than zero. That is, b_0 is the initial (non-zero) amplitude of the wavelet. Discrete stable time functions are discussed in Section 2.6.

Let the strength, or weighting factor, of the wavelet which arrives at time t be given by ξ_t. That is, ξ_t is a constant weighting factor which weights the entire wavelet whose arrival time is time t. The variable ξ_t is the theoretical knife-sharp impulse of which the particular wavelet (i.e. the one which arrives at time t) is the response. For example, if no wavelet arrives at a particular time t, then $\xi_t = 0$.

In our discussion of the nature of the seismic trace, we shall call the knife-sharp impulses ξ_t "random variables". Our use of the term "random variable ξ_t" does not imply that the variable ξ_t is one whose value is uncertain and can be determined by a "chance" experiment. That is, the variable ξ_t

is not random in the sense of the frequency interpretation of probability (Cramér, 1946), but is fixed by the geologic structure. Frechet (1937) describes this type of variable as "nombre certain" and "function certaine" and Neyman (1941) translates these terms by "sure number" and "sure function". Another example of a "sure number" is the ten thousandth digit of the expansion e = 2.71828..., which, although unknown, is a definite fixed number. Since the knowledge of working geophysicist about the entire deterministic setting is far from complete, we shall treat this incomplete knowledge from a statistical point of view. We thus call ξ_t a "random variable", although we keep in mind that it is a "sure number". Further discussions about this general type of problem may be found in the statistical literature with discussions about the theorem of the English clergyman Thomas Bayes and with discussions about statistical estimation (Cramér, 1946; Jeffreys, 1939). The relationship of the use of Bayes' Theorem in statistical estimation to other methods of statistical estimation is discussed by the author (1950).

Without loss of generality, we may center the knife-sharp impulses ξ_t so that their mean $E[\xi_t]$ is equal to zero. Nevertheless the following discussions may be carried out, by some minor modifications, without centering the ξ_t.

Our assumption about the unpredictability of the arrival times and strengths of wavelets means mathematically that the knife-sharp impulses ξ_t are mutually uncorrelated random variables; that is,

$$E[\xi_t \xi_s] = 0, \quad s \neq t. \tag{6.21}$$

An explanation of the expectation symbol E is given in Section 4.2; mutually uncorrelated variables, Section 4.5. Our assumption that the knife-sharp impulses ξ_t are mutually uncorrelated with each other is an orthogonality assumption, and is a weaker assumption than the assumption that the ξ_t are statistically independent, which we need not make.

Returning again, for the moment, to our discussion about the "sure" nature of the knife-sharp impulses ξ_t, we see that the assumption that they are mutually uncorrelated in time and in strength does not hold in completely deterministic system. Nevertheless, such an assumption is a reasonable one again for the working geophysicist whose knowledge of the entire deterministic setting is far from complete, and who is faced with essentially a statistical problem.

In other words, we assume that knowledge of the arrival time and strength of one wavelet does not allow us to predict the arrival time and strength of any other wavelets.

In particular, we assume that an arrival time and magnitude of a reflection from a certain reflecting horizon does not allow us to predict the arrival time and magnitude of a reflection from a deeper reflecting horizon.

The use of AVC recordings means mathematically that the strengths ξ_t have a constant variance, which without loss of generality we shall take to be unity,

$$E[\xi_t^2] = 1. \tag{6.211}$$

Finally since we assume that the seismogram trace x_t is additively composed of wavelets, all with shape b_t, and strengths ξ_t, we may write this wavelet complex mathematically as

$$x_t = \sum_{s=0}^{\infty} b_s \xi_{t-s} \text{ for } t_1 \leq t \leq t_2 \tag{6.22}$$

where the time interval ($t_1 \leq t \leq t_2$) comprises our basic section of seismic trace. This equation includes tails of wavelets with shape b_t, these wavelets being due to knife-sharp impulses ξ_{t_1}, ξ_{t_1-1}, ξ_{t_1-2},... which occur before time t_1. Equation (6.22) is illustrated in Figure 12, in which the top diagram shows the knife-sharp impulses ξ_t, the center diagram shows the Ricker wavelets b_t weighted by these impulses, and the bottom diagram shows the seismic trace x_t

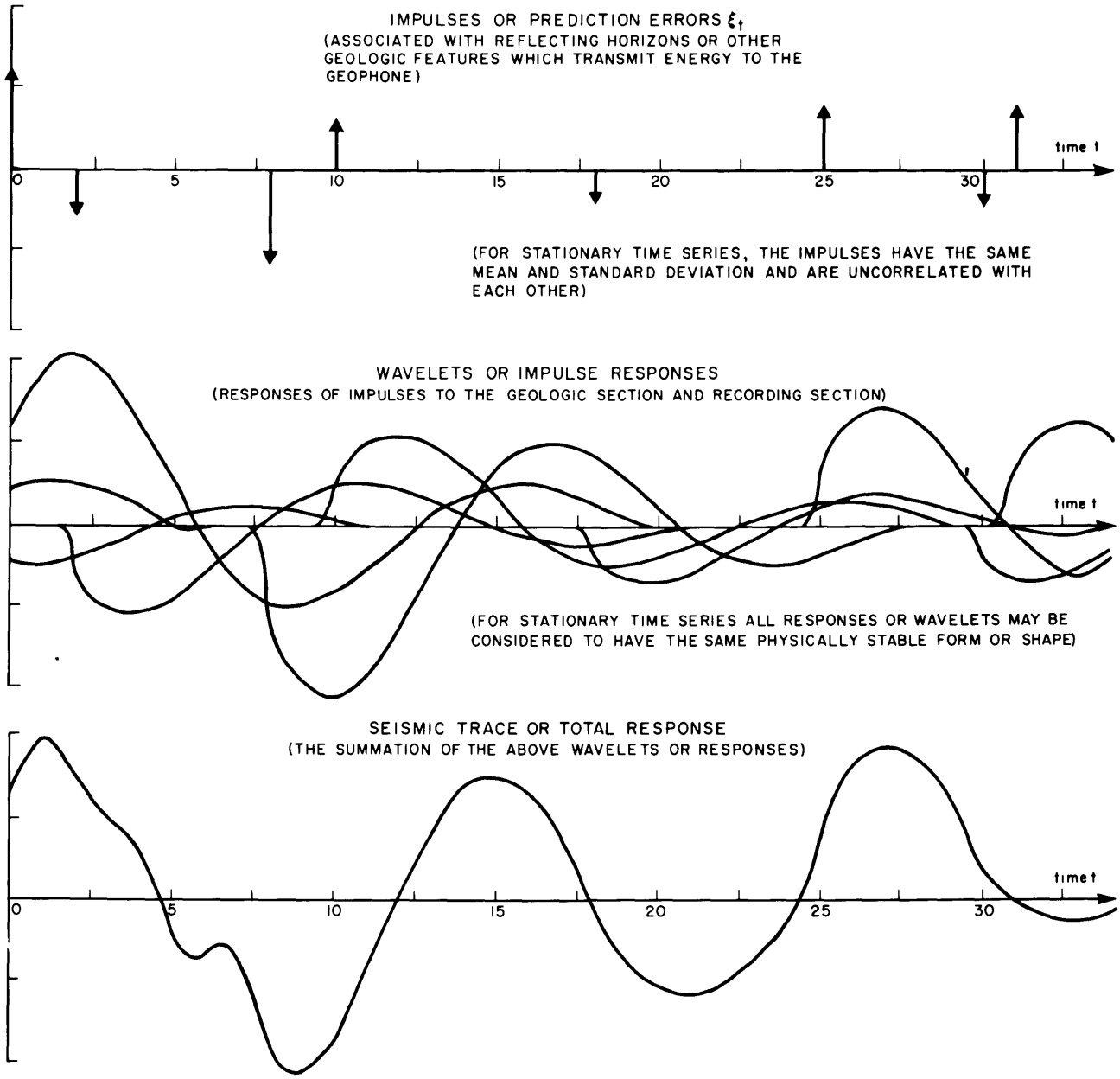


FIGURE 12 THE PREDICTIVE DECOMPOSITION OF A SEISMIC TRACE

which is obtained by adding the wavelets of the center diagram.

