A LINEAR PREDICTION APPROACH TO TWO-DIMENSIONAL SPECTRAL FACTORIZATION AND SPECTRAL ESTIMATION

by

THOMAS LOUIS MARZETTA

S.B., Massachusetts Institute of Technology
(1972)

M.S., University of Pennsylvania
(1973)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February, 1978

Signature of Author
Department of Electrical Engineering and
Computer Science, February 3, 1978

Certified by .. ..
Thesis Supervisor

Accepted by . ...
Chairman, Departmental Committee
on Graduate Students

A LINEAR PREDICTION APPROACH TO TWO-DIMENSIONAL SPECTRAL FACTORIZATION AND SPECTRAL ESTIMATION

by

THOMAS LOUIS MARZETTA

Submitted to the Department of Electrical Engineering and Computer Science on February 3, 1978, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Abstract

This thesis is concerned with the extension of the theory and computational techniques of time-series linear prediction to two-dimensional (2-D) random processes. 2-D random processes are encountered in image processing, array processing, and generally wherever data is spatially dependent. The fundamental problem of linear prediction is to determine a causal and causally invertible (minimum-phase), linear, shift-invariant whitening filter for a given random process. In some cases, the exact power density spectrum of the process is known (or is assumed to be known) and finding the minimum-phase whitening filter is a deterministic problem. In other cases, only a finite set of samples from the random process is available, and the minimum-phase whitening filter must be estimated. Some potential applications of 2-D linear prediction are Wiener filtering, the design of recursive digital filters, high-resolution spectral estimation, and linear predictive coding of images.

2-D linear prediction has been an active area of research in recent years, but very little progress has been made on the problem. The principal difficulty has been the lack of computationally useful ways to represent 2-D minimum-phase filters.

In this thesis research, a general theory of 2-D linear prediction has been developed. The theory is based on a particular definition for 2-D causality which totally orders the points in the plane. By paying strict attention to the ordering property, all of the major results of 1-D linear prediction theory are extended to the 2-D case.

Among other things, a particular class of 2-D, least-squares, linear, prediction error filters are shown to be minimum-phase, a 2-D version of the Levinson algorithm is derived, and a very simple interpretation for the failure of Shanks' conjecture is obtained.

From a practical standpoint, the most important result of this thesis is a new canonical representation for 2-D minimum-phase filters. The representation is an extension of the reflection coefficient (or partial correlation coefficient) representation for 1-D minimum-phase filters to the 2-D case. It is shown that associated with any 2-D minimum-phase filter, analytic in some neighborhood of the unit circles, is a generally infinite 2-D sequence of numbers, called reflection coefficients, whose magnitudes are less than one, and which decay exponentially to zero away from the origin. Conversely, associated with any such 2-D reflection coefficient sequence is a unique 2-D minimum-phase filter. The 2-D reflection coefficient representation is the basis for a new approach to 2-D linear prediction. An approximate whitening filter is designed in the reflection coefficient domain, by representing it in terms of a finite number of reflection coefficients. The difficult minimum-phase requirement is automatically satisfied if the reflection coefficient magnitudes are constrained to be less than one.

A remaining question is how to choose the reflection coefficients optimally; this question has only been partially addressed. Attention was directed towards one convenient, but generally suboptimal method in which the reflection coefficients are chosen sequentially in a finite raster scan fashion according to a least-squares prediction error criterion. Numerical results are presented for this approach as applied to the spectral factorization problem. The numerical results indicate that, while this suboptimal, sequential algorithm may be useful in some cases, more sophisticated algorithms for choosing the reflection coefficients must be developed if the full potential of the 2-D reflection coefficient representation is to be realized.

Thesis Supervisor: Arthur B. Baggeroer

Title: Associate Professor of Electrical Engineering
Associate Professor of Ocean Engineering

ACKNOWLEDGMENTS

I would like to take this opportunity to express my appreciation to my thesis advisor, Professor Arthur Baggeroer, and to my thesis readers, Professor James McClellan and Professor Alan Willsky. This research could not have been performed without their cooperation. It was Professor Baggeroer who originally suggested that I investigate this research topic; throughout the course of the research he maintained the utmost confidence that I would succeed in shedding light on what proved to be a difficult problem area. I had many useful discussions with Professor Willsky during the earlier stages of the research. Special thanks go to Professor McClellan who was my unofficial thesis advisor during Professor Baggeroer's sabbatical.

The daily contact and technical discussions with Mr. Richard Kline and Dr. Kenneth Theriault were an indispensable part of my graduate education.

I would like to thank Mr. Dave Harris for donating his programming skills to obtain the contour plot and the projection plot displayed in this thesis. Finally, I must mention the superb typing skills of Ms. Joanne Klotz.

This research was supported, in part, by a Vinton Hayes Graduate Fellowship in Communications.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

1.1 One-dimensional Linear Prediction

An important tool in stationary time-series analysis is linear prediction. The basic problem in linear prediction is to determine a causal and causally invertible linear shift-invariant filter that whitens a particular random process. The term "linear prediction" is used because if a causal and causally invertible whitening filter exists, it can be shown to be proportional to the least-squares linear prediction error filter for the present value of the process given the infinite past.

Linear prediction is an essential aspect of a number of different problems including the Wiener filtering problem [1], the problem of designing a stable recursive filter having a prescribed magnitude frequency response [2], the autoregressive (or "maximum entropy") method of spectral estimation [3], and the compression of speech by linear predictive coding [4]. The theory of linear prediction has been applied to the discrete-time Kalman filtering problem (for the case of a stationary signal and noise) to obtain a fast algorithm for solving for the time-varying gain matrix [5]. Linear prediction is closely related to the problem of solving the wave-equation in a nonuniform transmission line [6], [7].

In general there are two classes of linear prediction problems. In one case we are given the actual power density spectrum of the process, and the problem is to compute (or at least to find an approximation to) the causal and causally invertible whitening filter. We refer to this problem as the spectral factorization problem. The classical method of time-series spectral factorization (which is applicable whenever the spectrum is rational and has no poles or zeroes on the unit circle) involves first computing the poles and zeroes of the spectrum, and then representing the whitening filter in terms of the poles and zeroes located inside the unit circle [1].

In the second class of linear prediction problems we are given a finite set of samples from the random process, and we want to estimate the causal and causally invertible whitening filter. A considerable amount of research has been devoted to this problem for the special case where the whitening filter is modeled as a finite-duration impulse response (FIR) filter. We refer to this problem as the autoregressive model fitting problem. In the literature, this is sometimes called all-pole modeling. (A more general problem is concerned with fitting a rational whitening filter model to the data; this is called autoregressive moving-average or pole-zero modeling. Pole-zero modeling has received comparatively little attention in the literature. This is apparently due to the fact that there are no computational techniques for pole-zero modeling which are as effective or as convenient to use as the available methods of all-pole modeling.) The two requirements in autoregressive model fitting are that the FIR filter should closely represent the second-order statistics of the data, and that it should have a causal, stable inverse. (Equivalently, the zeroes of the filter should be inside the unit circle.) The two most popular methods of autoregressive model fitting are the so-called autocorrelation method [3] and the Burg algorithm [3]. Both algorithms are convenient to use, they tend to give good whitening filter estimates, and under certain conditions (which are nearly always attained in practice) the whitening filter estimates are causally invertible.

1.2 Two-dimensional Linear Prediction

Given the success of linear prediction in time-series analysis, it would be desirable to extend it to the analysis of multidimensional random processes, that is, processes parameterized by more than one variable. Multidimensional random processes (also called random fields) occur in image processing as well as radar, sonar, geophysical signal processing, and in general, in any situation where data is sampled spatially.

In this thesis we will be working with the class of two-dimensional (2-D) wide-sense stationary, scalar-valued random processes, denoted $x(k,\ell)$ where k and ℓ are integers. The basic 2-D linear prediction problem is similar to the 1-D problem: for a particular 2-D process, determine a causal and causally invertible linear shift-invariant whitening filter.

While many results in 1-D random process theory are easily extended to the 2-D case, the theory of 1-D linear prediction has been extremely difficult, if not impossible, to extend to the 2-D case. Despite the efforts of many researchers, very little progress has been made towards developing a useful theory of 2-D linear prediction. What has been lacking is a computationally useful way to represent 2-D causal and causally invertible filters.

Our contribution in this thesis is to extend virtually all of the known 1-D linear prediction theory to the 2-D case. We succeed in this by paying strict attention to the ordering properties of points in the plane.

From a practical standpoint, our most important result is a new canonical representation for 2-D causal and causally invertible linear, shift-invariant filters. We use this representation as the basis for new algorithms for 2-D spectral factorization and autoregressive model fitting.

1.3 Two-dimensional Causal Filters

We define a 2-D causal, linear, shift-invariant filter to be one whose unit sample response has the support illustrated in Fig. 1.1. (In the literature, such filters have been called "one-sided filters" and "non-symmetric half-plane filters," and the term "causal filter" has usually been reserved for the less-general class of quarter-plane filters. But there is no universally accepted terminology, and throughout this thesis we use our own carefully defined terminology.) The motivation for this definition of 2-D causality is that it leads to significant theoretical and practical results. We emphasize that the usefulness of the definition is independent of any physical properties of the 2-D random process under consideration. This same statement also applies, although to a lesser extent, to the 1-D notion of causality; often a 1-D causal recursive digital filter is used, not because its structure conforms to a physical notion of causality, but because of the computational efficiency of the recursive structure.

The intuitive idea of a causal filter is that the output of the filter at any point should only depend on the present and past values of the input. Equivalently the unit sample response of the filter must vanish at all points occurring in the past of the origin. Corresponding to our definition of 2-D causality is the definition of

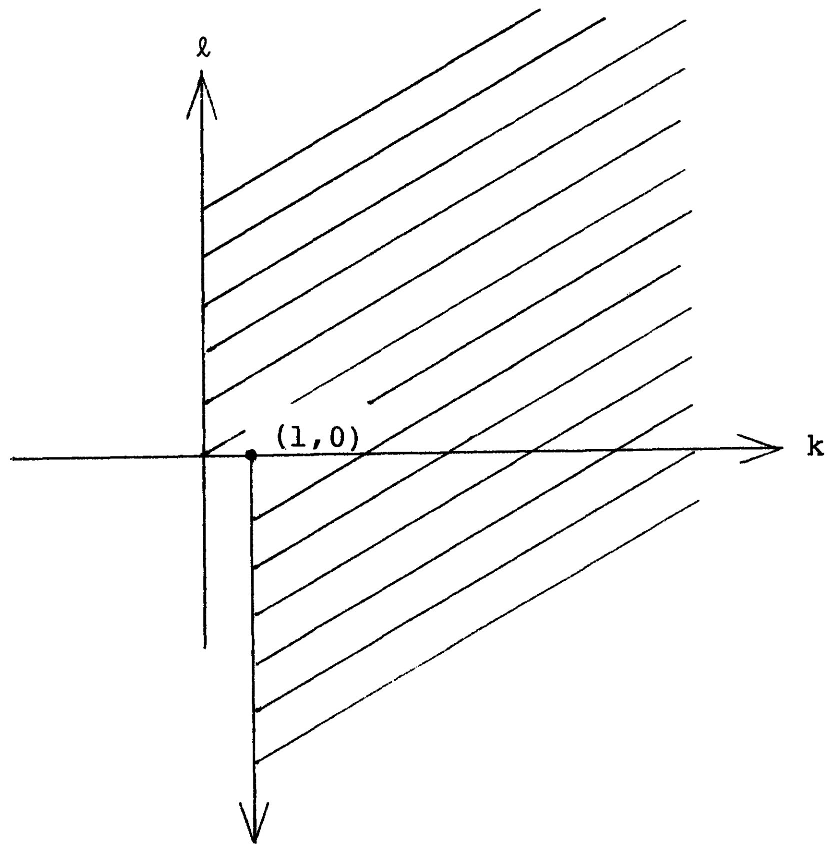


Fig. 1.1 Support for the unit sample response of a 2-D causal filter.

"past," "present," and "future" illustrated in Fig. 1.2 This definition of "past," "present," and "future" uniquely orders the points in the 2-D plane, the ordering being in the form of an infinite raster scan. It is this "total ordering" property that makes our definition of 2-D causality a useful one.

1.4 Two-dimensional Spectral Factorization and Autoregressive Model Fitting

As in the 1-D case, the primary 2-D linear prediction problems are 1) The determination (or approximation) of the 2-D causal and causally invertible whitening filter given the power density spectrum (spectral factorization); and 2) The estimation of the 2-D causal and causally invertible whitening filter given a finite set of samples from the random process (for an FIR whitening filter estimate, the autoregressive model fitting problem). Despite the efforts of many researchers, most of the theory and computational techniques of 1-D linear prediction have not been extended to the 2-D case.