For the purposes of our theoretical discussion, let us assume that our assumptions about the time series x_t, equation (6.22), now hold for all time. That is, we consider the mathematical abstraction in which equation (6.22) holds for all t, where ξ_t now represents a stationary mutually uncorrelated process (Section 4.5). Thus equation (6.22) becomes

$$x_t = \sum_{s=o}^{\infty} b_s \xi_{t-s} \quad \text{for } -\infty < t < \infty \,. \qquad (6.221)$$

Equation (6.221) is the mathematical representation of the Predictive Decomposition Theorem of Wold (1938) for a stationary time series with an absolutely continuous spectral distribution function. For further discussion of this theorem, see Section 5.2. Thus the infinite time series x_t given by equation (6.221) is a stationary time series with an absolutely continuous spectral distribution, and the finite time series x_t given by equation (6.22) represents a finite section of the infinite time series (6.221).

In other words, the Predictive Decomposition Theorem states that a stationary time series is the summation of the responses of a stable linear system to impulses ξ_t which have uncorrelated strengths and arrival times. The response to each impulse has the same stable shape or form b_t, and the

variance $E[\xi_t^2]$ of the impulses is constant with time.

In equation (6.221), the wavelet b_t represents the dynamics of the time series, whereas the impulses ξ_t represents the "random" nature of the time series. The basic problem which we wish to consider consists of the separation of the dynamic from the random components of the time series, or seismic trace.

Will the computation of the Fourier transform of the trace effect this separation? The answer is no because it merely transforms time information into equivalent frequency information. As an illustration, let us consider the following example.

To avoid difficulties with end effects, let us assume, for this example, that the wavelet b_t damps sufficiently rapidly so that we may let

$$b_t = 0 \text{ for } t > M. \tag{6.23}$$

Then the predictive decomposition becomes

$$x_t = \sum_{s=o}^{M} b_s \xi_{t-s}. \tag{6.231}$$

Also, for this example, let us assume the trace for $t_1 \leq t \leq t_2$ consists of only those responses to impulses ξ_t which arrive for times $t_1 \leq t \leq t_2 - M$. The Fourier transform of this section of the trace becomes

$$X(\omega) = \sum_{t_1}^{t_2} x_t e^{-i\omega t} = \sum_{t=t_1}^{t_2} \sum_{s=o}^{M} b_s \xi_{t-s}$$

$$= \sum_{s=o}^{M} b_s e^{-i\omega s} \sum_{t=t_1}^{t_2} \xi_{t-s} e^{-i\omega(t-s)}$$

$$= \left(\sum_{s=o}^{M} b_s e^{-i\omega s} \right) \left(\sum_{t=t_1}^{t_2-M} \xi_t e^{-i\omega t} \right)$$

$$= B(\omega)\, I(\omega) \qquad (6.232)$$

where $B(\omega)$ is the Fourier transform, or spectrum, of the Ricker wavelet and $I(\omega)$ is the Fourier transform of a realization of the uncorrelated knife-sharp impulses. Although the Fourier transform $X(\omega)$ contains the dynamic and random elements of a seismic trace, it does not help us to separate the dynamic component $B(\omega)$ from the random component $I(\omega)$ since $X(\omega)$ is the product of the two.

In order to separate the random components ξ_t from the dynamic component b_t of the seismic trace one may use statistical method of averaging. The basic probabilistic approach from a theoretical point of view consists of the

following operations on the mathematical abstraction of the seismic trace, (i.e., the stationary time series x_t, given by equation (6.221)):

(1) Average out the random components ξ_t so as to yield the wavelet shape b_t.

(2) Using the wavelet shape thus found, remove this wavelet shape from the trace, thereby leaving, as a residual, the random components ξ_t (which are the prediction errors for prediction distance $\alpha = 1$).

If one wishes to filter the seismic trace (see Section 5.4) one further step is added, namely:

(3) Reaverage the prediction errors ξ_t by means of a stable linear operator q_t so as to approximate the desired output or message $m_{t+\alpha}$. That is, compute

$$\hat{m}_{t+\alpha} = \sum_{\tau=0}^{\infty} q_{\tau+\alpha} \xi_{t-\tau} \qquad (5.441)$$

which is optimum filtered time series in the sense of least squares. In Section 5.4 we describe how the linear operator q_t is determined from the spectra and cross-spectra of message and noise.

The theoretical procedure for carrying out these operations has been treated in detail in our discussion of stationary time series (with absolutely continuous spectral distri-

butions) in the preceding chapters. Let us review this theoretical procedure for infinite stationary time series.

(1) Compute the autocorrelation function of the time series

$$\phi(\tau) = \lim_{T\to\infty} \frac{1}{2T+1} \sum_{t=-T}^{T} x_t x_{t+\tau} = E[x_t x_{t+\tau}]$$

$$= \sum_{t=o}^{\infty} b_t b_{t+\tau} = b_o b_\tau + b_1 b_{\tau+1} + b_2 b_{\tau+2} \cdots \quad (5.232)$$

This computation averages out the random elements ξ_t and preserves the dynamic elements b_t in the form of the autocorrelation function $\sum_{t=o}^{\infty} b_t b_{t+\tau}$ of the wavelet. That is, the autocorrelation of the time series x_t is the same function as the autocorrelation function of the wavelet b_t.

From this autocorrelation function, compute the shape b_t of the wavelet in the following manner. Take the Fourier transform of the autocorrelation function which yields the power spectrum $\Phi(\omega)$ of the time series x_t, which is also the energy spectrum $|B(\omega)|^2$ of the wavelet b_t; that is

$$\sum_{\tau=-\infty}^{\infty} \phi(\tau)\, e^{-i\omega\tau} = \Phi(\omega) = |B(\omega)|^2 \quad (6.24)$$

where

$$B(\omega) = \sum_{t=o}^{\infty} b_t e^{-i\omega t} \qquad (6.241)$$

is the Fourier transform of the wavelet b_t. Thus we have determined $|B(\omega)|^2$ but not $B(\omega)$. Although there may be many wavelet shapes which yield the energy spectrum $|B(\omega)|^2$, all of these wavelet shapes are unstable, except one. Therefore it is not unreasonable to assume that this unique stable wavelet is the wavelet generated by a physical phenomenon, and that the unstable ones are not. The Fourier transform, $B(\omega)$, of this stable wavelet may then be found by the Wold-Kolmogorov method of factoring the power spectrum (see Section 5.1) expressed by

$$\Phi(\omega) = |B(\omega)|^2 = B(\omega)\ \overline{B(\omega)} \qquad (5.181)$$

where $B(\lambda)$ is required to have no singularities or zeros in the lower half λ plane, where $\lambda = \omega + i\sigma$. In the language of the engineer, $B(\omega)$ is a transfer function with minimum phase-shift characteristic. Having thus determined $B(\omega)$, the stable wavelet (or linear operator) b_t is given by

$$b_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} B(\omega)\ e^{i\omega t}\ d\omega. \qquad (6.242)$$

(2) From this wavelet shape b_t, we find the inverse wavelet shape a_t, where a_t is equal to zero for t less than zero. If we let the b_t represent the coefficients of a linear operator, then the a_t are the coefficients of the inverse linear operator (Section 2.7). Thus the values of a_t are found by

$$a_t = 0 \quad \text{for } t < 0,$$

$$a_o b_o = 1,$$

$$\sum_{s=o}^{t} a_s b_{t-s} = 0 \text{ for } t = 1,2,3\ldots. \tag{2.785}$$

Since the wavelet b_t is stable, the inverse wavelet a_t is also stable. Let $A(\omega)$ be the Fourier transform of a_t, that is

$$A(\omega) = \sum_{s=o}^{\infty} a_s e^{-i\omega s}. \tag{2.791}$$

Then $A(\omega)$ and $B(\omega)$ are related by

$$A(\omega) = \frac{1}{B(\omega)} \tag{2.795}$$

and $A(\omega)$ also has minimum phase-shift characteristic. Swartz and Sokoloff (1953) in their Figures 12 and 13 plot the a_t (in reverse manner) and $|A(\omega)|$ for empirical prediction operators.