Considering the spectral factorization problem, the 1-D method of factoring a rational spectrum by computing its poles and zeroes does not extend to the 2-D case [8], [9]. Specifically, a rational 2-D spectrum almost never has a rational factorization (though under certain conditions it does have an infinite-order

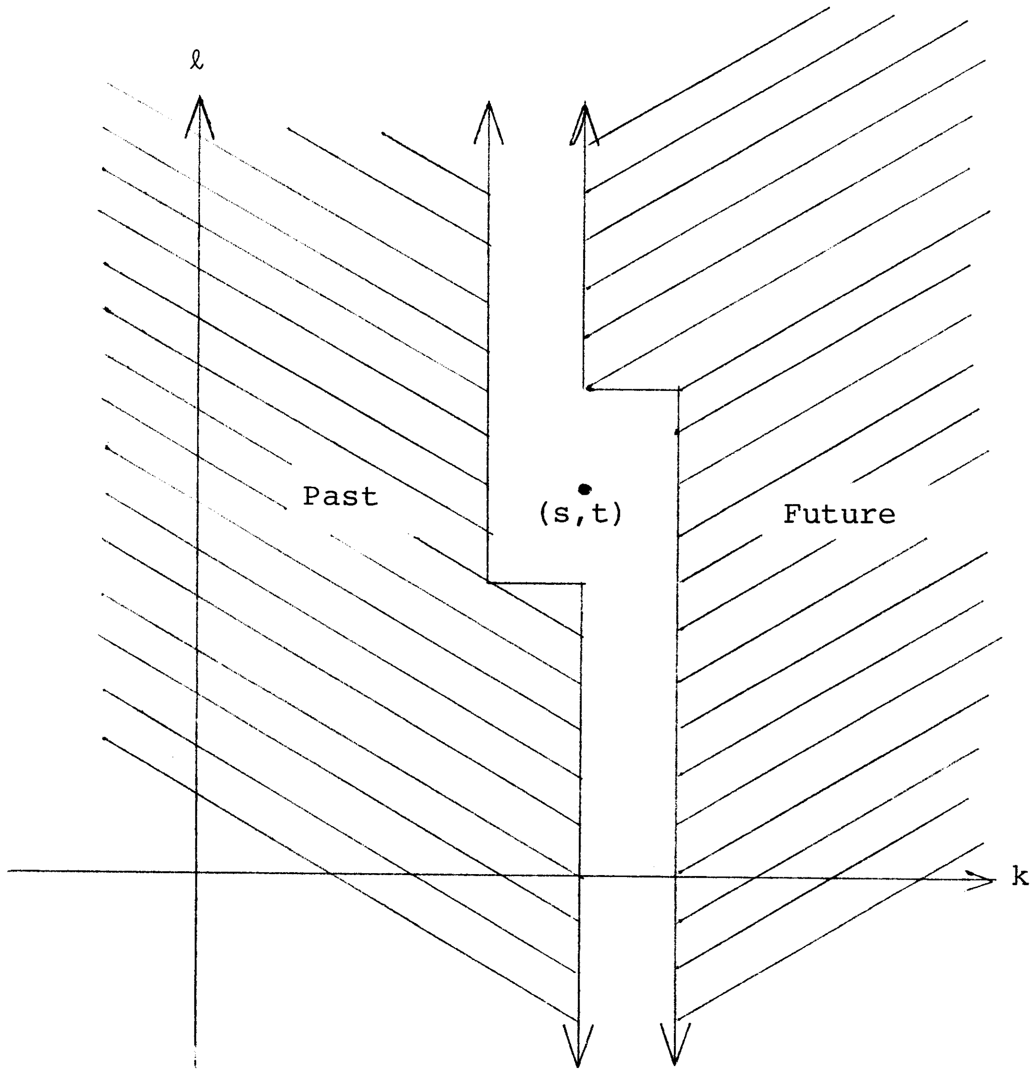


Fig. 1.2 Associated with any point (s,t) is a unique "past" and "future."

factorization). The implication of this is that in most cases we can only approximately factor a 2-D spectrum.

Shanks proposed an approximate method of 2-D spectral factorization which involves computing a finite-order least-squares linear prediction error filter [10]. Unfortunately, Shanks method, unlike an analogous 1-D method, does not always produce a causally invertible whitening filter approximation [11].

Probably the most successful method of 2-D spectral factorization to be proposed, is the Hilbert transform method (sometimes called the cepstral method or the homomorphic transformation method [8], [12], [13], [14]). The method relies on the fact that the phase and the log-magnitude of a 2-D causal and causally invertible filter are 2-D Hilbert transform pairs. While the method is theoretically exact, it can only be implemented approximately, and it has some practical difficulties.

Considering the autoregressive model fitting problem, neither the autocorrelation method nor the Burg algorithm has been successfully extended to the 2-D case. The 2-D autocorrelation method fails for the same reason that Shanks method fails. The Burg algorithm is essentially a stochastic version of the Levinson algorithm, which was originally derived as a fast method of inverting a Toeplitz covariance matrix [15]. Until now, no one has discovered a 2-D version of the Levinson algorithm that would enable a 2-D Burg algorithm to be devised.

1.5 New Results in 2-D Linear Prediction Theory

In this thesis we consider a special class of 2-D causal, linear, shift-invariant filters that has not previously been studied. The form of this class of filters is llustrated in Fig. 1.3. It can be seen that these filters are infinite-order in one variable, and finite-order in the other variable. Of greater significance is the fact that according to our definition of 2-D causality, the support for the unit sample response of these filters consists of the points (0,0) and (N,M), and all points in the future of (0,0) and in the past of (N,M). The basic theoretical result of this thesis is that by working with 2-D filters of this type, we can extend virtually all of the known 1-D linear prediction theory to the 2-D case. Among other things we can prove the following:

1) Given a 2-D, rational power density spectrum, $S(z_1,z_2)$, which is strictly positive and bounded on the unit circles, we can find a causal whitening filter for the random process which is a ratio of two filters, each of the form illustrated in Fig. 1.3. Both the numerator and the denominator polynomials of the whitening filter are analytic in the neighborhood of the unit circles (so the filters

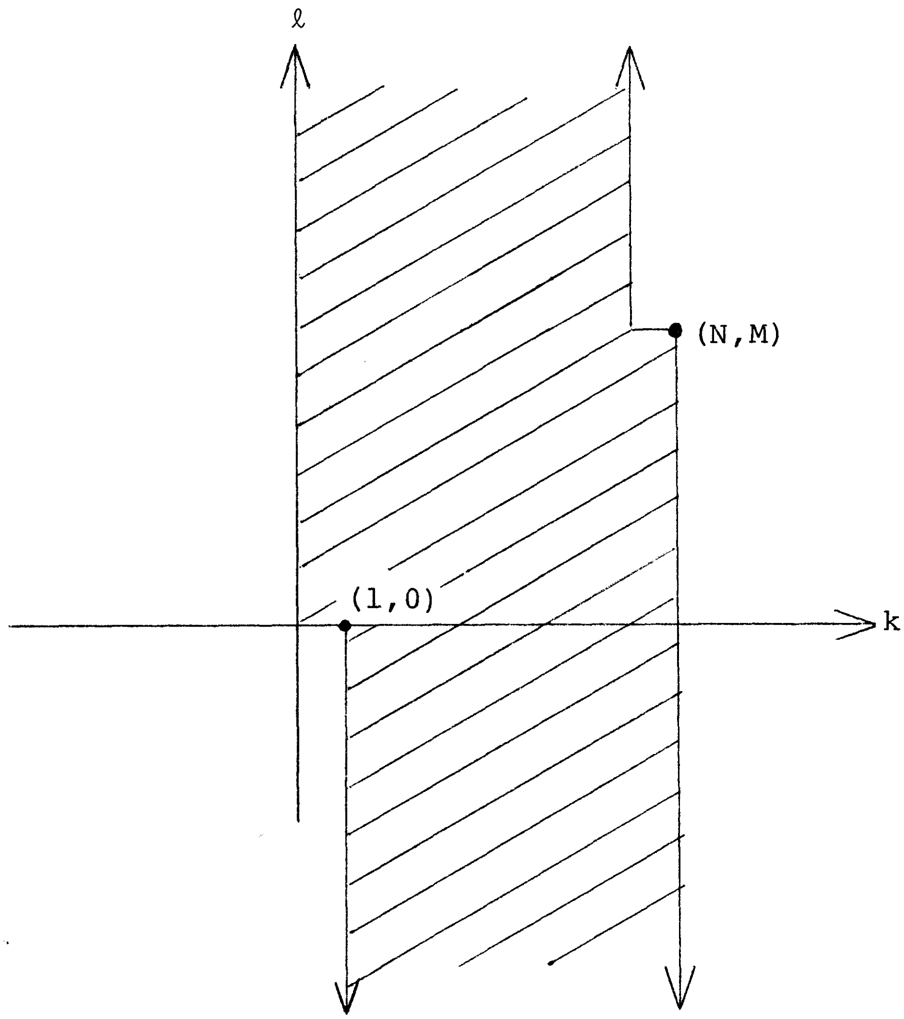


Fig. 1.3 A particular class of 2-D causal filters. The support consists of the points (0,0), (N,M), and all points in the future of (0,0) and in the past of (N,M).

are stable), and they have causal, analytic inverses (so the inverse filters are stable).

2) Consider the 2-D prediction problem illustrated in Fig. 1.4. The problem is to find the least-squares linear estimate for the point x(s,t) given the points shown in the shaded region. The solution of this problem involves solving an infinite set of linear equations. This problem is the same as that considered by Shanks, except that Shanks was working with a finite-order prediction-error filter, and here we are working with an infinite-order prediction error filter of the form illustrated in Fig. 1.3. Given certain conditions on the 2-D autocorrelation function (a sufficient condition is that the power density spectrum is analytic in the neighborhood of the unit circles, and strictly positive on the unit circles), we can prove that the prediction error filter is analytic in the neighborhood of the unit circles (and therefore stable) and that it has a causal and analytic (therefore stable) inverse.

3) From a practical standpoint, the most important theoretical result that we obtain is a canonical representation for a particular class of causal and causally invertible 2-D filters. The representation is an extension of the well-known 1-D reflection coefficient (or "partial correlation coefficient") representation for FIR minimum-phase filters [18] to the 2-D case.

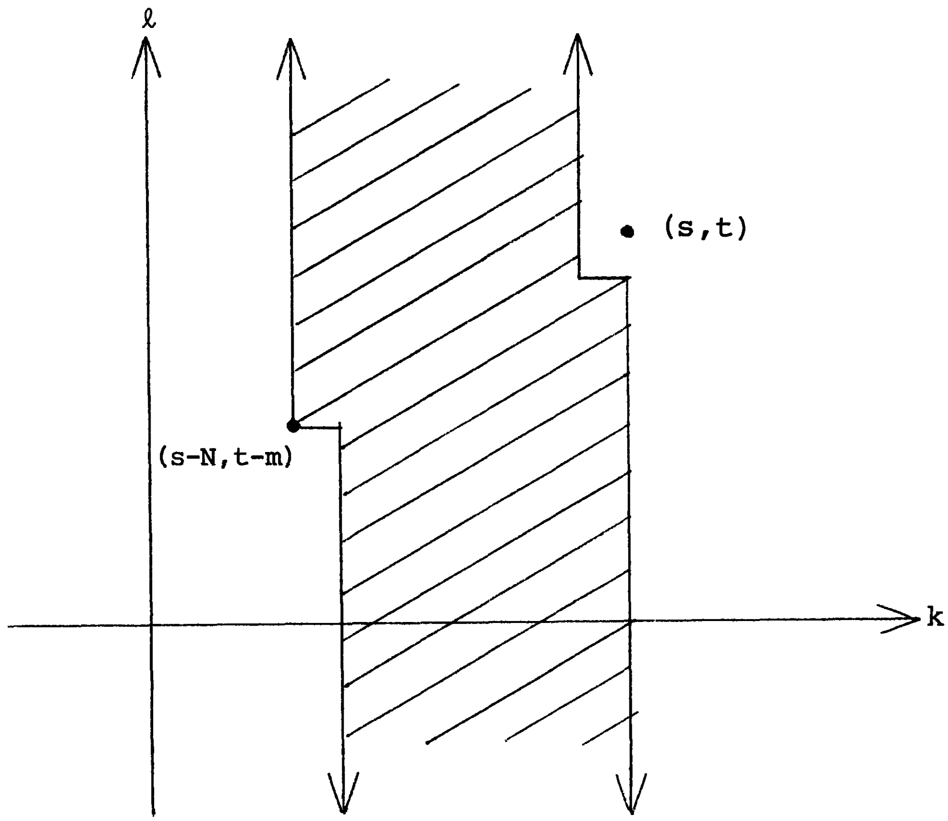


Fig. 1.4 The problem is to find the least-squares, linear estimate for the point x(s,t) given the points shown in the shaded region. Given certain conditions on the 2-D autocorrelation function, the prediction error filter is stable, and it has a causal, stable inverse.

We consider the class of 2-D filters having the support illustrated in Fig. 1.5(a). The filters themselves may be either finite-order or infinite-order. In addition we require that a) the filters be analytic in some neighborhood of the unit circles; b) the filters have causal inverses, analytic in some neighborhood of the unit circles; c) the filter coefficients at the origin be one. Then associated with any such filter is a unique 2-D sequence, called a reflection coefficient sequence, of the form illustrated in Fig. 1.5(b). The reflection coefficient sequence is obtainable from the filter by a recursive formula. The elements of the reflection coefficient sequence (called reflection coefficients) satisfy two conditions: their magnitudes are less than one, and they decay exponentially fast to zero as ℓ goes to plus or minus infinity. The relation between the class of filters and the class of reflection coefficient sequences is one-to-one.

In most cases, if the filter is finite-order, then the reflection coefficient sequence is infinite order. Fortunately, if the reflection coefficient sequence is finite-order then the filter is finite-order as well.

The practical significance of the 2-D reflection coefficient representation is that it provides a new domain in which to design 2-D FIR filters. Our point is that by formulating 2-D linear prediction problems (either

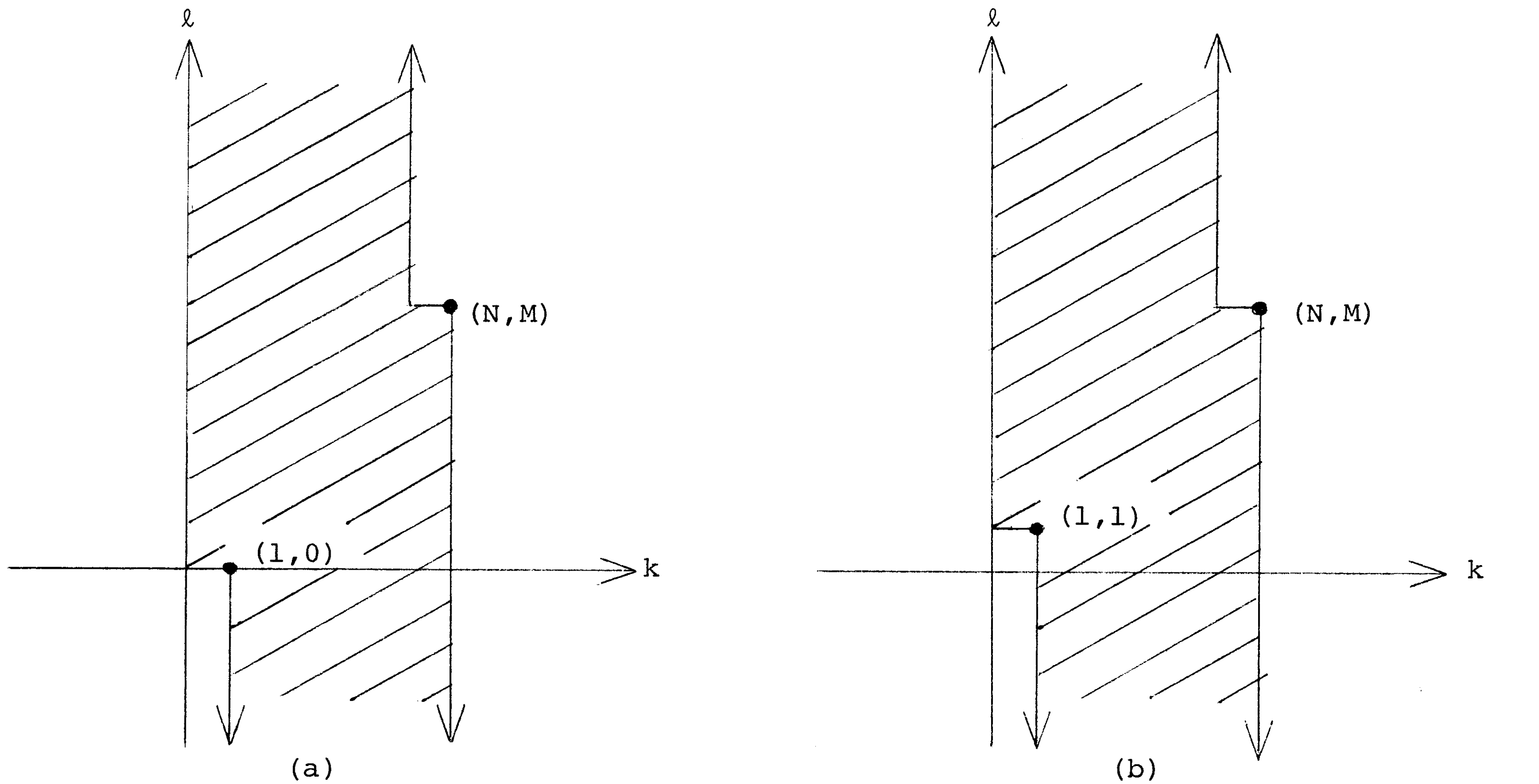


Fig. 1.5 2-D Reflection Coefficient Representation;
a) Filter (analytic with a causal, analytic inverse),
b) Reflection coefficient sequence.

spectral factorization or autoregressive model fitting) in the reflection coefficient domain, we can automatically satisfy the previously intractable requirement that the FIR filter be causally invertible. The idea is to attempt to represent the whitening filter by means of an FIR filter corresponding to a finite set of reflection coefficients, and to optimize over the reflection coefficients subject to the relatively simple constraint that the reflection coefficient magnitudes are less than one. As we prove later, if the power density spectrum is analytic in the neighborhood of the unit circles, and positive on the unit circles, then the whitening filter can be approximated arbitrarily closely in this manner (in a uniform sense) by using a large enough reflection coefficient sequence.