The reciprocal of $|A(\omega)|$ then gives $|B(\omega)|$ which is the absolute value of the wavelet spectrum.

We use the inverse wavelet shape a_t to remove the wavelets, which are of shape b_t, from the time series x_t, by compressing the wavelets into the knife-sharp impulses ξ_t. That is, the linear operator a_t is the prediction operator for unit prediction distance, and the prediction errors ξ_t are yielded by the computation

$$\sum_{s=0}^{\infty} a_s x_{t-s}. \tag{6.25}$$

To see that this computation does yield the ξ_t we use the predictive decomposition (6.221) for x_{t-s} and thus obtain

$$\sum_{s=0}^{\infty} a_s x_{t-s} = \sum_{s=0}^{\infty} a_s \sum_{\tau=0}^{\infty} b_\tau \xi_{t-s-\tau}$$

$$= \sum_{s=0}^{\infty} a_s \sum_{n=s}^{\infty} b_{n-s} \xi_{t-n}. \tag{6.251}$$

Recalling that $b_r = 0$ for $r < 0$, and using equation (2.785), we have

$$\sum_{s=0}^{\infty} a_s x_{t-s} = \sum_{n=0}^{\infty} \left(\sum_{s=0}^{\infty} a_s b_{n-s} \right) \xi_{t-n} = \xi_t. \tag{6.252}$$

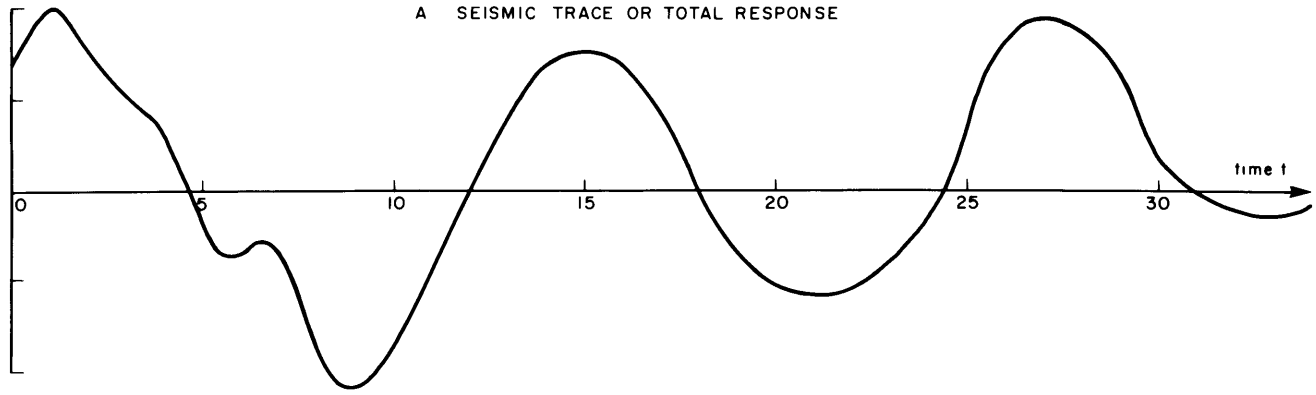
Thus we have

$$\sum_{s=o}^{\infty} a_s x_{t-s} = \xi_t \qquad (6.253)$$

which is the prediction error, or knife-sharp impulse.

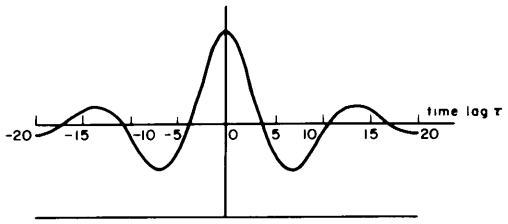
Thus by these theoretical steps we may separate the dynamic component, represented by the response function or wavelet shape b_t, from the random component, represented by the knife-sharp impulses ξ_t which represent arrival times and strengths of the wavelets which comprise the time series. These theoretical steps are illustrated in Figure 13. In this Figure, as in others, we plot the discrete time functions as points and then draw smooth curves through these points.

The practical solution of the problem of separating the dynamic and random components of a finite section of seismic trace involves statistical estimation. One method consists of estimating the prediction operator, or inverse wavelet shape, directly from the finite section of seismic trace. For this purpose one may use the Gauss method of least squares as described in Wadsworth, et al (1953). Since the method described there is more general, let us write down the equations to be used for our specific problem in which we consider only one trace x_t for a prediction distance equal to one.



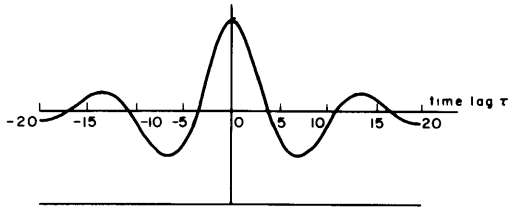
B AUTOCORRELATION FUNCTION OF TRACE

(UNIQUELY DETERMINED FROM INFINITE TIME SERIES, BUT MAY BE ESTIMATED FROM FINITE SECTION OF TRACE)



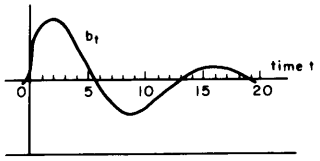
C AUTOCORRELATION FUNCTION OF INDIVIDUAL WAVELET

(SAME AS AUTOCORRELATION FUNCTION OF TRACE)



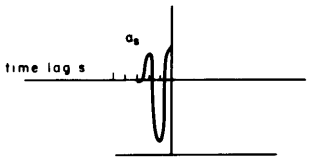
D PHYSICALLY STABLE FORM OR SHAPE OF INDIVIDUAL WAVELET

(UNIQUELY DETERMINED FROM WAVELET AUTOCORRELATION BY METHOD OF THE FACTORIZATION OF THE SPECTRUM)



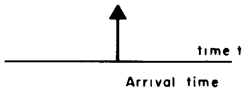
E STABLE PREDICTION OPERATOR WITH MINIMUM PHASE CHARACTERISTIC

(UNIQUELY DETERMINED FROM WAVELET FORM)



F IMPULSE OR PREDICTION ERROR

YIELDED BY PREDICTION OPERATOR ACTING ON WAVELET THIS IMPULSE OCCURS AT THE ARRIVAL TIME OF WAVELET



G SERIES OF IMPULSES OR PREDICTION ERRORS

YIELDED BY PREDICTION OPERATOR ACTING ON SEISMIC TRACE (WHICH IS A SUM OF WAVELETS, WEIGHTED BY THE IMPULSE STRENGTHS) THE PREDICTION ERRORS OR IMPULSES ξ_t OCCUR AT THE ARRIVAL TIMES OF THE WAVELETS

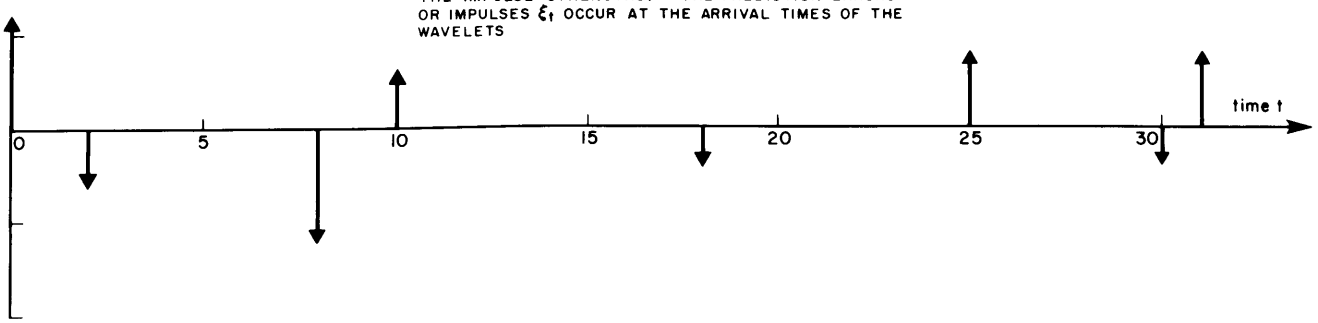


FIGURE 13 ANALYSIS OF A SEISMIC TRACE BY A PREDICTION OPERATOR

Let us note these differences in notation: the prediction distance k of Wadsworth, et al (1953) is our α; the operator coefficients a_s(s=0,1,...M) of Wadsworth, et al, are respectively our operator coefficients k_s (s=0,1,...M) of equation (2.21); and the operator coefficients b_s for the y-trace of Wadsworth, et al, are not our operator coefficients b_s in this thesis. In other words, our use of the symbols a_s and b_s is different from the use of the symbols a_s and b_s in Wadsworth, et al (1953).

Then utilizing our notation, equation (37) of Wadsworth, et al (1953) becomes for the special case of our problem:

$$\hat{x}_{t+1} = c + \sum_{s=o}^{M} k_s \; x_{t-s} \qquad (6.26)$$

(37)

where we use the same notation except for the differences we have just noted. According to our convention, we have let the spacing h=1, so that the running index t is the same as the running index i. Equation (6.26) is equation (2.21) with α=1, except that a constant c also appears in equation (6.26) to take account of the mean value of the time series, since now we do not require the mean to be zero.

The operator time interval (Wadsworth, et al, 1953) is chosen to be the time interval of the section of our hypo-

thetical trace which we assume to be a section of a stationary time series. The normal equations (40) become

(40)

$$cn + \sum_s k_s \sum_t x_{t-s} = \sum_t x_{t+1}$$

$$c \sum_t x_{t-r} + \sum_s \left(k_s \sum_t x_{t-r}\, x_{t-s}\right) = \sum_t x_{t-r}\, x_{t+1}, \text{ for } r=0,1,\ldots M \qquad (6.261)$$

where, as in Wadsworth, et al (1953), summations on the index t are for t = N-1 to t = N+n-2 since α=1, and all summations on the index s are for s=0 to s=M. The solution of the normal equations (6.261) yields the operator coefficients c, k_o, $k_1, \ldots k_M$. Then the inverse wavelet shape, a_t, is given by equation (2.281) which is

$$a_o = 1,\ a_1 = -k_o,\ a_2 = -k_1, \ldots,\ a_m = -k_M \qquad (2.281)$$

with m = M+1. As we have noted in Section 2.2, although both a_t and k_t represent the coefficients of the same operator, as seen by equation (2.281), we call a_t the standard

form of the prediction operator. Also as we have noted, our k_o, $k_1, k_2, \ldots k_M$ in equation (2.281) are respectively the $a_o, a_1, a_2, \ldots a_M$ of Wadsworth, et al, (1953). The constant c of equation (6.26), which adjusts for the mean value of the empirical trace, is not used in determining the shape a_t of the wavelet. Since $a_o=1$, the inverse wavelet a_t may be called a "unit" inverse wavelet.

The shape b_t of the Ricker wavelet may be readily computed by means of equations (2.785), which may be rewritten in terms of the prediction operator k_s of equation (6.26) as

$$b_t = 0 \text{ for } t < 0$$

$$b_o = 1$$

$$b_{t+1} = -\sum_{s=1}^{m} a_s\, b_{t+1-s} = \sum_{s=o}^{M} k_s b_{t-s}, \text{ for } t > 0. \qquad (6.262)$$

That is, the wavelet shape b_t for $t > 0$ is determined by successive step-by-step predictions from its past values, where we let the initial values be $b_t = 0$ for $t < 0$ and $b_o = 1$.

As we have seen, the Gauss method of least squares described yields an empirical estimate of the theoretical prediction operator, or inverse wavelet shape. This empirical estimate has certain optimum statistical properties under general conditions. For a treatment of the optimum properties of linear least-squares estimates, see Wold (1953).

A good estimate of the prediction operator should yield prediction errors which are not significantly autocorrelated.

In other words, the prediction errors ξ_t should be mutually uncorrelated at some preassigned level of significance. Let it be noted that we are confining our attention to the hypothetical section of the trace which we assumed to be a section of a stationary time series; that is, we are dealing with the prediction errors in the so-called operator time interval. For example, if the prediction errors are significantly autocorrelated, more coefficients may be required in the empirical prediction operator.

In Figure 16, in the left hand diagram, we show the prediction operator a_s computed for trace N650 for the time interval 0.350 seconds to 0.475 seconds on MIT Record No. 1 (supplied by the Magnolia Petroleum Co.). This seismogram is illustrated and described in Wadsworth, et al (1953). In the computation of this inverse wavelet shape, a_s, we used equations (6.261) to find the k_s, and then used equations (2.281) to find the a_s. In the right hand diagram of Figure 16, we show the inverse prediction operator, which is the shape of the Ricker wavelet b_t. The shape of the Ricker wavelet was "predicted" by means of equation (6.262). In these computations, we used discrete time series where the spacing h = 2.5 milliseconds. In plotting a_s and b_t we followed our usual procedure which is to plot discrete time

functions, such as a_s and b_t, as discrete points, and then to draw a smooth curve through these discrete points. Also in Figure 16, the time axes are shifted by one discrete time unit (which is 2.5 milliseconds), which is not the convention we have used in our other Figures. Thus in Figure 16, $a_{-1} = 0$ is plotted at time lag $s = 0$, and $a_0 = 1$ is plotted at time lag $s = 0.0025$ seconds. Similarly $b_{-1} = 0$ is plotted at time $t = 0.0000$ seconds. As is our usual convention, the prediction operator a_s is plotted in the reverse manner, as described by Swartz and Sokoloff (1954); that is, the time lag s runs in the positive direction toward the left. Swartz and Sokoloff (1954) also describe the filtering action of discrete linear operators, and their relation to the continuous response functions of electric filters.

Here we have described a statistical method to determine the shape of a Ricker wavelet. Alternatively, from other considerations, one may know the shape of the seismic wavelet. Then the prediction operator, or inverse wavelet shape, may be computed by means of equations (2.785).

So far we have confined ourselves to a section of seismic trace which we assume to be approximately stationary. The prediction operator transforms this section of trace into the uncorrelated prediction errors ξ_t, the mean square value of

which is a minimum. As we have seen the operator cannot predict from past values of the trace the initial arrival of a new wavelet, and thus a prediction error ξ_t is introduced at the arrival time of each wavelet. Nevertheless, for times subsequent to the arrival time of the wavelet, the prediction operator which is the inverse to the wavelet can perfectly predict this wavelet, thereby yielding zero error of prediction.

Nevertheless, a seismic trace is not made up of wavelets which have exactly the same form and which differ only in amplitudes and arrival times. Thus if a prediction operator, which is the unique inverse of a certain wavelet shape, encounters a different wavelet shape, the prediction error will no longer be an impulse, but instead will be a transient time function. Thus the prediction errors yielded by this prediction operator acting on a time series additively composed of wavelets of different shapes will not have a minimum mean square value. Since reflected wavelets in many cases have different shapes than the wavelets comprising the seismic trace in a given non-reflection interval, a prediction operator determined from this non-reflection interval will yield high errors of prediction at such reflections. Such a procedure provides a method for the detection of reflections, (Wadsworth,

et al, 1953). In Figures 14 and 15, running averages of the squared prediction errors are plotted. The peaks on these prediction error curves indicate reflections on the seismogram. Since two-trace operators were used, the empirical coherency existing between the two traces was utilized in the determination of these prediction errors. The arrow indicates the operator time interval. Further description of these Figures is given in MIT GAG Report No. 6. Since only the information existing in the operator time interval is utilized in the determination of linear operators by this method, one may expect greater resolving power if more information on the seismogram is utilized in the determination of various other types of operators. Further research on this general subject is now being carried out by the Geophysical Analysis Group.