The remaining practical question concerns how to choose the reflection coefficients in an "optimal" way. For the spectral factorization problem, a convenient (but generally suboptimal) method consists of sequentially choosing the reflection coefficients subject to a least-squares criterion (In the 1-D case this algorithm reduces to the Levinson algorithm.) We present two numerical examples of this algorithm. For the autoregressive model fitting problem a similar suboptimal algorithm for sequentially choosing the reflection coefficients can be derived which, in the 1-D case, becomes the Burg algorithm.

It is believed that the full potential of the 2-D reflection coefficient representation can only be realized by using more sophisticated methods for choosing the reflection coefficients.

1.6 Preview of Remaining Chapters

Chapter 2 is a survey of the theory and computational techniques of 1-D linear prediction. While it contains no new results, it provides essential background for our discussion of 2-D linear prediction.

We begin our discussion of 2-D linear prediction in Chapter 3. We discuss the existing 2-D linear prediction theory, including the classical "failures" of 1-D results to extend to the 2-D case, and we review the available computational techniques of 2-D linear prediction. We introduce some terminology, and we prove some theorems that we use in our subsequent theoretical work. We discuss some potential applications of 2-D linear prediction.

Chapter 4 contains most of our new theoretical results. We state and prove 2-D versions of all of the 1-D theorems stated in Chapter 2.

In Chapter 5 we apply the 2-D reflection coefficient representation to the spectral factorization and autoregressive model fitting problems. We present numerical results involving our sequential spectral factorization algorithm.

CHAPTER 2

SURVEY OF ONE-DIMENSIONAL LINEAR PREDICTION

In this chapter we summarize some well-known 1-D linear prediction results. The theory that we review concerns the equivalence of three separate domains: the class of positive-definite Toeplitz covariance matrices, the class of minimum-phase FIR prediction error filters and positive prediction error variances, and the class of finite duration reflection coefficient sequences and positive prediction error variances. We illustrate the practical significance of this theory by showing how it applies to several methods of spectral factorization and autoregressive model fitting.

2.1 1-D Linear Prediction Theory

Throughout this chapter we assume that we are working with a real, discrete-time, zero-mean, wide-sense stationary random process $x(t)$, where t is an integer. We denote the autocorrelation function by

$$r(\tau) = E\{x(t+\tau)x(t)\} \quad , \tag{2.1}$$

and the power density spectrum by

$$S(z) = \sum_{\tau=-\infty}^{\infty} r(\tau) z^{-\tau} \quad . \tag{2.2}$$

We consider the problem of finding the minimum mean-square error linear predictor for the point x(t) given the N preceding points:

$$[\hat{x}(t)\,|\,x(t-1),x(t-2),\ldots,x(t-N)] = \sum_{i=1}^{N} h(N;i)x(t-i) \quad . \tag{2.3}$$

We determine the optimum predictor coefficients by applying the Orthogonality Principle, according to which the least-squares linear prediction error is orthogonal to each data point [16]:

$$E\{[x(t) - \sum_{i=1}^{N} h(N;i)x(t-i)]x(t-s)\}$$

$$= [r(s) - \sum_{i=1}^{N} h(N;i)r(s-i)] = 0 \quad , \quad 1\leq s\leq N \; . \tag{2.4}$$

These equations are called the normal equations, or the Yule-Walker equations. We denote the optimum mean-square prediction error by

$$P_N = E\{[x(t) - \sum_{i=1}^{N} h(N;i)x(t-i)]^2\}$$

$$= [r(0) - \sum_{i=1}^{N} h(N;i)r(-i)] \quad . \tag{2.5}$$

Writing the normal equations in matrix form we have

$$\begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(N) \\ r(1) & r(0) & r(1) & \cdots & r(N-1) \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ r(N) & r(N-1) & r(N-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ -h(N;1) \\ \cdot \\ \cdot \\ \cdot \\ -h(N;N) \end{bmatrix} = \begin{bmatrix} P_N \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \quad (2.6)$$

The matrix is a symmetric, non-negative definite Toeplitz covariance matrix. The following theorem can be proved.

Theorem 2.1(a): Assume that the covariance matrix in (2.6) is positive definite. Then

1) (2.6) has a unique solution for the filter coefficients, $\{h(N;1),\ldots,h(N;N)\}$, and the prediction error variance, P_N;

2) P_N is positive;

3) The prediction error filter (PEF)

$$H_N(z) = [1 - \sum_{i=1}^{N} h(N;i)z^{-i}] \quad (2.7)$$

is minimum-phase (that is, the magnitudes of its poles and zeroes are less than one) [17], [18].

A converse to Theorem 2.1(a) can also be proved:

Theorem 2.1(b): Given any positive P_N, and any minimum-phase $H_N(z)$, where $H_N(z)$ is of the form (2.7), there is exactly one $(N+1)\times(N+1)$ positive-definite, Toeplitz covariance matrix such that (2.6) is satisfied. The elements of the covariance matrix are given by the formula

$$r(\tau) = \frac{1}{2\pi j} \oint_{|z|=1} \frac{z^{(\tau-1)} P_N dz}{H_N(z) H_N(1/z)} , \quad |\tau| \leq N . \tag{2.8}$$

[7], [3], [19]

The normal equations are a set of (N+1) simultaneous linear equations. Using the Gaussian elimination technique they can be solved with about $N^3/3$ computations. Levinson devised an algorithm for solving the normal equations, taking advantage of the Toeplitz structure of the covariance matrix, that requires only about N^2 computations. The algorithm operates by successively computing PEFs of increasing order.

Theorem 2.2 (Levinson algorithm): Suppose that the covariance matrix in (2.6) is positive-definite; then (2.6) can be solved by performing the following steps:

1) $$\rho(1) = \frac{r(1)}{r(0)} , \tag{2.9}$$

$$h(1;1) = \rho(1) , \tag{2.10}$$

$$P_1 = r(0)[1 - \rho^2(1)] ; \tag{2.11}$$

2) $$\rho(n) = \frac{1}{P_{n-1}} E\{[x(t) - \sum_{i=1}^{(n-1)} h(n-1;i)x(t-i)] \cdot [x(t-n) - \sum_{i=1}^{(n-1)} h(n-1;i)x(t-n+i)]\}$$

$$= \frac{1}{P_{n-1}} [r(n) - \sum_{i=1}^{(n-1)} h(n-1;i)r(n-i)] \quad , \quad (2.12)$$

$$h(n;n) = \rho(n) \quad , \quad (2.13)$$

$$h(n;i) = [h(n-1;i) - \rho(n)h(n-1;n-i)] \quad , \quad 1 \leq i \leq (n-1) \quad , \quad (2.14)$$

$$P_n = P_{n-1}[1 - \rho^2(n)] \quad ; \quad (2.15)$$

$$2 \leq n \leq N \quad . \quad [15], [20]$$

The numbers $\rho(n)$, given by (2.9) and (2.12), are called "reflection coefficients," and their magnitudes are always less than one. (The term "reflection coefficient" is used because a physical interpretation for the Levinson algorithm is that it solves for the structure of a 1-D layered medium (i.e., the reflection coefficients) given the medium's reflection response [6]. The reflection coefficients are also called partial correlation coefficients, since they are partial correlation coefficients between forward and backward prediction errors.)

Equations (2.10), (2.13), and (2.14) can be written in the more convenient Z-transform notation as follows:

$$H_0(z) = 1 \quad , \tag{2.16}$$

$$H_n(z) = [H_{n-1}(z) - \rho(n) z^{-n} H_{n-1}(1/z)] \quad , \quad 1 \le n \le N \; , \tag{2.17}$$

where $$H_n(z) = [1 - \sum_{i=1}^{n} h(n;i) z^{-i}], \quad 1 \le n \le N \; . \tag{2.18}$$

One interpretation of the Levinson algorithm is that it solves for the PEF, $H_N(z)$, by representing the filter in terms of the reflection coefficient sequence, $\{\rho(1), \rho(2), \ldots, \rho(N)\}$, and by sequentially choosing the reflection coefficients in an optimum manner. This reflection coefficient representation is a canonical representation for FIR minimum-phase filters:

Theorem 2.3(a): Given any reflection coefficient sequence, $\{\rho(1), \rho(2), \ldots, \rho(N)\}$, where the reflection coefficient magnitudes are less than one, there is a unique sequence of minimum-phase filters, $\{H_0(z), H_1(z), \ldots, H_N(z)\}$, of the form (2.18), satisfying the following recursion:

$$H_0(z) = 1 \quad , \tag{2.19}$$

$$H_n(z) = [H_{n-1}(z) - \rho(n) z^{-n} H_{n-1}(1/z)] \quad ,$$

$$1 \leq n \leq N \quad . \qquad (2.20)$$

Theorem 2.3(b): Given any minimum-phase filter, $H_N(z)$, of the form (2.18), there is a unique reflection coefficient sequence, $\{\rho(1), \rho(2), \ldots, \rho(N)\}$, where the reflection coefficient magnitudes are less than one, and a unique sequence of minimum-phase filters, $\{H_0(z), H_1(z), \ldots, H_{N-1}(z)\}$, of the form (2.18), satisfying the following recursion:

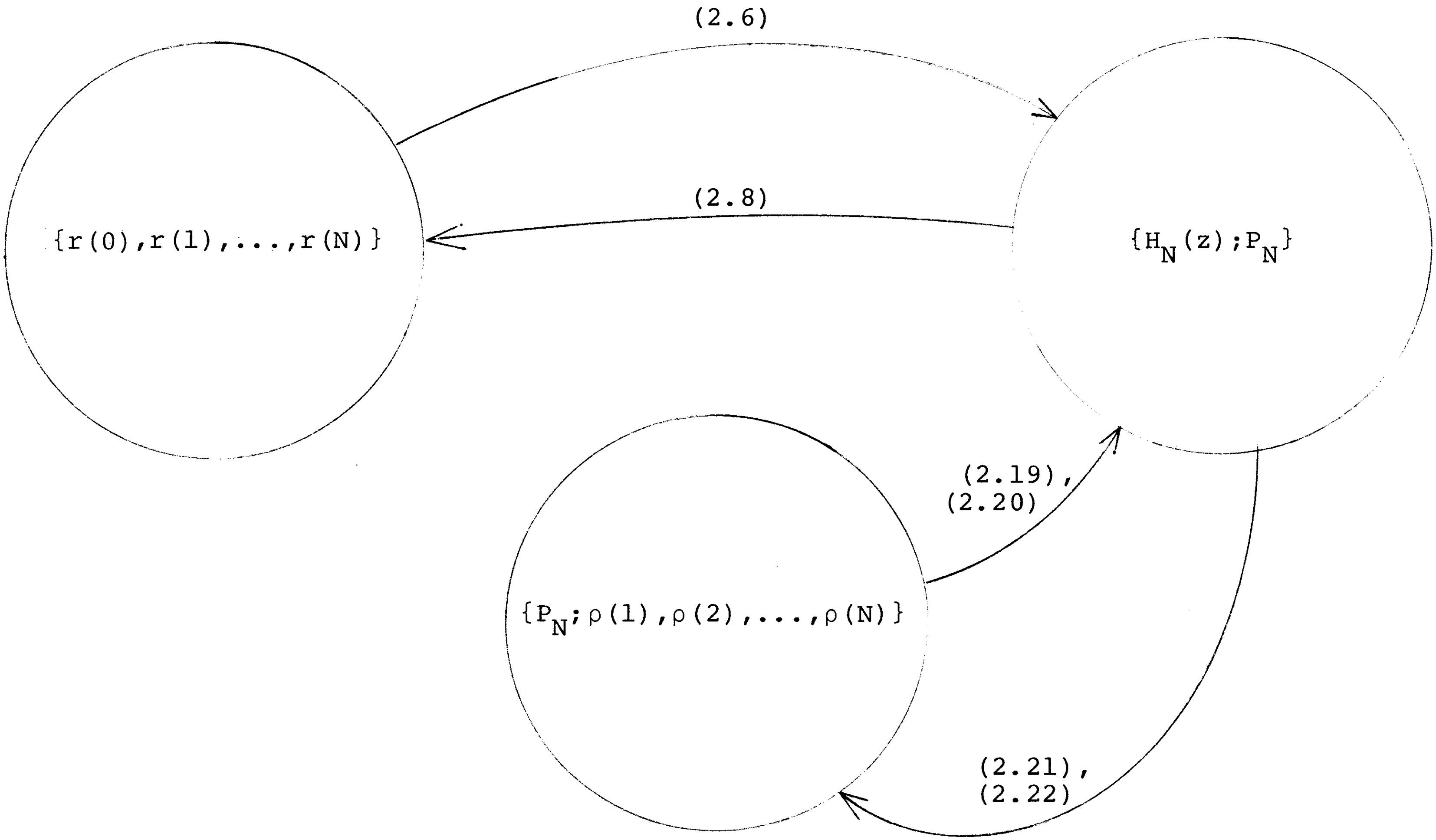
$$\rho(n) = h(n;n) \quad , \qquad (2.21)$$

$$H_{n-1}(z) = \frac{1}{[1-\rho^2(n)]} [H_n(z) + \rho(n) z^{-n} H_n(1/z)] \quad , \qquad (2.22)$$

$$N \geq n \geq 1 \quad .$$ [18], [7], [3]

Theorems 2.1 and 2.3 are summarized in Fig. 2.1.

Finally, we want to discuss the behavior of the sequence of PEFs, $H_N(z)$, as N goes to infinity. The basic result is that by imposing some conditions on the power density spectrum, the sequence $H_N(z)$ converges uniformly to the causal and causally invertible whitening filter for the random process.



<u>Fig. 2.1</u> The correspondence among 1-D positive-definite autocorrelation sequences, FIR minimum-phase PEFs and positive prediction error variances, and reflection coefficient sequences and positive prediction error variances.

Theorem 2.4: If the power density spectrum, S(z), is analytic in some neighborhood of the unit circle, and strictly positive on the unit circle, then

1) The sequence of minimum-phase PEFs, $H_N(z)$, converges uniformly in some neighborhood of the unit circle to a limit filter

$$\lim_{N\to\infty} H_N(z) = H_\infty(z) \quad ; \tag{2.23}$$

2) $H_\infty(z)$ is analytic in some neighborhood of the unit circle, it has a causal analytic inverse, and it is the unique (to within a multiplicative constant) causal and causally invertible whitening filter for the random process;

3) The reflection coefficient sequence decays exponentially fast to zero as N goes to infinity

$$|\rho(N)| < (1+\varepsilon)^{-N} \quad , \quad \varepsilon>0 \quad ; \tag{2.24}$$

4) The sequence of prediction error variances converges to a positive limit

$$\lim_{N\to\infty} P_N = P_\infty$$

$$= \exp[\frac{1}{2\pi j} \oint_{|z|=1} z^{-1} \log S(z)dz] \quad . \tag{2.25}$$

[21], [7]

2.2 1-D Spectral Factorization

As we stated in the introduction, the spectral factorization problem is the following: given a spectrum $S(z)$, find a causal and causally invertible whitening filter for the random process. Equivalently, the problem is to write the spectrum in the form

$$S(z) = \frac{1}{G(z)G(1/z)} \tag{2.26}$$

where $G(z)$ is causal and stable, and has a causal and stable inverse. A sufficient condition for a spectrum to be factorizable is that it is analytic in some neighborhood of the unit circle, and positive on the unit circle. In this section we discuss two approximate methods of spectral factorization, the Hilbert transform method, and the linear prediction method.

Considering first the Hilbert transform method, if the spectrum is analytic in some neighobrhood of the unit circle, and positive on the unit circle, it can be shown that the complex logarithm of the spectrum is also analytic in some neighborhood of the unit circle, and it therefore has a Laurent expansion in that region [1]

$$\log S(z) = \sum_{n=-\infty}^{\infty} c_n z^{-n} \quad , \tag{2.27}$$

where $$c_n = c_{-n} = \frac{1}{2\pi j} \oint_{|z|=1} z^{n-1} \log S(z)dz. \tag{2.28}$$

(The sequence c_n is called the "real cepstrum.")

Or $$\log S(z) = C(z) + C(1/z) \tag{2.29}$$

where $$C(z) = \left(\frac{c_0}{2} + \sum_{n=1}^{\infty} c_n z^{-n} \right) . \tag{2.30}$$

Therefore,

$$S(z) = \frac{1}{G(z)G(1/z)} \tag{2.31}$$

where $$G(z) = \exp[-C(z)] . \tag{2.32}$$

It is straightforward to prove that G(z) is causal and analytic in the neighborhood of the unit circle, and that it has a causal and analytic inverse.

While the Hilbert transform method is a theoretically exact method, it can only be implemented approximately by means of discrete Fourier transform (DFT) operations. The basic difficulty is that the exact cepstrum is virtually always infinite-order, and it can only be approximated by a finite-order cepstrum. A finite cepstrum always produces an infinite-order filter, according to (2.32), but again this infinite-order filter is truncated in practice. Consequently in using the Hilbert transform method, there are always two separate truncations involved. Both truncations can distort the frequency response of the whitening filter approximation, and the second truncation can even product a nonminimum-phase filter. These difficulties can always be overcome by performing the DFTs with a sufficiently fine sample spacing, but one can never predict in advance how fine this spacing should be. Particular difficulties are encountered whenever the spectrum has poles or zeroes close to the unit circle.

The basic idea of the linear prediction method of spectral factorization is to approximate the causal and causally invertible whitening filter by a finite-order PEF, $H_N(z)$, for some value of N. If the spectrum is analytic and positive, then according to Theorem 2.1(a), $H_N(z)$ is minimum-phase, and according to Theorem 2.4, this approximation can be made as accurate as desired by making N large enough. The principle difficulty is choosing N. One possible criterion is to choose N large enough so that the prediction error variance, P_N, is sufficiently close to its limit, P_∞ (which can be precomputed by means of the formula (2.25)).

2.3 1-D Autoregressive Model Fitting

We recall that the problem of autoregressive model fitting is the following: given a finite set of samples from the random process, estimate the causal and causally invertible whitening filter. In contrast to spectral factorization which is a deterministic problem, the problem of autoregressive model fitting is one of stochastic estimation. Two convenient and effective methods of autoregressive model fitting are the autocorrelation method [3] and the Burg alogrithm [3].

Given a finite segment of the random process, $\{x(0),x(1),...,x(T)\}$, the autocorrelation method first uses the data samples to estimate the autocorrelation function to a finite lag. Then an N-th order PEF and prediction error variance, $\hat{H}_N(z)$ and $\hat{P}_N$, are computed for some $N \leq T$ by solving the normal equations associated with the estimated autocorrelation sequence. The autocorrelation estimate commonly used is

$$\hat{r}(\tau) = \frac{1}{(T+1)} \sum_{t=0}^{(T-|\tau|)} x(t+|\tau|)x(t) \quad , \quad |\tau| \leq T. \tag{2.33}$$

If the true autocorrelation sequence is positive definite, then $\hat{r}(\tau)$ is positive definite with probability one, and according to Theorem 2.1(a), $\hat{H}_N(z)$ is minimum-phase and $\hat{P}_N$ is positive. Furthermore, according to Theorem 2.1(b) the autoregressive spectrum,

$$\hat{S}(z) = \frac{\hat{P}_N}{\hat{H}_N(z)\hat{H}_N(1/z)} \quad , \tag{2.34}$$

is consistent with the autocorrelation estimate, $\hat{r}(\tau)$, for $|\tau| \leq N$. (The autoregressive spectrum is sometimes called the maximum-entropy spectrum; it can be shown that among all spectra consistent with $\hat{r}(\tau)$ for $|\tau| \leq N$, the N-th order autoregressive spectrum has the greatest entropy, where the entropy is defined by the formula

$$\frac{1}{2\pi j} \oint_{|z|=1} z^{-1} \log \hat{S}(z) dz \quad [3] \quad .) \tag{2.35}$$

The Burg algorithm operates by successively fitting higher order PEFs directly to the data. The basic idea is to mimic the Levinson algorithm. The whitening filter estimate, $\hat{H}_N(z)$, is found by sequentially choosing its reflection coefficients subject to a particular optimality criterion. Given the random process segment, $\{x(0), x(1), \ldots, x(T)\}$, the algorithm proceeds as follows:

1) $$\hat{H}_0(z) = 1 \quad , \tag{2.36}$$

$$\hat{P}_0 = \frac{1}{(T+1)} \sum_{t=0}^{T} x^2(t) \quad ; \tag{2.37}$$

2) At the beginning of the n-th stage we have $\hat{H}_{n-1}(z)$ and $\hat{P}_{n-1}$. The only new parameter to estimate is the new reflection coefficient, $\hat{\rho}(n)$. The PEF and the prediction error variance are then updated by the formulas

$$\hat{H}_n(z) = \hat{H}_{n-1}(z) - \hat{\rho}(n) z^{-n} \hat{H}_{n-1}(1/z) \tag{2.38}$$

and $$\hat{P}_n = \hat{P}_{n-1}[1 - \hat{\rho}^2(n)] \quad . \tag{2.39}$$

The new reflection coefficient is chosen to minimize the sum of the squares of the n-th order forward and backward prediction errors,

$$\sum_{t=n}^{T} \{[\varepsilon_n^{(+)}(t)]^2 + [\varepsilon_n^{(-)}(t-n)]^2\} \quad , \tag{2.40}$$

where

$$\varepsilon_n^{(+)}(t) = [x(t) - \sum_{i=1}^{n} \hat{h}(n;i)x(t-i)] \quad , \quad n \leq t \leq T \quad , \tag{2.41}$$

and $$\varepsilon_n^{(-)}(t) = [x(t) - \sum_{i=1}^{n} \hat{h}(n;i)x(t+i)] \quad , \quad 0 \leq t \leq (T-n) \quad . \tag{2.42}$$

The expression (2.40) is minimized by choosing $\hat{\rho}(n)$ according to the formula

$$\hat{\rho}(n) = \frac{2 \sum_{t=n}^{T} [\varepsilon_{n-1}^{(+)}(t)][\varepsilon_{n-1}^{(-)}(t-n)]}{\sum_{t=n}^{T} \{[\varepsilon_{n-1}^{(+)}(t)]^2 + [\varepsilon_{n-1}^{(-)}(t-n)]^2} \quad . \tag{2.43}$$

It can be shown that the magnitude of $\hat{\rho}(n)$ is less than one, and therefore $\hat{H}_n(z)$ is minimum-phase and $\hat{P}_n$ is positive.

The forward and backward prediction errors do not have to be directly computed at each stage of the algorithm; instead they can be recursively computed by the formulas

$$\varepsilon_n^{(+)}(t) = [\varepsilon_{n-1}^{(+)}(t) - \hat{\rho}(n)\varepsilon_{n-1}^{(-)}(t-n)] \quad , \qquad n \leq t \leq T \quad , \tag{2.44}$$

and

$$\varepsilon_n^{(-)}(t) = [\varepsilon_{n-1}^{(-)}(t) - \hat{\rho}(n)\varepsilon_{n-1}^{(+)}(t+n)] \quad , \qquad 0 \leq t \leq (T-n) \quad . \tag{2.45}$$

The Burg algorithm has been found experimentally to give better resolution than the correlation method in cases where the final order of the PEF is comparable to the length of the data segment [22]. This is apparently due to the bias of the autocorrelation estimate used in the correlation method.

Regardless of which method is used, the most difficult problem in autoregressive model fitting is choosing the order of the model. While in special cases we may know this in advance, that is not usually the case. In practice, the order of the model is made large enough so that $\hat{P}_N$ appears to be approaching a lower limit. At present there is no universally optimal way to make this decision.

CHAPTER 3

TWO-DIMENSIONAL LINEAR PREDICTION - BACKGROUND

3.1 2-D Random Processes and Linear Prediction

For the remainder of this thesis we assume that we are working with a 2-D, wide-sense stationary random process, $x(k,\ell)$, where k and ℓ are integers. $x(k,\ell)$ is further assumed to be zero-mean, real, and scalar-valued. The autocorrelation function is denoted

$$r(s,t) = E\{x(k+s,\ell+t)x(k,\ell)\} \quad , \qquad (3.1)$$

and the power density spectrum is denoted

$$S(z_1,z_2) = \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} r(s,t) z_1^{-s} z_2^{-t} \quad . \qquad (3.2)$$

As stated in the introduction, the 2-D linear prediction problem concerns the determination of a causal, stable whitening filter for $x(k,\ell)$ which has a causal, stable inverse.

3.2 Two-dimensional Causality

As we mentioned in the introduction, our 2-D linear prediction results are based on a particular notion of 2-D causality. For any point (s,t) we define the <u>past</u> to be the set of points

$$\{(k,\ell) \mid k=s, \ell<t; \; k<s, -\infty<\ell<\infty\} \quad ,$$

and the <u>future</u> to be the set of points

$$\{(k,\ell)\,|\,k=s,\ell>t;\ k>s,-\infty<\ell<\infty\} \quad .$$

This is illustrated in Fig. 1.2. It is straightforward to verify two implications of this definition:

1) If (k_1,ℓ_1) is in the past of (k_2,ℓ_2), then (k_2,ℓ_2) is in the future of (k_1,ℓ_1) ;

2) If (k_1,ℓ_1) is in the past of (k_2,ℓ_2), and (k_2,ℓ_2) is in the past of (k_3,ℓ_3), then (k_1,ℓ_1) is in the past of (k_3,ℓ_3).

In other words, our definition of causality totally orders the points in the plane.

As a matter of notation, if (k_1,ℓ_1) is in the past of (k_2,ℓ_2), we denote this by

$$(k_1,\ell_1) < (k_2,\ell_2) \quad ,$$

or equivalently

$$(k_2,\ell_2) > (k_1,\ell_1) \quad .$$

Therefore we define a causal 2-D linear, shift-invariant filter to be one whose unit sample response vanishes at all points in the past of the origin. Equivalently, a 2-D filter, $A(z_1,z_2)$, is <u>causal</u> if its Z-transform can be written in the form

$$A(z_1,z_2) = \sum_{\ell=0}^{\infty} a(0,\ell) z_2^{-\ell} + \sum_{k=1}^{\infty} \sum_{\ell=-\infty}^{\infty} a(k,\ell) z_1^{-k} z_2^{-\ell}$$

$$= \sum\sum_{(k,\ell) \geq (0,0)} a(k,\ell) z_1^{-k} z_2^{-\ell} \quad . \qquad (3.3)$$

The geometry of a 2-D causal filter is illustrated in Fig. 1.1.

3.3 The 2-D Minimum-phase Condition

We define a 2-D, stable, linear, shift-invariant filter (either causal or non-causal) to be one whose unit sample response is absolutely summable. Equivalently, the Z-transform of a stable 2-D filter converges absolutely on the unit circles (for $|z_1| = 1 = |z_2|$). It can be shown that such a filter is stable in the bounded input, bounded output sense; that is, if the input to the filter is bounded, then the output is bounded as well [23].