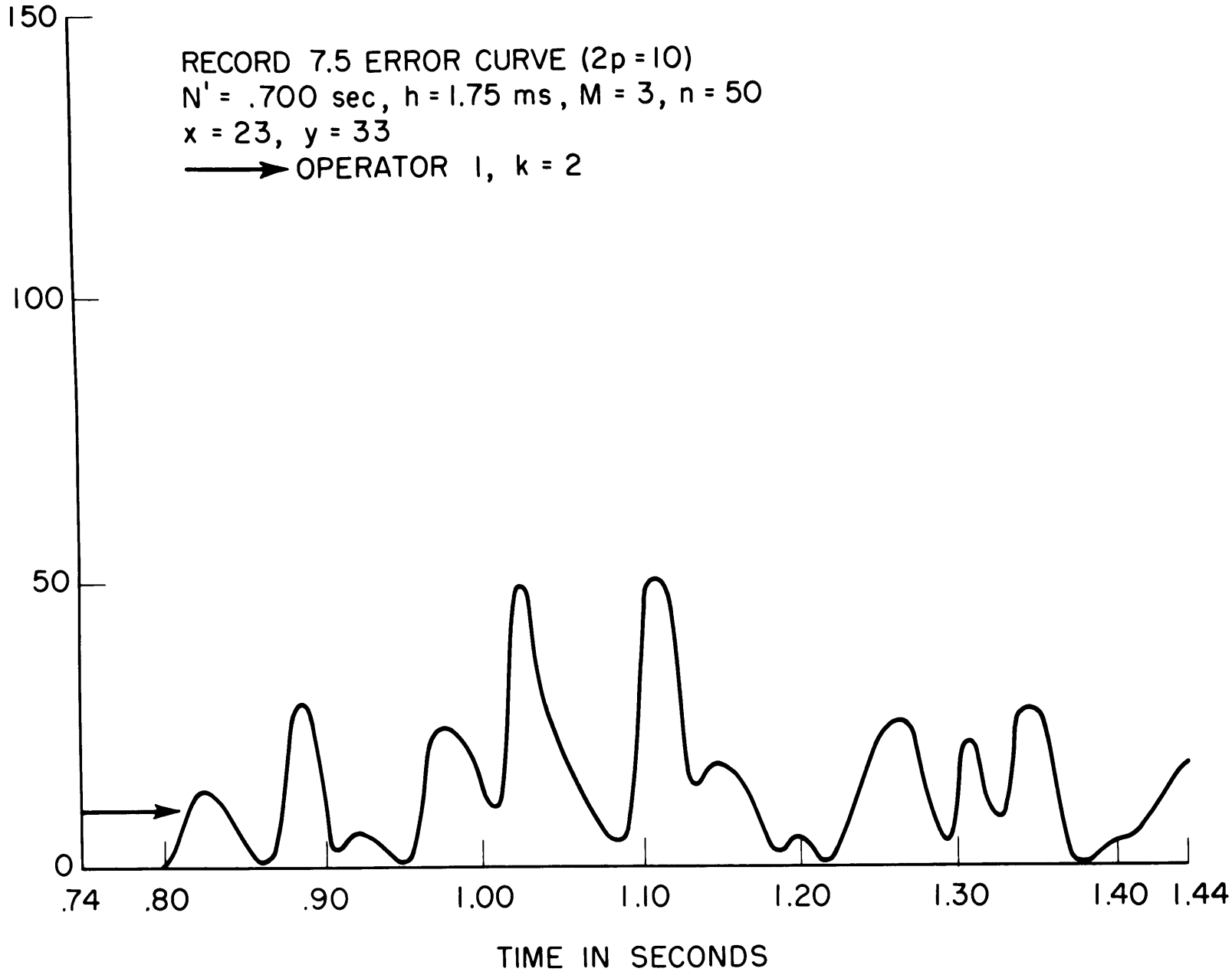


FIGURE 14 PREDICTION ERROR CURVE

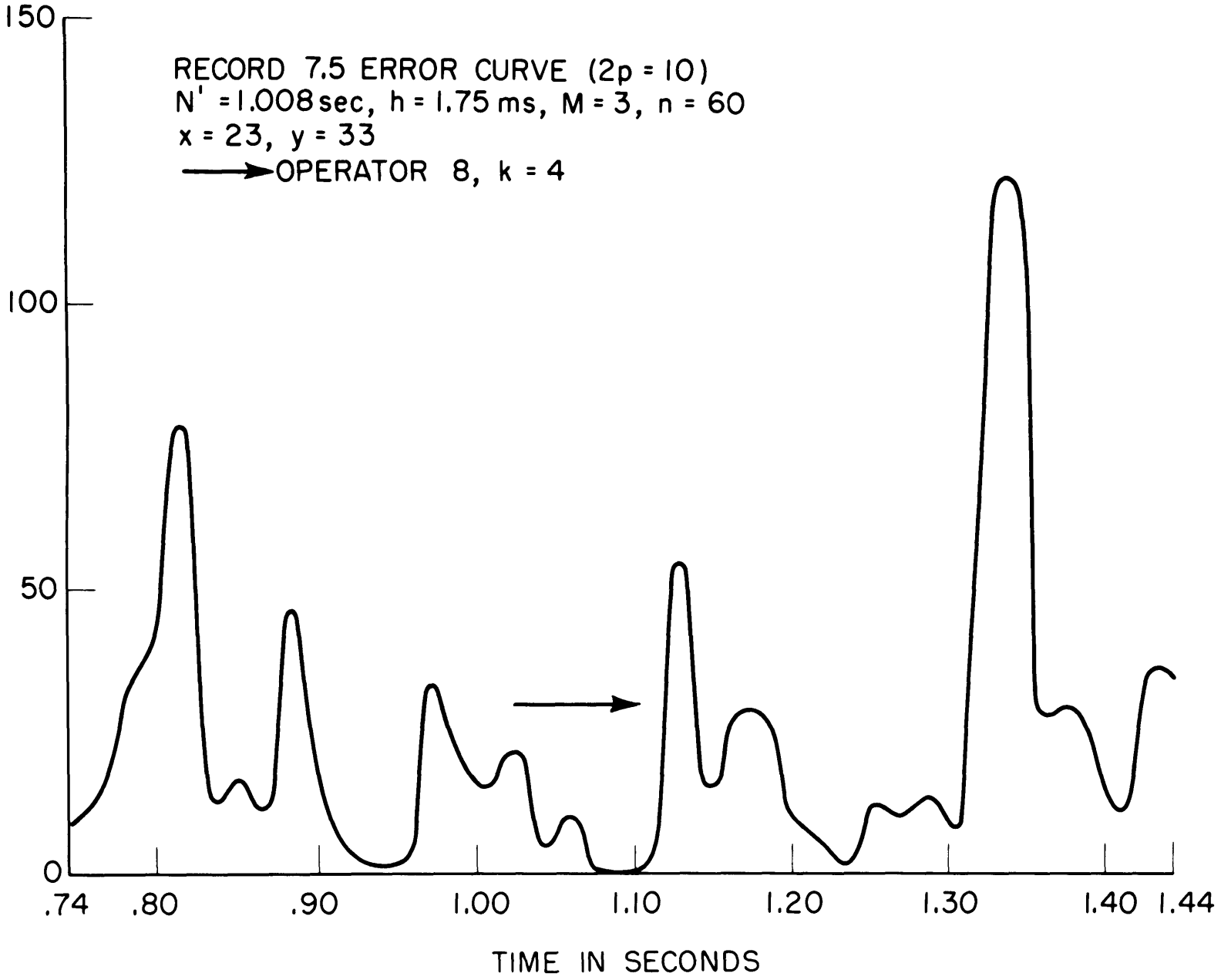


FIGURE 15 PREDICTION ERROR CURVE

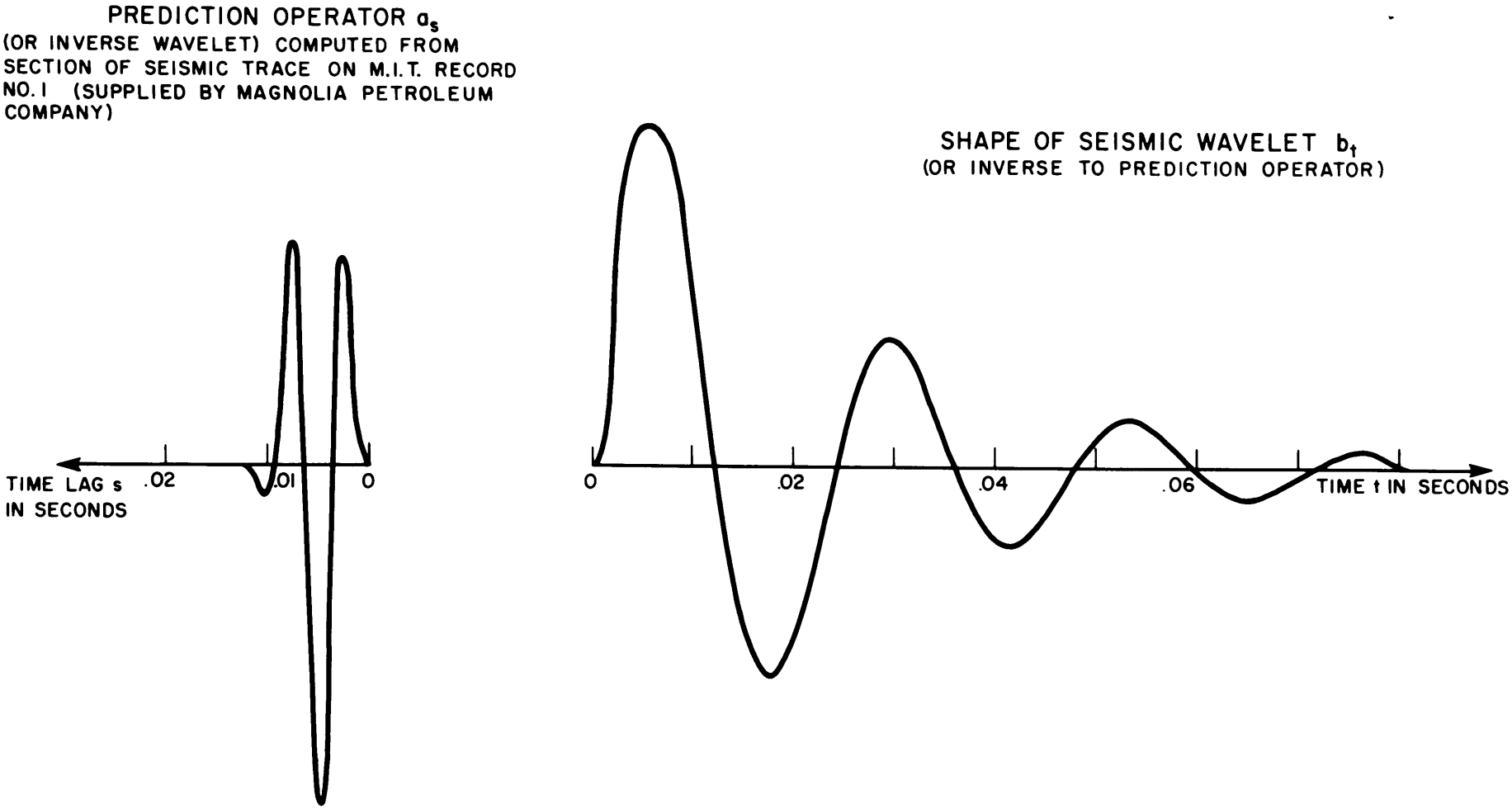


FIGURE 16 COMPUTATION OF WAVELET FROM SECTION OF TRACE ON M.I.T. RECORD NO. 1 (SUPPLIED BY THE MAGNOLIA PETROLEUM COMPANY)

REFERENCES

Aitken, A.C., 1935, On least squares and linear combination of observations: Proc. Roy. Soc. Edinburgh, v. 55, 42-48.

Bartlett, M.S., 1936, The information available in small samples, Proc. Camb. Phil. Soc., 32, p 560.

Bartlett, M., 1946, On the theoretical specification and sampling properties of autocorrelated time-series: J. Roy. Stat. Soc. (B), v.8, 27-41.

Bode, H.W., and Shannon, C.E., 1950, A simplified derivation of linear least square smoothing and prediction theory: Proc. Inst. Rad. Eng., v. 38, 417-425.

Booton, R.C., 1952, Nonstationary statistical theory associated with time varying linear systems: ScD Thesis, MIT.

Bullen, K.E., 1953, An introduction to the theory of seismology, (2nd ed), Camb. Univ. Press.

Champernowne, D.G., 1948, Sampling theory applied to autoregressive sequences: J. Roy. Stat. Soc., B10, 204-231 (-242 discussion).

Cochrane, D., and Orcutt, G.H., 1949, Application of least-squares regression to relationships containing autocorrelated error terms: J. Am. Stat. Assoc., v. 44, 32-61.

Cohen, R., 1948, Some analytical and practical aspects of Weiner's theory of prediction: Res. Lab. Electronics (MIT), Technical Report No. 69.

Cramér, H., 1940, On the theory of stationary random processes: Ann. Math, 41, 215-230.