We define a 2-D minimum-phase filter to be a 2-D causal, stable, linear, shift-invariant filter which has a causal, stable inverse. If a filter is minimum-phase then it is easy to show that its inverse is unique.

Throughout this thesis we will be mainly concerned with a special class of 2-D minimum-phase filters that we call analytic minimum-phase filters. We define a 2-D filter to be an analytic minimum-phase filter if 1) the filter is minimum-phase, 2) the filter is analytic in some neighborhood of the unit circles (that is for some $(1-\varepsilon)<|z_1|,|z_2|<(1+\varepsilon))$, and 3) the inverse filter is analytic in some neighborhood of the unit circles. (The third condition is redundant, since any function which is analytic and non-zero in a region has an analytic inverse in the same region.) Not every 2-D minimum-phase filter is analytic. Consider for example the filter

$$A(z_1,z_2) = \{1 + \frac{1}{10} z_1^{-1} [\sum_{\ell=1}^{\infty} \frac{1}{\ell^2} (z_2^{-\ell}+z_2^{\ell})] \} \quad . \quad (3.4)$$

Recalling the identity

$$\sum_{\ell=1}^{\infty} \frac{1}{\ell^2} = \frac{\pi^2}{6} \quad (3.5)$$

we see that $A(z_1,z_2)$ is causal and stable. But (3.4) diverges whenever z_2 is off the unit circle, so the filter is not analytic. Nevertheless, it does have a causal, stable inverse. Using a formal geometric series, we have

$$A^{-1}(z_1,z_2) = \{ \sum_{k=0}^{\infty} \frac{(-1)^k}{10^k} [\sum_{\ell=1}^{\infty} \frac{1}{\ell^2} (z_2^{-\ell}+z_2^{\ell})]^k z_1^{-k} \} \quad (3.6)$$

where the series converges uniformly for $|z_2| = 1$.

From a practical standpoint, 2-D minimum-phase filters which are not analytic are of no importance. In practice, rational 2-D minimum-phase filters are the only type of 2-D minimum-phase filters that would ever be implemented, and these filters are analytic.

The 2-D minimum-phase condition is considerably more complicated than the 1-D minimum-phase condition. This is primarily due to the fact that 2-D filters are characterized by an uncountably infinite number of poles and zeroes. There is a great amount of literature devoted to various algebraic minimum-phase tests for 2-D FIR one-quadrant filters [10], [24], [25], [26]. This author has shown how to extend these tests to the more general class of causal 2-D FIR filters whose support occupies more than one quadrant [27]. Another approach to testing the 2-D minimum-phase condition is a numerical one based on the cepstral representation for mininum-phase filters [14]. However, we do not discuss any of the above tests, since at no point in this thesis do we actually need to test the minimum-phase condition for a particular filter. What we do need is a theoretical tool that will enable us to prove inductively the condition for a particular class of filters. To that end we now state and prove a theorem which is a 2-D extension of a well-known 1-D theorem [22].

Theorem 3.1: If $A(z_1,z_2)$ is an analytic minimum-phase filter, and $\delta(z_1,z_2)$ is a causal filter, analytic in some neighborhood of the unit circles, whose magnitude is less than the magnitude of $A(z_1,z_2)$ when z_1 and z_2 are on the unit circles,

$$|\delta(z_1,z_2)| < |A(z_1,z_2)| \quad , \qquad |z_1| = |z_2| = 1 \quad , \tag{3.7}$$

then the sum of the two filters is an analytic minimum-phase filter. (Note that $\delta(z_1,z_2)$ does not have to be minimum-phase.)

Proof: The sum of the two filters is analytic in some neighborhood of the unit circles. Since the sum is non-zero on the unit circles, continuity implies that the sum is non-zero in some neighborhood of the unit circles. Therefore the inverse filter is analytic in some neighborhood of the unit circles. To prove that the inverse filter is causal we proceed as follows:

$$\begin{aligned}[A(z_1,z_2) + \delta(z_1,z_2)]^{-1} \\ &= A^{-1}(z_1,z_2)[1 + A^{-1}(z_1,z_2)\delta(z_1,z_2)]^{-1} \\ &= A^{-1}(z_1,z_2) \sum_{n=0}^{\infty} (-1)^n [A^{-1}(z_1,z_2)\delta(z_1,z_2)]^n \quad ,\end{aligned} \tag{3.8}$$

where the series converges uniformly in some neighborhood of the unit circles. It is easy to prove that a product of 2-D causal filters is also causal. Therefore each term in the series is causal, so the uniform limit is causal.

3.4 Properties of 2-D Minimum-phase Whitening Filters

In this section we assume that a particular 2-D random process has a minimum-phase whitening filter (ignoring for the present the question of existence) and we state and prove some properties of this whitening filter. We denote the minimum-phase whitening filter by $A(z_1,z_2)$, and we denote its inverse by $B(z_1,z_2)$. We then have

$$w(k,\ell) = \sum\sum_{(s,t)\geq(0,0)} a(s,t)x(k-s,\ell-t) \quad , \qquad (3.9)$$

and

$$x(k,\ell) = \sum\sum_{(s,t)\geq(0,0)} b(s,t)w(k-s,\ell-t) \quad , \qquad (3.10)$$

where $w(k,\ell)$ is a white-noise process with variance σ^2:

$$E\{w(k+s,\ell+t)w(k,\ell)\} = \sigma^2\delta_s\delta_t \quad . \qquad (3.11)$$

(The generally infinite sums in (3.9) and (3.10) are interpreted in the mean-square sense.) We can show the following:

1) The minimum-phase whitening filter is related to the power spectrum by the formula

$$S(z_1,z_2) = \frac{\sigma^2}{A(z_1,z_2)A(1/z_1,1/z_2)} \quad ; \qquad (3.12)$$

2) $A(z_1,z_2)$ is unique to within a multiplicative constant;

3) $A(z_1,z_2)$ is proportional to the least-squares linear prediction error filter for $x(k,\ell)$ given the infinite past.

To prove (1) we substitute (3.9) into (3.11) and we obtain

$$\sigma^2\delta_s\delta_t = \sum\sum_{(k_1,\ell_1)\geq(0,0)} \sum\sum_{(k_2,\ell_2)\geq(0,0)} a(k_1,\ell_1)a(k_2,\ell_2) \cdot r(s-k_1+k_2,t-\ell_1+\ell_2) \quad . \tag{3.13}$$

Taking the z-transform of both sides of the equation, we have

$$\sigma^2 = S(z_1,z_2)A(z_1,z_2)A(1/z_1,1/z_2) \quad , \tag{3.14}$$

which reduces to (3.12).

To prove (2) we assume the existence of some other minimum-phase whitening filter, $A'(z_1,z_2)$. We have

$$S(z_1,z_2) = \frac{\sigma^2}{A(z_1,z_2)A(1/z_1,1/z_2)} = \frac{\sigma'^2}{A'(z_1,z_2)A'(1/z_1,1/z_2)} \quad . \tag{3.15}$$

Or

$$\frac{A'(z_1,z_2)\sigma^2}{A(z_1,z_2)} = \frac{A(1/z_1,1/z_2)\sigma'^2}{A'(1/z_1,1/z_2)} \quad . \tag{3.16}$$

The left-hand side of (3.16) is a causal filter, and the right-hand side is an anti-causal filter. Clearly (3.16) can be true only if each side of the equation equals a constant. Therefore,

$$A'(z_1,z_2) = cA(z_1,z_2) \tag{3.17}$$

where c is a constant.

To prove (3) we first normalize the whitening filter as follows:

$$H(z_1,z_2) = \frac{1}{a(0,0)} A(z_1,z_2)$$

$$= [1 - \sum\sum_{(k,\ell)>(0,0)} h(k,\ell) z_1^{-k} z_2^{-\ell}] \quad . \tag{3.18}$$

We then have

$$S(z_1,z_2) = \frac{P}{H(z_1,z_2)H(1/z_1,1/z_2)} , \tag{3.19}$$

where $$P = \frac{\sigma^2}{a^2(0,0)} \; . \tag{3.20}$$

Our claim is that $H(z_1,z_2)$ is the least-squares linear prediction error filter for $x(k,\ell)$ given the infinite past, and that P is the prediction error variance. We have

$$H(z_1,z_2)S(z_1,z_2) = \frac{P}{H(1/z_1,1/z_2)} \quad . \tag{3.21}$$

Taking the inverse Z-transform of both sides of (3.21), and using the fact that the right-hand side is an anti-causal filter, we have

$$[r(s,t) - \sum\sum_{(k,\ell)>(0,0)} h(k,\ell)r(s-k,t-\ell)] = P\delta_s\delta_t \quad ,$$

$$(s,t)\geq(0,0) \quad , \tag{3.22}$$

or

$$E\{[x(u,v) - \sum\sum_{(k,\ell)>(0,0)} h(k,\ell)x(u-k,v-\ell)]x(u-s,v-t)\}$$

$$= P\delta_s\delta_t \quad , \quad (s,t)\geq(0,0) \quad . \tag{3.23}$$

According to (3.23), $H(z_1,z_2)$ operates on the random process $x(k,\ell)$ to produce a white-noise process that is uncorrelated with all past values of $x(k,\ell)$. Therefore the Orthogonality Principle is satisfied, so $H(z_1,z_2)$ has the linear prediction interpretation that we claim for it. (Using (3.23) it is easy to prove that first, no other PEF can perform better than $H(z_1,z_2)$, and second, that any PEF which performs as well as $H(z_1,z_2)$ must be equal to $H(z_1,z_2)$.)

3.5 2-D Spectral Factorization

We recall that the problem of 2-D spectral factorization is the following: given a 2-D spectrum, find the minimum-phase whitening filter. We begin by discussing the 2-D Hilbert transform method of spectral factorization. This is justifiably considered to be one of the most significant results in 2-D systems theory. (The ironic fact is that it was first reported in 1954 [8]. Apparently it was then forgotten until it was rediscovered in recent years by several other researchers [12], [13], [14].)

As a theoretical tool, the 2-D Hilbert transform method is the means of proving sufficient conditions for a 2-D spectrum to be factorizable, and as a computational tool it is an approximate method of 2-D spectral factorization which has some practical difficulties. The method is applicable if the spectrum is analytic in some neighborhood of the unit circles, and strictly positive on the unit circles. We call any spectrum which satsifies these conditions a <u>positive analytic</u> spectrum.

The 2-D Hilbert transform method is precisely analogous to the 1-D Hilbert transform method. Given a positive analytic spectrum, it can be shown that the complex logarithm of the spectrum is analytic in some neighborhood of the unit circles, and it therefore has a Laurent series expansion in that region [14]:

$$\log S(z_1,z_2) = \sum_{k=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} c_{k,\ell} z_1^{-k} z_2^{-\ell} , \qquad (3.24)$$

where

$$c_{-k,-\ell} = c_{k,\ell}$$

$$= \frac{1}{(2\pi j)^2} \oint_{|z_1|=1} \oint_{|z_2|=1} z_1^{k-1} z_2^{\ell-1} \log S(z_1,z_2) dz_1 dz_2 \quad . \tag{3.25}$$

Or $$\log S(z_1,z_2) = c_{0,0} + C(z_1,z_2) + C(1/z_1,1/z_2), \tag{3.26}$$

where $$C(z_1,z_2) = \sum\sum_{(k,\ell)>(0,0)} c_{k,\ell} z_1^{-k} z_2^{-\ell} \quad . \tag{3.27}$$

Therefore

$$S(z_1,z_2) = \frac{P}{H(z_1,z_2)H(1/z_1,1/z_2)} \quad , \tag{3.28}$$

where $$H(z_1,z_2) = \exp[-C(z_1,z_2)]$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} [-C(z_1,z_2)]^n$$

$$= [1 - \sum\sum_{(k,\ell)>(0,0)} h(k,\ell) z_1^{-k} z_2^{-\ell}] \quad , \tag{3.29}$$

and $$P = \exp[c_{0,0}]$$

$$= \exp[\frac{1}{(2\pi j)^2} \oint_{|z_1|=1} \oint_{|z_2|=1} z_1^{-1} z_2^{-1}$$

$$\log S(z_1,z_2) dz_1 dz_2] \quad . \tag{3.30}$$

Since $C(z_1,z_2)$ is analytic it follows that $H(z_1,z_2)$ is analytic as well, and from the series expansion of the exponential function in (3.29) we see that $H(z_1,z_2)$ is causal, since each term of the uniformly converging series is causal. The inverse of $H(z_1,z_2)$ is also analytic and causal, since it can be written as

$$H^{-1}(z_1,z_2) = \exp[C(z_1,z_2)] \quad . \tag{3.31}$$

As an approximate computational method of 2-D spectral factorization (implemented with 2-D discrete Fourier transforms), the 2-D Hilbert transform method has all of the drawbacks of the 1-D method. In particular if the DFTs are performed with an insufficiently fine sample spacing, the resulting filter may be an inaccurate approximation to the true whitening filter, and it may even be nonminimum-phase.

As we mentioned in the introduction, the root method of factoring a rational spectrum does not extend to the 2-D case. Consider for example the rational spectrum

$$S(z_1,z_2) = (5 + z_1^{-1} + z_1 + z_2^{-1} + z_2) \quad . \tag{3.32}$$

If the spectrum had a rational factorization, then it could be written in the form

$$S(z_1,z_2) = \lambda(z_1,z_2)\lambda(1/z_1,1/z_2) \quad , \tag{3.33}$$

where $$\lambda(z_1,z_2) = [\sum_{\ell=0}^{N} \lambda_{0,\ell} z_2^{-\ell} + \sum_{k=1}^{N} \sum_{\ell=-N}^{N} \lambda_{k,\ell} z_1^{-k} z_2^{-\ell}] \quad , \tag{3.34}$$

for some finite N. But it can be shown that (3.33) cannot be satisfied for any finite value of N; there are always more constraints to satisfy than there are parameters to choose [9]. This is by no means an isolated example; if one chooses a rational 2-D spectrum "at random" then there is virtually no chance of it having a rational factorization.

However considering (3.32), if z_2 is held constant with its magnitude equal to one, then we have a 1-D rational spectrum in z_1. This suggests that we can find a factorization which is finite-order in z_1. Considering in general a positive 2-D spectrum of the form

$$S(z_1,z_2) = \sum_{k=-N}^{N} \sum_{\ell=-M}^{M} r(k,\ell) z_1^{-k} z_2^{-\ell} \quad , \tag{3.35}$$

we have the infinite-order factorization

$$S(z_1,z_2) = \lambda(z_1,z_2)\lambda(1/z_1,1/z_2), \tag{3.36}$$

where $$\lambda(z_1,z_2) = \sum\sum_{(k,\ell)\geq(0,0)} \lambda_{k,\ell} z_1^{-k} z_2^{-\ell} \quad . \tag{3.37}$$

Or $$\lambda(z_1,z_2) = \lambda^{-1}(1/z_1,1/z_2) S(z_1,z_2) \quad . \tag{3.38}$$

Denoting $\lambda^{-1}(z_1,z_2)$ as follows:

$$\lambda^{-1}(z_1,z_2) = \sum\sum_{(k,\ell)\geq(0,0)} \mu_{k,\ell} z_1^{-k} z_2^{-\ell} , \qquad (3.39)$$

we have

$$\lambda(z_1,z_2) = [\sum\sum_{(k,\ell)\leq(0,0)} \mu_{-k,-\ell} z_1^{-k} z_2^{-\ell}]$$

$$\cdot [\sum_{k=-N}^{N} \sum_{\ell=-M}^{M} r(k,\ell) z_1^{-k} z_2^{-\ell}]$$

$$= [\sum\sum_{(k,\ell)\leq(0,0)} \mu_{-k-\ell} z_1^{-k} z_2^{-\ell}]$$

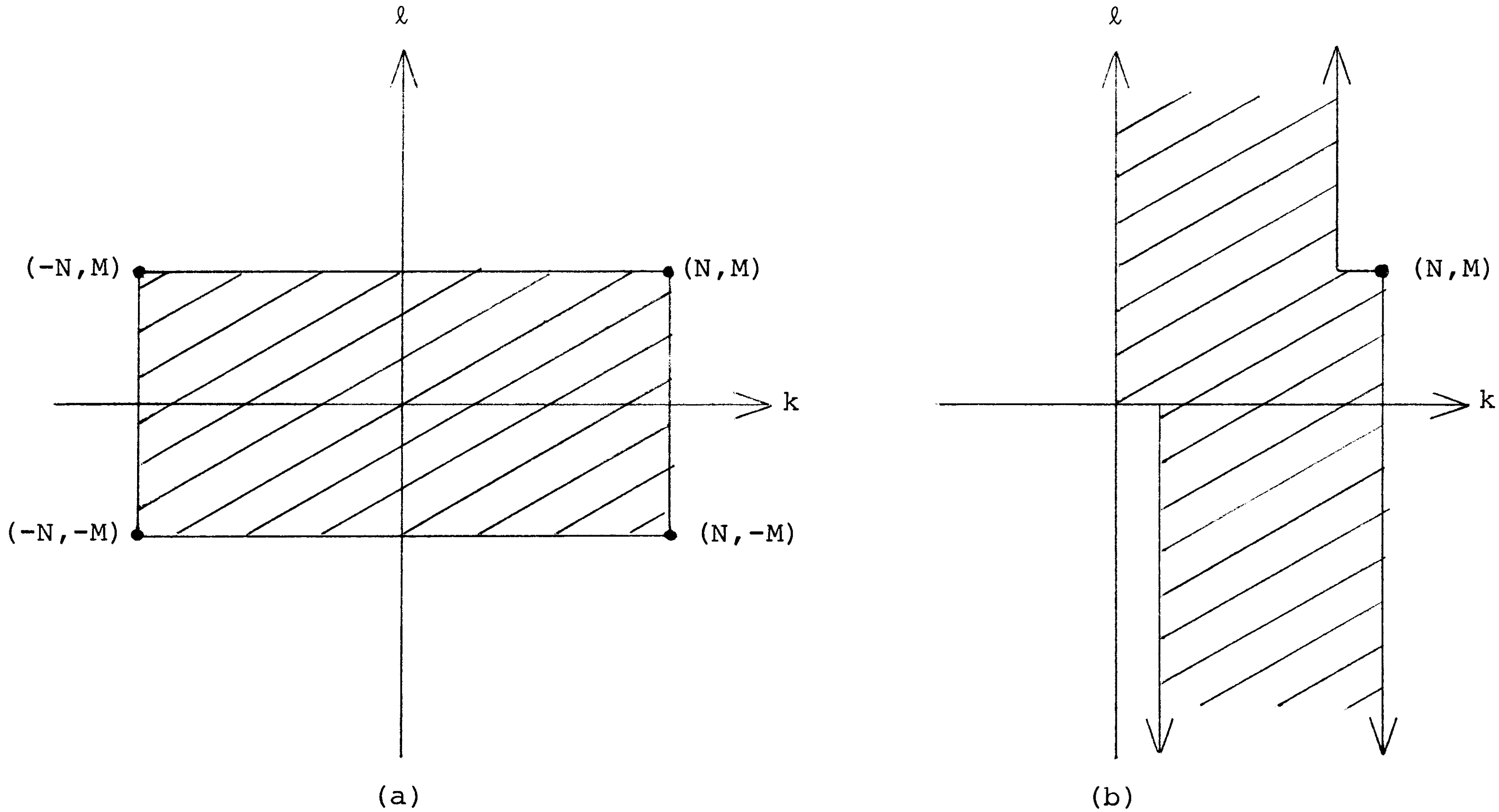
$$\cdot [\sum\sum_{(-N,-M)\leq(k,\ell)\leq(N,M)} r(k,\ell) z_1^{-k} z_2^{-\ell}] . \qquad (3.40)$$

Therefore, $\lambda(z_1,z_2)$ must be of the form

$$\lambda(z_1,z_2) = \sum\sum_{(0,0)\leq(k,\ell)\leq(N,M)} \lambda_{k,\ell} z_1^{-k} z_2^{-\ell} . \qquad (3.41)$$

$S(z_1,z_2)$ and $\lambda(z_1,z_2)$ are illustrated in Fig. 3.1.

We conclude this section by discussing Shanks method of 2-D spectral factorization. Until now there has been no simple explanation as to why this method fails. The basic idea of Shanks method is to approximate the minimum-phase whitening filter by computing an FIR least-squares, linear prediction error filter [10]. For example, consider the linear predictor,



<u>Fig. 3.1</u> The positive spectrum $S(z_1,z_2)$ shown in (a) has the factorization, $S(z_1,z_2) = \lambda(z_1,z_2)\lambda(1/z_1,1/z_2)$, where $\lambda(z_1,z_2)$ is an analytic minimum-phase filter shown in (b).

$$[\hat{x}(k,\ell)\,|\,x(k,\ell-1),x(k-1,\ell),x(k-1,\ell-1)]$$

$$= [h(0,1)x(k,\ell-1) + h(1,0)x(k-1,\ell) + h(1,1)x(k-1,\ell-1)] \ . \tag{3.42}$$

The geometry of the PEF,

$$H(z_1,z_2) = [1 - h(0,1)z_2^{-1} - h(1,0)z_1^{-1} - h(1,1)z_1^{-1}z_2^{-1}] \ , \tag{3.43}$$

is illustrated in Fig. 3.2(a). Applying the Orthogonality Principle, we obtain the set of linear equations that the filter coefficients and prediction error variance must satisfy:

$$\begin{bmatrix} r(0,0) & r(0,1) & r(1,0) & r(1,1) \\ r(0,1) & r(0,0) & r(1,-1) & r(1,0) \\ r(1,0) & r(1,-1) & r(0,0) & r(0,1) \\ r(1,1) & r(1,0) & r(0,1) & r(0,0) \end{bmatrix} \begin{bmatrix} 1 \\ -h(0,1) \\ -h(1,0) \\ -h(1,1) \end{bmatrix} = \begin{bmatrix} P \\ 0 \\ 0 \\ 0 \end{bmatrix} \ . \tag{3.44}$$

If the covariance matrix is positive-definite then there is a unique solution for the filter coefficients and P, and P is positive. But the PEF, $H(z_1,z_2)$, is not always minimum-phase [11]. Moreover, even if the PEF is minimum-phase it can be shown that the transformation between the covariance matrix, and the filter coefficients and P is not invertible. Specifically, an infinite number of

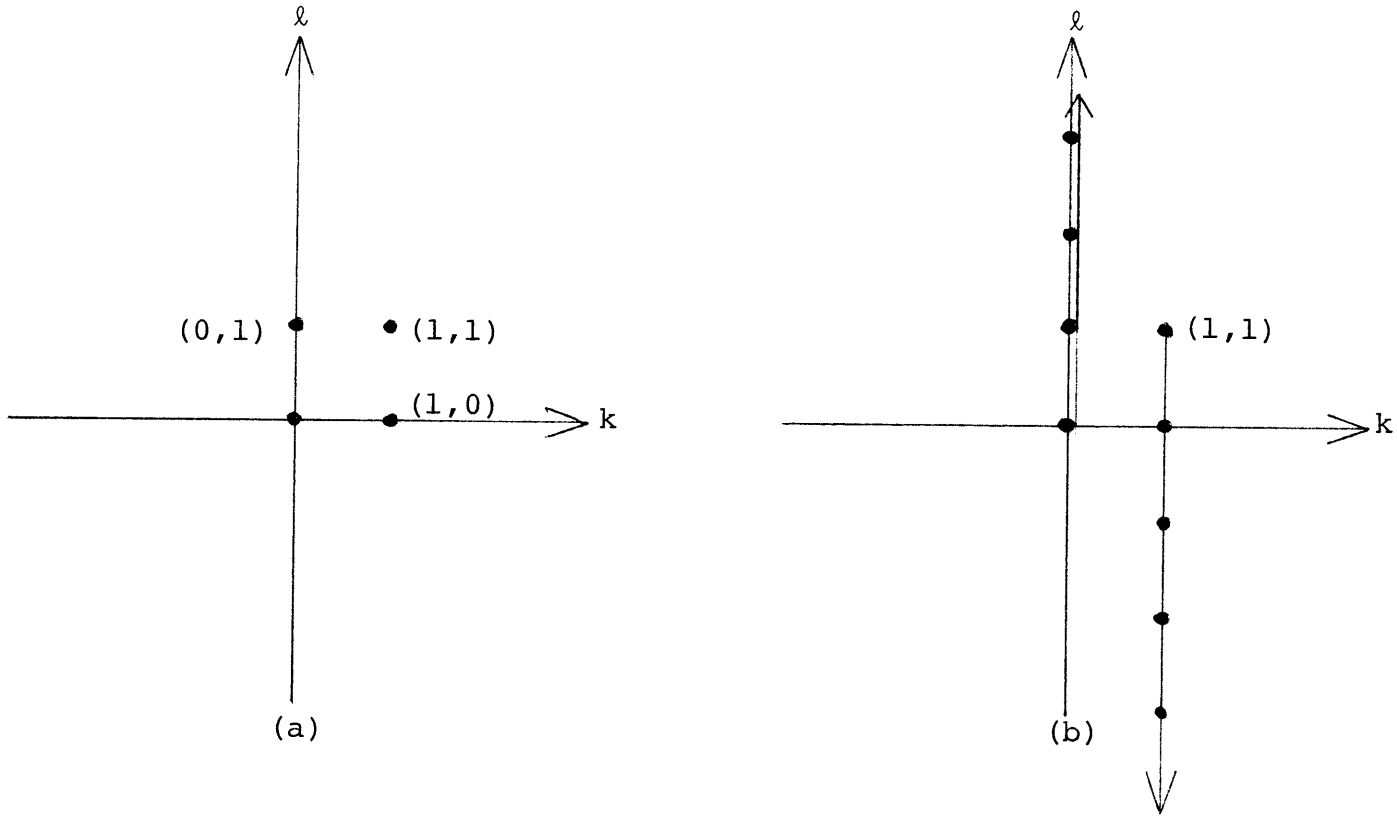


Fig. 3.2 (a) The FIR PEF is not always minimum-phase;
(b) The infinite-order PEF is minimum-phase, given certain conditions on the autocorrelation sequence.

different positive-definite covariance matrices can generate the same PEF and prediction error variance. This is easy to see; referring to (3.44), the covariance matrix contains five different parameters, while the PEF and P together consist of only four parameters. (The five parameters of the covariance matrix are not completely independent since the matrix is required to be positive definite.) We can summarize the failure of Shanks method by saying that Theorem 2.1 fails to extend to 2-D FIR PEFs.

If we examine Shanks method explicitly in terms of our definition of 2-D causality, we can obtain a very simple interpretation for the failure of the method. Essentially we can show that 2-D FIR PEFs are the 2-D analogs to a class of 1-D FIR PEFs which are not guaranteed to be minimum-phase. The predictor (3.42) utilizes three points in the past of $x(k,\ell)$, $\{x(k,\ell-1), x(k-1,\ell), x(k-1,\ell-1)\}$, and the points are ordered as follows:

$$(k-1,\ell-1) < (k-1,\ell) < (k,\ell-1) < (k,\ell) \quad . \qquad (3.45)$$

The predictor does not use the infinite number of points lying between $x(k,\ell-1)$ and $x(k-1,\ell)$:

$$\{x(s,t); (k-1,\ell)<(s,t)<(k,\ell-1)\} \quad .$$

Therefore the data sequence in this prediction problem is "discontinuous." The analogous 1-D situation occurs in the case of the "discontinuous" predictor,

$$[\hat{x}(t)\,|\,x(t-1),x(t-3),x(t-4)] = [h(1)x(t-1) + h(3)x(t-3) + h(4)x(t-4)] \quad , \tag{3.46}$$

where the PEF,

$$H(z) = [1 - h(1)z^{-1} - h(3)z^{-3} - h(4)z^{-4}] \quad , \tag{3.47}$$

is not guaranteed to be minimum-phase.

Our point is that in view of the ordering of points in the plane, 2-D FIR PEFs are not the "natural" 2-D analogs to 1-D FIR PEFs, so we should not be surprised that Theorem 2.1 fails to extend to 2-D FIR PEFs. As we prove in Chapter 4, the PEF,

$$H(z_1,z_2) = [1 - \sum\sum_{(0,0)<(k,\ell)\leq(1,1)} h(k,\ell)z_1^{-k}z_2^{-\ell}] \quad , \tag{3.48}$$

illustrated in Fig. 3.2(b), <u>is</u> minimum-phase if the autocorrelation function satisfies certain conditions. The distinguishing feature of this PEF is that it uses the "continuous" data sequence, $\{x(s,t);\ (k-1,\ell-1)\leq(s,t) <(k,\ell)\}$ to predict $x(k,\ell)$.

3.6 Applications of 2-D Linear Prediction

In this final section we briefly describe some potential applications of 2-D linear prediction. We discuss the problem of 2-D recursive filter design and the 2-D Wiener filtering problem which are applications of spectral factorization, and we discuss 2-D autoregressive spectral estimation and linear predictive coding of images which are applications of autoregressive model fitting.