_____, 1942, On harmonic analysis in certain functional spaces: Ark. Mat. Astr. Fys. 28B, no. 12, 17pp.

_____, 1945, Mathematical methods of statistics: Uppsala, Almqvist and Wiksells, (Parrallel ed. 1946; Univ. Press, Princeton, 575 pp.)

_____, 1947, Problems in probability theory: Annals Math. Stat., v. 18, 165-193.

_____, 1951, A contribution to the theory of stochastic processes: Proc. Sec. Berkeley Symp. Math. Statistics and Prob., Berkeley, 329-339.

Davis, R.C., 1952, On the theory of prediction of nonstationary stochastic processes: J. Appl. Phys. v. 23, p 1047.

Dobrin, M.B., 1952, Introduction to geophysical prospecting, McGraw-Hill Book Company, Inc.

Doob, J.L,, 1949, Time series and harmonic analysis: Proc. Berkeley Symp. Math. Statistics and Prob. Berkeley, 303-343.

_____, 1953, Stochastic processes: New York, John Wiley and Sons, Inc.

Feller, W., 1950, An introduction to probability theory and its applications: New York, Vol.1.

Fisher, R.A., 1921, On the mathe. foundations of theoretical statistics, Phil. Trans. Roy. Soc. 222-309 pp.

_____, 1925, Statistical methods for research workers: (11th ed. 1950) Oliver and Boyd, Edinburgh, 354 pp.

_____, 1935, The design of experiments: (4th ed. 1947) Edinburgh, Oliver and Boyd, 260 pp.

Frank, H.R., and Doty, W.E.M., 1953, Signal-to-noise ratio improvements by filtering and mixing: Geophysics, p 587.

Frechet, M., 1937, Recherches theoriques modernes sur la theorie des probabilites: Paris, Gauthier-Villars.

Goldman, Stanford, 1953, Information theory, New York, Prentice-Hall, Inc.

Hamerle, J.F., 1951, Linear prediction of discrete stationary time series, S.M. Thesis MIT.

Hardy, G.H., Littlewood, J.E., and Pólya, G., 1934, Inequalities: Cambridge, Univ. Press, 314 pp.

Heiland, C.A., 1940, Geophysical exploration, Prentice-Hall.

Jakosky, I.J., 1950, Exploration geophysics, Trija Publishing Co.

Jeffreys, H. , 1939, Theory of Probability, Oxford.

_____, 1952, The Earth,(Third Ed.) Camb. Univ. Press.

Kac, M. and Siegert, A.J.F., 1947a, An explicit representation of a stationary Gaussian process; Ann. Math. Stat., v. 18, 438-442.