1. 2-D Recursive Filter Design: We have a specified magnitude-squared frequency response, $S_D(z_1,z_2)$ and the problem is to design a stable, recursive filter with approximately the same magnitude-squared frequency response. 2-D recursive filters are of considerable interest in both image processing and array processing. Denoting the filter input by $y_i(k,\ell)$, and the filter output by $y_0(k,\ell)$, we have

$$y_0(k,\ell) = \sum_s \sum_t h(s,t) y_0(k-s,\ell-t) + \sqrt{P}\; y_i(k,\ell), \tag{3.49}$$

where
$$H(z_1,z_2) = [1 - \sum_s \sum_t h(s,t) z_1^{-s} z_2^{-t}] \tag{3.50}$$

is a minimum-phase, FIR filter. (Given appropriate boundary conditions, the difference equation, (3.49), can be recursively solved [13].) The transfer function of the recursive filter is

$$\frac{Y_0(z_1,z_2)}{Y_i(z_1,z_2)} = \frac{\sqrt{P}}{H(z_1,z_2)} . \tag{3.51}$$

Therefore the design problem is a spectral factorization problem:

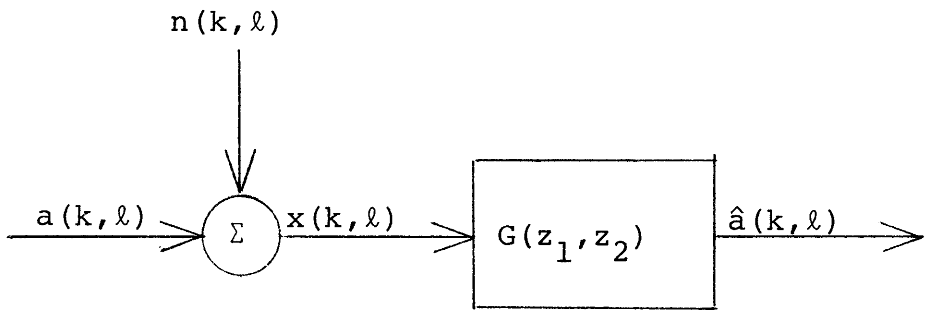
$$S_D(z_1,z_2) \simeq \frac{P}{H(z_1,z_2)H(1/z_1,1/z_2)} . \tag{3.52}$$

2. <u>The 2-D Wiener Filtering Problem</u>: We observe a signal, $x(k,\ell)$, which is the sum of a message, $a(k,\ell)$, and noise, $n(k,\ell)$. We model the message and the noise as wide-sense stationary random processes, and we assume that the power density spectra are known. The problem is to design a filter, $G(z_1,z_2)$, which will give the optimum least-squares linear estimate for the message. The problem is illustrated in Fig. 3.3(a). 2-D Wiener filtering is of interest in image processing as well as array processing.

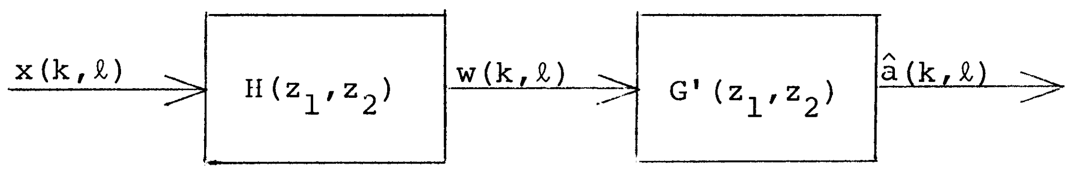
The classical solution to the problem involves first finding the minimum-phase whitening filter for the observed signal by solving the corresponding spectral factorization problem:

$$S_x(z_1,z_2) = \frac{P}{H(z_1,z_2)H(1/z_1,1/z_2)} . \tag{3.53}$$

The idea is that the optimum filter, $G(z_1,z_2)$, can be represented as the product of the whitening filter and



(a)



(b)

Fig. 3.3 a) the 2-D Wiener filtering problem;
b) the classical whitening filter solution;

some other filter, $G'(z_1,z_2)$. This is illustrated in Fig. 3.3(b). Since the minimum-phase whitening filter is causally invertible, no loss in information is incurred by performing the whitening operation. The remaining problem is to design $G'(z_1,z_2)$; this is considerably easier than the original problem, since we are now working with a white-noise process.

In some cases $G(z_1,z_2)$ is required to be causal, and in other cases $G(z_1,z_2)$ can be non-causal. If we further assume that the message and noise are uncorrelated, and that the noise is white with variance N_0, then we can find $G(z_1,z_2)$ explicitly in terms of the whitening filter. In the case where $G(z_1,z_2)$ is causal (the filtering problem), we have

$$G(z_1,z_2) = [1 - \frac{N_0}{P} H(z_1,z_2)] \quad . \tag{3.54}$$

In the case where $G(z_1,z_2)$ is allowed to be non-causal (the smoothing problem), we have

$$G(z_1,z_2) = [1 - \frac{N_0}{P} H(z_1,z_2)H(1/z_1,1/z_2)] \quad . \tag{3.55}$$

3. <u>Autoregressive Spectral Estimation</u>: We are given a finite set of samples from a random process, $x(k,\ell)$, and the idea is to estimate the power density spectrum by modeling the minimum-phase whitening filter as an FIR minimum-phase filter. 2-D autoregressive spectral

estimation would be especially useful in many array processing problems where we wish to find a high-resolution frequency-wavenumber spectral estimate. Equivalently the problem is to fit the following autoregressive model to the data,

$$x(k,\ell) = \sum_s \sum_t \hat{h}(s,t)x(k-s,\ell-t) + w(k,\ell) \quad , \tag{3.56}$$

where $\hat{H}(z_1,z_2) = [1 - \sum_s \sum_t \hat{h}(s,t)z_1^{-s}z_2^{-t}]$ (3.57)

is an FIR minimum-phase filter, and where $w(k,\ell)$ is modelled as a white-noise process with variance $\hat{P}$. Having obtained the model parameters, the autoregressive spectral estimate is given by the formula

$$\hat{S}(z_1,z_2) = \frac{\hat{P}}{\hat{H}(z_1,z_2)\hat{H}(1/z_1,1/z_2)} \quad . \tag{3.58}$$

4. Linear Predictive Coding of Images: We have a sampled image, $x(k,\ell)$, and the idea is to obtain autoregressive models for the image over relatively small regions of the plane. Each piece of the image is then passed through its approximate whitening filter, the whitened image is then transmitted over a communciations channel, and the original image is reconstructed at the other side of the channel. The motivation for using a whitening filter for source encoding is that it removes linear redundancy. The scheme is illustrated in Fig. 3.4.

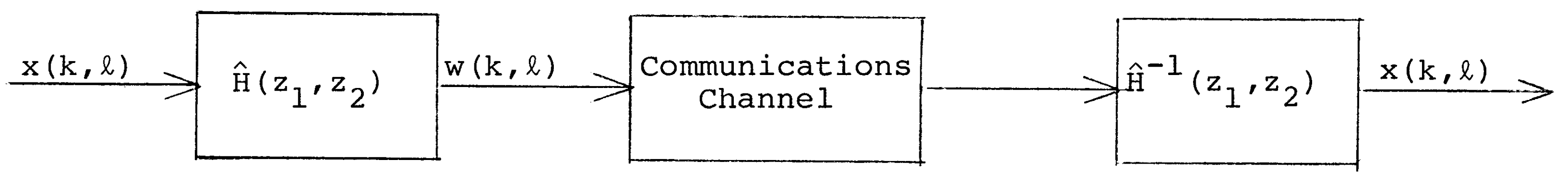


Fig. 3.4 Linear predictive coding of images.

CHAPTER 4

NEW RESULTS IN 2-D LINEAR PREDICTION THEORY

In this chapter we discuss the new 2-D linear prediction theory that has been developed in this thesis research. We extend all of the theorems of Chapter 2 to the 2-D case.

We begin by proving that a particular class of 2-D PEFs is always minimum-phase if the power spectrum is positive analytic. Unfortunately, these filters are infinite-order in z_2 so they cannot be implemented in practice. Nevertheless, we show that any such PEF can be approximated arbitrarily closely by an FIR minimum-phase filter which, in turn, can be represented in terms of a finite set of numbers that we refer to as reflection coefficients.

The practical implication of these theoretical results is that approximate 2-D minimum-phase whitening filters can be designed in the 2-D reflection coefficient domain. We show that if the reflection coefficient magnitudes are less than one, the difficult minimum-phase requirement is automatically satisfied.

4.1 The Correspondence between 2-D Positive-Definite Analytic Autocorrelation Sequences and 2-D Analytic Minimum-phase PEFs

We begin by considering the following least-squares linear prediction problem

$$[\hat{x}(k,\ell)\,|\,\{x(s,t);\ (k-N,\ell-M)\leq(s,t)<(k,\ell)\}]$$

$$= \sum\sum_{0<(s,t)\leq(N,M)} h(N,M;s,t)\,x(k-s,\ell-t) \quad , \quad (N,M)>(0,0) \quad . \tag{4.1}$$

The geometry of the problem is illustrated in Fig. 1.4. We denote the prediction error variance by $P_{N,M}$, and we denote the PEF by

$$H_{N,M}(z_1,z_2) = [1 - \sum\sum_{(0,0)<(k,\ell)\leq(N,M)} h(N,M;k,\ell)\,z_1^{-k}z_2^{-\ell}] \quad . \tag{4.2}$$

The PEF is illustrated in Fig. 1.3.

We apply the Orthogonality Principle to obtain the normal equations that the filter coefficients and $P_{N,M}$ must satisfy:

$$[r(s,t) - \sum\sum_{(0,0)<(k,\ell)\leq(N,M)} h(N,M;k,\ell)\,r(s-k,t-\ell)] = P_{N,M}\delta_s\delta_t \quad ,$$

$$(0,0)\leq(s,t)\leq(N,M) \quad . \tag{4.3}$$

The normal equations are an infinite set of linear equations, and unless we impose certain conditions on the autocorrelation sequence, there is no guarantee that there is a stable solution. To obtain our results we require that the 2-D autocorrelation sequence be <u>positive-definite</u>

and <u>analytic</u>. We say that the autocorrelation sequence, $\{r(k,\ell);\ (0,0)\leq(k,\ell)\leq(N,M)\}$, is positive-definite and analytic if

1) $r(k,\ell)$ decays at least exponentially fast to zero as ℓ goes to plus or minus infinity;

2) the following Toeplitz matrix is strictly positive-definite for $|z_2| = 1$:

$$\begin{bmatrix} R_0(z_2) & R_1(1/z_2) & R_2(1/z_2) & \cdots & R_{N-1}(1/z_2) \\ R_1(z_2) & R_0(z_2) & R_1(1/z_2) & \cdots & R_{N-2}(1/z_2) \\ \vdots & & & & \\ R_{N-1}(z_2) & R_{N-2}(z_2) & R_{N-3}(z_2) & \cdots & R_0(z_2) \end{bmatrix}, \tag{4.4}$$

where $$R_k(z_2) = \sum_{\ell=-\infty}^{\infty} r(k,\ell)\, z_2^{-\ell} \; ; \tag{4.5}$$

3) the prediction error variance associated with any FIR PEF having the same support as $H_{N,M}(z_1,z_2)$ has a positive lower bound.

We note that the first condition implies that the $R_k(z_2)$ are analytic in the neighborhood of the unit circle. For $|z_2| = 1$, the Toeplitz matrix (4.4) is Hermitian. Finally we observe that a sufficient (but not necessary) condition for the autocorrelation sequence to be positive-definite and analytic is that the power density spectrum of the random process is positive analytic.

We can now state the following theorem.

Theorem 4.1(a): Given a positive-definite analytic autocorrelation sequence, $\{r(k,\ell);\ (0,0)\leq(k,\ell)\leq(N,M)\}$, then:

1) the normal equations (4.3) have a unique solution for $H_{N,M}(z_1,z_2)$ and $P_{N,M}$;

2) $P_{N,M}$ is positive;

3) $H_{N,M}(z_1,z_2)$ is analytic and minimum-phase.

Outline of Proof: The fact that $P_{N,M}$ is positive follows trivially from the assumption that the autocorrelation sequence is positive-definite and analytic. The uniqueness part of the theorem is easy to prove; the existence of more than one solution to the normal equations would imply the existence of a PEF with zero prediction error.

The difficult parts of the proof are the existence proof, and the proof that the PEF is analytic minimum-phase. The outline of the proofs is as follows (the details are in Appendix A1):

1) We first prove that a solution for a lower-order PEF,

$$H_{N,-\infty}(z_1,z_2) = H_{N-1,+\infty}(z_1,z_2) \quad , \tag{4.6}$$

exists, and that the PEF is analytic minimum-phase. Due to the simple structure of the filter, the solution can be obtained by working with transforms in z_2.

2) We constructively prove, for all sufficiently small values of m ($m \to -\infty$), that there is an analytic solution for $H_{N,m}(z_1,z_2)$. The solution is a Neumann series solution involving the lower-order PEF, $H_{N,-\infty}(z_1,z_2)$.

3) As part of the Neumann series solution, we prove that the sequence of PEFs, $H_{N,m}(z_1,z_2)$, converges uniformly, in the neighborhood of the unit circles, to the limit filter, $H_{N,-\infty}(z_1,z_2)$, as m goes to minus infinity:

$$\lim_{m \to -\infty} H_{N,m}(z_1,z_2) = H_{N,-\infty}(z_1,z_2) \quad . \tag{4.7}$$

Applying Theorem 3.1 we can then prove that for all sufficiently small values of m, $H_{N,m}(z_1,z_2)$ is analytic minimum-phase.

4) Finally, using a 2-D version of the Levinson algorithm, we precursively obtain a solution for $H_{N,m}(z_1,z_2)$ for <u>all</u> $m \le M$. Using Theorem 3.1 in conjunction with the 2-D Levinson algorithm, we inductively prove that the $H_{N,m}(z_1,z_2)$ are analytic minimum-phase.

Theorem 4.1(a) has a converse.