_____, 1947b, On the theory of noise in radio receivers with square law detectors, J. Appl. Phys., v. 18, 383-397.

Karhunen, Kari, 1947, Uber lineare Methoden in der Wahrscheinlichkeitsrechnung: Ann. Acad. Sci. Fennicae Ser. A, I. Math. Phys. 37, 79 pp.

Kendall, M., 1946, The advanced theory of statistics, v. II, London, Griffin, 521 pp.

_____, 1946, Contributions to the study of oscillatory time-series, Natl. Inst. of Ec. of Soc. Research. Occasional Papers IX, Cambridge Univ. Press.

Khintchine, A. 1933, Asymptotische Gessetze der Wahrscheinlich-Keitsrechung, Ergebnisse der Mathe. und ihrer Grenzgebiete 2 no. 4.

_____, 1934, Korrelationstheorie der stationaren stochastischen Prozassa, Mathematiskhe Annalen 109, 604-615.

Kolmogorov, A., 1933, Grundbegriffe der Wahrscheinlichkeitsrechnung: Ergebnisse der Mathe. und ihrer Grenzyebiete, v. 2,3, (also separately at Springer, Berlin. Amer. Ed. 1950, Chelsea, New York) 74 pp.

_____, 1939, Sur l'interpolation et extrapolation des suites stationnaires: C.R. Acad. Sci., v. 208, Paris, 2043-2045.

_____, 1941, Interpolation und Extrapolation von stationaren zufalligen Folgen: Bull. Acad. Sci. U.R.S.S. Ser. Math., V. 5, 3-14.

Koopmans, T.C., ed. 1950, Statistical inference in dynamic economic models. Cowles Commission Monograph no. 10, Wiley, New York, 438 pp.

_____, 1951, Activity analysis of production and allocation: New York, Wiley, 404 pp. Cowles commission Monograph No. 13 Wiley, New York 404 pp.

Lawson, J.E., and Uhlenbeck, G.E., 1950, Threshold signals: New York, McGraw Hill.

Lee, Y. W., 1949, Communication applications of correlation analysis, Woods Hole Symp. Applic. Autocorr. Analy. Phys. Prob. 4-23, Office of Naval Research, NAVEXOS-P-735.

_____, 1950, Application of statistical methods to communication problems: Res. Lab. Electronics (MIT), Technical Report No. 181.

Lee, Y.W., and Stutt, C.A., 1949, Statistical prediction of noise: Mass. Inst. of Tech., Res. Lab. Electronics, Tech. Rpt. 129.

Levinson, Norman, 1947a, The Wiener RMS (root mean square) error criterion in filter design and prediction: J. Math. Phys. (MIT) v. 25, 261-278.

_____, 1947b, A heuristic exposition of Wiener's mathematical theory of prediction and filtering: J. Math. Phys., v. 26, 110-119.

Loeve, Michel, 1946, Quelques proprietes des fonctions aleatoires de second ordre: C.R. Acad. Sci. Paris, v. 222, 469-470.

Middleton, D., 1953, Statistical criteria for the detection of pulsed carriers in noise I, II: J. Appl. Phys. v 24, p 371-391.

MIT Geophysical Analysis Group, Wadsworth, G.P., and Robinson, E.A., Bryan, J. G., and Hurley, P. M., Oct. 1, 1952, Results of an auto-correlation and cross-correlation analysis of seismic records. Presnted at MIT Industrial Liaison Conference of August 6, 1952. MIT GAG Report No. 1 (Reprinted in Wadsworth, et al (1953)).

_____, Nov. 5, 1952, A prospectus on the application of linear operators to seismology, MIT GAG Report No. 2. (Reprinted in Wadsworth, et al (1953)).

MIT Geophysical Analysis Group, Robinson, E.A., Director, Simpson, S.M., and Smith, M.K., (July 8, 1953), Case study of Henderson County seismic record, Part I, MIT GAG Report No. 3.

_____, (July 21, 1953) Linear operator study of a Texas Company seismic profile, Part I, MIT GAG Report No. 4.

_____, (August 4, 1953), On the theory and practice of linear operators in seismic analysis, MIT GAG Report No. 5.

_____, (March 10, 1954), Further research on linear operators in seismic analysis, MIT GAG Report No. 6.

Nettleton, L. L., 1940, Geophysical prospecting for Oil; McGraw Hill.

Neyman, J., 1941, Fiducial argument and the theory of confidence: Biometrika, v. 32, p.128.

Neyman, J., and Pearson, E.S., 1933, On the problem of the most efficient tests of statistical hypotheses: Phil. Trans. Roy. Soc. (A) 231, 289-338.

_____, 1936, Sufficient statistics and uniformly most powerful tests of statistical hypothesis, Statist. Research Mem., 1, 113 pp.

Paley, R.E.A.C., and Wiener, N., 1934, Fourier transforms in the complex domain, American Math. Soc. Colloquium, Publication Vol. 19.

Piety, R. G., 1942, Interpretation of the transient behavior of the reflection seismograph: Geophysics, v.7, p 123-132.

Quenouille, M.H., 1947, A large-sample test for the goodness of fit of autoregressive schemes: JRSS (A) 110, 123-129.

Rice, S.O., 1944, Mathematical analysis of random noise: Bell Sys. Tech. Jour., v. 23, 282-332.

_____, 1945, Mathematical analysis of random noise: Bell Sys. Tech. Jour. v.24, 46-156.

Ricker, N., 1940, The form and nature of seismic waves and the structure of seismograms: Geophysics, v.5, No. 4, 348-366.

_____, 1941, A note on the determination of the viscosity of shale from the measurement of wavelet breadth: Geophysics, v. 6, No. 3, 254-258.

_____, 1943, Further developement in the wavelet theory of seismogram structue, Bull. of the Seis. Soc. America, Vol. 33, No. 3, 197-228.

_____, 1945, The computation of output disturbances from amplifiers for true wavelet inputs, Geophysics, v. 10, No. 2, 207-220.

_____, 1949, Attenuation and amplitudes of seismic waves, Tran. Amer. Geophy. Union, v. 30, No. 2, 184-186.

_____, 1953a, The form and laws of propagation of seismic wavelet, Geophysics, v. 18, No. 1, 10-40.

_____, 1953b, Wavelet contraction, wavelet expansion, and the control of seismic resolution, Geophysics, v. 18, No. 4, 769-792.

Robinson, E.A., 1950, The theories of fiducial limits and confidence limits: S.B. Thesis, MIT.

_____, 1952, Eight new applications of the theory of stochastic processes to economic analysis: S.M. Thesis, MIT.

Samuelson, P.A., 1941a,The stability of equilibrium: Comparative statics and dynamics: Econometrica 9, 97-120.

_____ , 1941b, Conditions that the roots of a polynomial be less than unity in absolute value: Annals of Mathematical Statistics, v. 12, 360-366.

_____, 1941c, A method of determining explicitly the coefficients of the characteristic equation: Annals of Mathematical Statistics, v. 12, 424-429.

_____, 1947, Foundations of economic analysis, Cambridge Harvard Univ. Press, 447 pp.

Seiwell, H.R. and Wadsworth, G.P., 1949, A new development in ocean wave research: Science, v.109, 271-274.

Simpson, S.M., 1953, Statistical approaches to certain problems in geophysics: Ph.D. Thesis, MIT.

Slutsky, E., 1927, The summation of random causes as the source of cyclic processes. (Russion with English summary.) Problems of Economic Conditions (Inst. Econ. Conjuncture, Moscow), 3. Revised English ed. 1937 in Econometrica 5, 105-146.

_____, 1928, Sur les fonctions eventuelles continues, integrables et derivables dans le sens stochastique: C.R. Acad. Sci. Paris, v. 187, 878-880.

Smith, M.K., 1954, Filter theory of linear operators with seismic applications: Ph.D. Thesis, MIT(Sections to which references are made are reprinted in MIT GAG Report No. 6).

Swartz, C.A., 1954, Some geometrical properties of residual maps: Geophysics, v. 19, p46-70.

Swartz, C.A. and Sokoloff, V.M. 1954, Filtering associated with selective sampling of geophysical data: Geophysics, v. 19, pp 402-419.

Tintner, G., 1952, Econometrics, New York, Wiley, 370 pp.

Tukey, John W., 1949, The sampling theory of power spectrum estimates, Woods Hole Symp Applic Autocorr Phys Prob., 47-67, Office of Naval Research - NAVEXOS-P-735.

Tukey, J.W. and Hamming, R.W., 1949, Measuring noise color, I, Bell Telephone Laboratories Memo. MM-49-110-119.

vonNeumann, John, 1941, Distribution of the mean square succession difference to the variance, Ann. Math. Stat. 12, 367-395.

vonNeumann, J., and Morgenstern, O., 1944, Theory of games and economic behaviour, (2nd ed. 1947) Princeton, Univ. Press, 625 pp.

Wadsworth, G.P., Robinson, E.A., Bryan, J.G., and Hurley, P.M., 1953, Detection of reflections on seismic records by linear operators, Geophysics, v 18, pp 539-586.

Wadsworth, G.P., 1947, Short range and extended forecasting by statistical methods, Air Weather Service Tech. Rept. 105-37.

Whittaker, E.T., and Robinson, G. 1924, The calculus of observations: (4th ed. 1944) London, Blackie, 397 pp.

Wiener, N., 1923, Differential space: J. Math. Phys. Math. Inst. Tech., v.2, 131-174.

_____, 1930, Generalized harmonic analysis: Acta Math. v. 55, 117-258.

_____, 1933, The Fourier integral and certain of its applications, Cambridge University Press.

_____, 1942, The extraploation, interpolation, and smoothing of stationary time series with engineering applications: Cambridge, MIT DIC Contract 6037, National Defense Research Council (Section D2)(Reprinted in 1949, New York, John Wiley and Sons, Inc. to which references to page numbers refer).

_____, 1948, Cybernetics: New York, John Wiley and Sons, Inc.

Wold, H., 1938, A study in the analysis of stationary time series: Uppsala. Almqvist and Wiksells, 214 pp.

_____, 1948, On prediction in stationary time series: Annals of Math Statistics 19, 559-567.

Wold, H., 1953, in association with L. Jureen, Demand Analysis, John Wiley and Sons, New York, 358 pp.

Worzel, J.L., and Ewing, M., 1948, Explosion sounds in shallow water: Geol. Soc. Amer. Memoir 27.

Yule, G.U., 1907, On the theory of correlation for any number of variables treated by a new system of notation: Proc. Roy. Soc. v. 79, 182-193.

_____, 1921, On the time-correlation problem, with especial reference to the variate-difference correlation method. J. Roy. Stat. Soc., v.84, 497-526.

_____, 1926, Why do we sometimes get nonsense-correlations between time-series? J. Roy. Stat. Soc. v. 89, 1-64.

_____, 1927, On a method of investigating periodicities in in disturbed series, with special reference to Wolfer's sunspot numbers. Phil. Trans. Roy. Soc. 226, 267-298.

Zadeh, L.A., 1950, The determination of the impulsive response of variable networks: J. Appl. Phys., v. 21, pp 642-655.

_____, 1953, Optimum nonlinear filters: J. Appl. Phys. v. 24, pp 396-404.

Zadeh, L. A., and Ragazzini, J.R., 1950, An extension of Wiener's theory of prediction, J. Appl. Phys., v. 21, 645-655.

Zygmund, Antoni, 1935, Trigonometrical series: Warsaw-Lwow, 331 pp.

Kosulajeff, P. A., 1941, Sur les problemes d'interpolation et d'estrapolation des suites stationnaires: Comptes rendus de l'academie des sciences de U.R.S.S. v 30, 13-17.

Schuster, A., 1898, On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena: Terrestrial Magnetism, v 3.

_____, 1900, The periodogram of the magnetic declination as obtained from the records of the Greenwich Observatory during the years 1871-1895: Trans Cambridge Phil. Soc. v 18.

Dix, C. H. Seismic prospecting for oil: Harper and Brothers, New York 1952.

ERRATA

Page	Line	Reads	Should Read
32	First equation	α 0, M 0	$\alpha \geqslant 0$, $M \geqslant 0$
32	Second equation	m o, $a_o=1$	$m \geqslant o$, $a_o=1$
32	Third equation	M 0	$M \geqslant 0$
32	Fourth equation	M 0	$M \geqslant 0$
32	Fifth equation	m o, $c_o=1$	$m \geqslant o$, $c_o=1$
36	3 and 4	(Levinson, 1946	(Levinson, 1947a
37	4b	$\xi_t = \sum_{s=o}^{m} a_s x_{1-s}$	$\xi_t = \sum_{s=o}^{m} a_s x_{t-s}$
68	6	$\rho_o+2 \sum_{\tau=o}^{\infty} \rho_\tau \cos \omega\tau$	$\rho_o+2 \sum_{\tau=1}^{\infty} \rho_\tau \cos \omega\tau$
68	9	$\rho_o+2 \sum_{\tau=o}^{N} (1 - \frac{\tau}{N}) \rho_\tau \cos \omega\tau$	$\rho_o+2 \sum_{\tau=1}^{N} (1- \frac{\tau}{N}) \rho_\tau \cos \omega\tau$
88	3	Levinson (1947)	Levinson (1947b)
140	9	eletric	electric
144	6b	representation(5.251)	representation(5.26)
149	1b	$\sum_{n=o}^{\infty} a_m x_{s-n}$	$\sum_{n=o}^{\infty} a_n x_{s-n}$
151	12	$-a_{\alpha+1}x_{t-1}-a_{\alpha=2}x_{t-2}$	$-a_{\alpha+1}x_{t-1}-a_{\alpha+2}x_{t-2}-\cdots$
165	3b	1 s=t	2π, s=t
169	6	$= \sum_{\tau=a}^{\infty} q_\tau \xi_{t-\tau-\alpha}$	$= \sum_{\tau=\alpha}^{\infty} q_\tau \xi_{t-\tau-\alpha}$

ERRATA

Page	Line	Reads	Should Read
169	8	$= \sum_{-\infty}^{\infty} q_\tau^2 + \sum_{-\infty}^{\infty} r_\tau^2$	$= \sum_{-\infty}^{\alpha} q_\tau^2 + \sum_{-\infty}^{\infty} r_\tau^2$
178	1b	for $\|\tau\|$ T	for $\|\tau\| > M$
182	9	Kosulajeff (1940)	Kosulajeff (1941)
185	2b	May	may
194	6	(see Figure 2)	(see Figure 3)
213	1b	Wiener (1942),Section	Wiener (1942,Section
214	4	alogue	analogue
214	3b	b_{-t} $t < 0$	β_{-t} , $t < 0$
214	1b	residual b_{-t}	residual β_{-t}
221	13	prediction and problem	prediction problem
235	10	and recording section	and recording system
238	11b	transform or spectrum,	transform, or spectrum,
239	15	$m_{t+\alpha} = \sum_{\tau=0}^{\infty} q_{\tau+\alpha} \xi_{t-\tau}$	$\hat{m}_{t+\alpha} = \sum_{\tau=0}^{\infty} q_{\tau+\alpha} \xi_{t-\tau}$
241	9	B(ω), or this	B(ω), of this
242	1b	and for empirical	and $\|A(\omega)\|$ for empirical
250	8	t=0.0025 seconds. As	t=0 seconds and b_o=1 is plotted at time t=0.0025 seconds. As

INDEX

Biographical Note: Enders Anthony Robinson

Received a S.B. degree in Mathematics in 1950 and a S.M. degree in Economics in 1952 at the Massachusetts Institute of Technology. During 1950 and 1951, employed as a Research Assistant in the MIT Department of Mathematics, and from 1952 to the present time, employed as a Research Associate in the MIT Department of Geology and Geophysics in charge of operations of the Geophysical Analysis Group.