Theorem 4.1(b): Given any positive $P_{N,M}$, and any analytic mimimum-phase filter, $H_{N,M}(z_1,z_2)$, for a particular (N,M) where

$$H_{N,M}(z_1,z_2) = [1 - \sum\sum_{(0,0)<(k,\ell)\leq(N,M)} h(N,M;k,\ell) z_1^{-k} z_2^{-\ell}] \quad , \tag{4.8}$$

there is a unique positive-definite analytic autocorrelation sequence, $\{r(k,\ell);\ (0,0)\leq(k,\ell)\leq(N,M)$, such that the normal euqations (4.3) are satisfied. The autocorrelation sequence is given by the formula

$$r(k,\ell) = \frac{1}{(2\pi j)^2} \oint_{|z_1|=1} \oint_{|z_2|=1} z_1^{k-1} z_2^{\ell-1} \cdot \frac{P_{N,M} dz_1 dz_2}{H_{N,M}(z_1,z_2) H_{N,M}(1/z_1,1/z_2)} \quad ,$$

$$(0,0)\leq(k,\ell)\leq(N,M) \quad . \tag{4.9}$$

Outline of Proof: The existence part of the proof is trivial; the quantity

$$\frac{P_{N,M}}{H_{N,M}(z_1,z_2) H_{N,M}(1/z_1,1/z_2)} \tag{4.10}$$

is a positive analytic spectrum which is already in factored form. Therefore $H_{N,M}(z_1,z_2)$ is the minimum-phase whitening filter for a random process having the spectrum (4.10).

Recalling the linear prediction interpretation for the minimum-phase whitening filter it is clear that (4.9) is satisfied.

The uniqueness part of Theorem 4.1(b) is comparatively difficult to prove, and it is discussed in Appendix A1.

An interesting interpretation of Theorem 4.1 is that it specifies a method of extrapolating a particular class of 2-D autocorrelation sequences. We begin with the positive-definite analytic autocorrelation sequence, $\{r(k,\ell);\ (0,0)\leq(k,\ell)\leq(N,M)\}$, and we compute $H_{N,M}(z_1,z_2)$ and $P_{N,M}$ by solving the normal equations. We then form the spectrum (4.10). The inverse Z-transform of this spectrum is an autocorrelation function, $r(k,\ell)$, which is equal to the original autocorrelation sequence for $(0,0)\leq(k,\ell)\leq(N,M)$, and which is an extrapolation of the autocorrelation sequence for $(k,\ell)>(N,M)$. The extrapolation is the <u>maximum-entropy</u> extrapolation; it can be shown that of all spectra which agree with the original autocorrelation sequence, the spectrum (4.10) has the greatest entropy, where the entropy is given by the formula

$$\frac{1}{(2\pi j)^2} \oint_{|z_1|=1} \oint_{|z_2|=1} z_1^{-1} z_2^{-1} \log S(z_1,z_2)\,dz_1 dz_2 \quad . \tag{4.11}$$

As we mentioned earlier, an important step in the proof of Theorem 4.1 depends on a 2-D version of the Levinson algorithm.

<u>Theorem 4.2 (2-D Levinson algorithm)</u>: Suppose that we have a positive-definite analytic autocorrelation sequence, $\{r(k,\ell);\ (0,0)\leq(k,\ell)\leq(N,M)\}$. We further assume that we have the solution to the normal equations for $H_{n,m-1}(z_1,z_2)$ and $P_{n,m-1}$ for some $(n,m-1)$ where $(0,0)\leq(n,m-1)<(N,M)$, and that $H_{n,m-1}(z_1,z_2)$ is stable. Then the solution for $H_{n,m}(z_1,z_2)$ and $P_{n,m}$ is given by the formulas

$$H_{n,m}(z_1,z_2) = [H_{n,m-1}(z_1,z_2) - \rho(n,m)z_1^{-n}z_2^{-m}H_{n,m-1}(1/z_1,1/z_2)] \quad , \tag{4.12}$$

and $P_{n,m} = P_{n,m-1}[1 - \rho^2(n,m)]$, (4.13)

where

$$\rho(n,m) = \frac{1}{P_{n,m-1}} E\{[x(k,\ell) - \sum\sum_{(0,0)<(s,t)\leq(n,m-1)} h(n,m-1;s,t)x(k-s,\ell-t)]$$

$$\cdot\ [x(k-n,\ell-m) - \sum\sum_{(0,0)<(s,t)\leq(n,m-1)} h(n,m-1;s,t)$$

$$\cdot\ x(k-n+s,\ell-m+t)]\}$$

$$= \frac{1}{P_{n,m-1}} [r(n,m) - \sum\sum_{(0,0)<(s,t)\leq(n,m-1)} h(n,m-1;s,t)r(n-s,m-t)] \ . \tag{4.14}$$

Furthermore, if $H_{n,m-1}(z_1,z_2)$ is analytic minimum-phase, then $H_{n,m}(z_1,z_2)$ is also analytic minimum-phase.

Proof: Equation (4.12) can be written algebraically as follows:

$$h(n,m;n,m) = \rho(n,m) \quad , \tag{4.15}$$

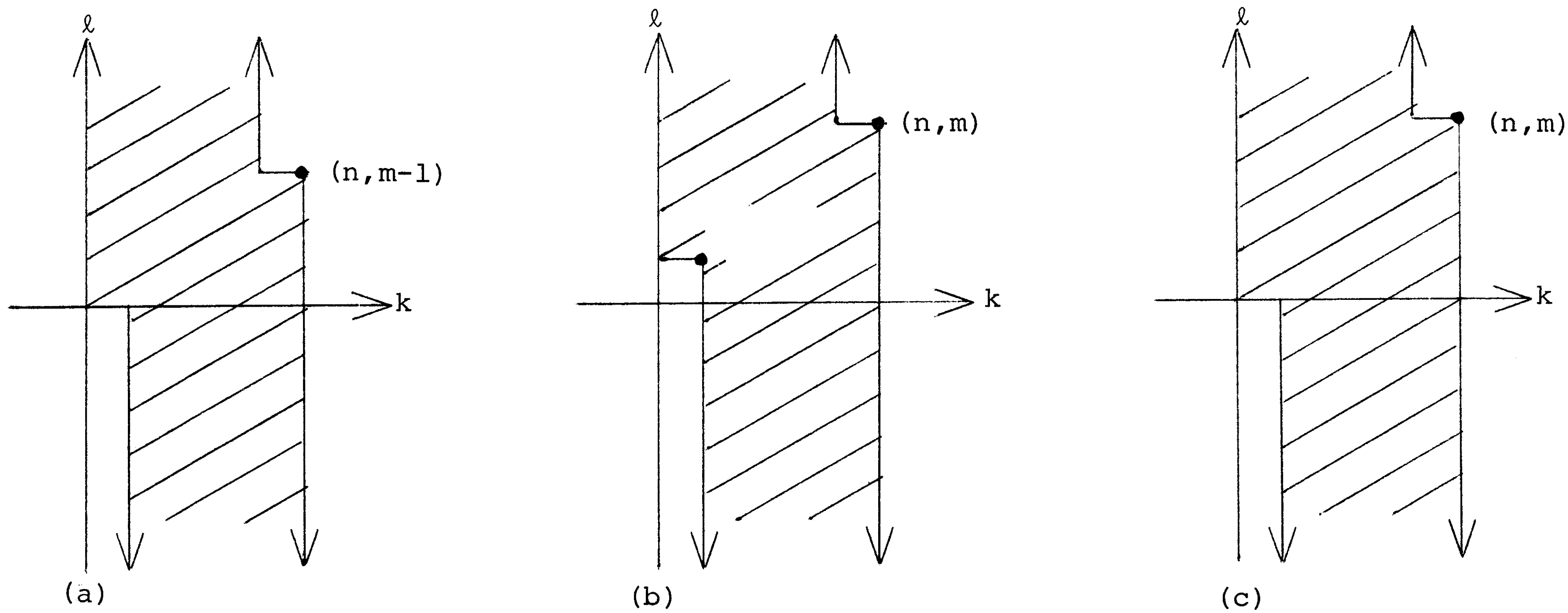
$$h(n,m;k,\ell) = [h(n,m-1;\ k,\ell) - \rho(n,m)h(n,m-1;n-k,m-\ell)] \ , \quad (0,0)<(k,\ell)\leq(n,m-1) \ . \tag{4.16}$$

The key idea of the 2-D Levinson algorithm is that the "new" PEF, $H_{n,m}(z_1,z_2)$, is equal to a linear combination of the "old" forward PEF, $H_{n,m-1}(z_1,z_2)$, and the "old" delayed backward PEF, $z_1^{-n}z_2^{-m}H_{n,m-1}(1/z_1,1/z_2)$. The geometry of the recursion is illustrated in Fig. 4.1.

Given that $H_{n,m-1}(z_1,z_2)$ and $P_{n,m-1}$ satisfy the "old" normal equations,

$$[r(s,t) - \sum\sum_{(0,0)<(k,\ell)\leq(n,m-1)} h(n,m-1;k,\ell)r(s-k,t-\ell)]$$

$$= P_{n,m-1}\delta_s\delta_t \quad , \quad (0,0)\leq(s,t)\leq(n,m-1) \quad , \tag{4.17}$$



<u>Fig. 4.1</u> Geometry of 2-D Levinson algorithm;

a) $H_{n,m-1}(z_1,z_2)$; b) $[z_1^{-n}z_2^{-m}H_{n,m-1}(1/z_1,1/z_2)]$;

c) $H_{n,m}(z_1,z_2)$

it is a matter of straightforward substitution to show that the solution given by (4.12), (4.13), and (4.14) satisfies the "new" normal equations,

$$[r(s,t) - \sum\sum_{(0,0)<(k,\ell)\leq(n,m)} h(n,m;k,\ell)\, r(s-k,t-\ell)] = P_{n,m}\delta_s\delta_t \quad ,$$

$$(0,0)\leq(s,t)\leq(n,m) \quad . \qquad (4.18)$$

The number $\rho(n,m)$ is called a "reflection coefficient," and its magnitude is always less than one; this can be seen from (4.13) and the fact that $P_{n,m-1}$ and $P_{n,m}$ are positive. (Alternatively, we see from (4.14) that $\rho(n,m)$ is a partial correlation coefficient.)

Considering (4.12) we see that for all values of z_1 and z_2 on the unit circles,

$$|H_{n,m-1}(z_1,z_2)|>|\rho(n,m)z_1^{-n}z_2^{-m}H_{n,m-1}(1/z_1,1/z_2)| \quad . \qquad (4.19)$$

Therefore if $H_{n,m-1}(z_1,z_2)$ is analytic minimum-phase, then Theorem 3.1 implies that $H_{n,m}(z_1,z_2)$ is also analytic minimum-phase.

As we pointed out in section 2.1, the 1-D Levinson algorithm has a physical interpretation related to the propagation of a wave in a layered medium. Unfortunately there is no such physical interpretation for the 2-D

Levinson algorithm. While it should be possible to find a formal scattering theory interpretation for the 2-D Levinson algorithm, it would not correspond to any physical scattering mechanism because the propagation would follow an infinite raster scan.

4.2 A Canonical Representation for 2-D Analytic Minimum-phase Filters

Recalling the 2-D Levinson algorithm, and in particular equation (4.12), we expect to find a reflection coefficient representation for 2-D analytic minimum-phase filters. The basic idea of the representation is that associated with every analytic minimum-phase filter, $H_{N,M}(z_1,z_2)$, is a unique 2-D reflection coefficient sequence, $\{\rho(k,\ell);\ (0,0)<(k,\ell)\leq(N,M)\}$, where the reflection coefficient magnitudes are less than one, and the reflection coefficients decay exponentially fast to zero as ℓ goes to plus or minus infinity. Conversely, given any such reflection coefficient sequence, there is a unique analytic minimum-phase filter. The geometry of the filter and the reflection coefficient sequence is illustrated in Fig. 1.3 and Fig. 1.4.

The importance of the 2-D reflection coefficient representation cannot be overemphasized. As we demonstrate later, it is a computationally useful representation for 2-D minimum-phase filters.

Theorem 4.3(a): Given any 2-D reflection coefficient sequence, $\{\rho(k,\ell);\ (0,0)<(k,\ell)\leq(N,M)\}$, such that

$$|\rho(k,\ell)| < (1+\varepsilon)^{-|\ell|} \quad , \quad \varepsilon>0 \quad , \tag{4.20}$$

there is a unique 2-D sequence of 2-D analytic minimum-phase filters, $\{H_{n,m}(z_1,z_2);\ (0,0)\leq(n,m)\leq(N,M)\}$, of the form

$$H_{n,m}(z_1,z_2) = [1 - \sum\sum_{(0,0)<(k,\ell)\leq(n,m)} h(n,m;k,\ell)\, z_1^{-k} z_2^{-\ell}] \quad , \tag{4.21}$$

such that:

1) $$H_{0,0}(z_1,z_2) = 1 \quad ; \tag{4.22}$$

2) $$H_{n,m}(z_1,z_2) = [H_{n,m-1}(z_1,z_2) - \rho(n,m)\, z_1^{-n} z_2^{-m} H_{n,m-1}(1/z_1,1/z_2)] \quad , \quad (0,0)<(n,m)\leq(N,M) \quad ; \tag{4.23}$$

3) $$H_{n,-\infty}(z_1,z_2) = H_{n-1,+\infty}(z_1,z_2) \quad , \quad 1\leq n\leq N \quad ; \tag{4.24}$$

4) $$\lim_{m\to+\infty} H_{n,m}(z_1,z_2) = H_{n,+\infty}(z_1,z_2) \quad , \quad 0\leq n\leq N-1 \tag{4.25}$$

(the convergence is uniform in some neighborhood of the unit circles);

5) $\lim_{m \to -\infty} H_{n,m}(z_1,z_2) = H_{n,-\infty}(z_1,z_2), \quad 1 \le n \le N$; (4.26)

(the convergence is uniform in some neighborhood of the unit circles).

Theorem 4.3(b): Given any analytic minimum-phase filter, $H_{N,M}(z_1,z_2)$, of the form (4.21), there is a unique reflection coefficient sequence, $\{\rho(k,\ell);\ (0,0)<(k,\ell)\le(N,M)\}$, and a unique 2-D sequence of 2-D analytic minimum-phase filters, $\{H_{n,m}(z_1,z_2);\ (0,0)\le(n,m)\le(N,M-1)\}$, such that equations (4.20) - (4.26) are satisfied.

Given the reflection coefficient sequence, we can compute the filter sequence, $H_{n,m}(z_1,z_2)$, recursively (at least conceptually); the order in which the filters are computed follows an infinite master scan. We first recursively compute the $H_{0,m}(z_1,z_2)$ beginning with $H_{0,0}(z_1,z_2) = 1$, and ending with $H_{0,+\infty}(z_1,z_2)$, using (4.23). We then recursively compute the $H_{1,m}(z_1,z_2)$ beginning with the initial condition, $H_{1,-\infty}(z_1,z_2) = H_{0,+\infty}(z_1,z_2)$, and ending with $H_{1,+\infty}(z_1,z_2)$. The remainder of the recursion follows in exactly the same manner.

A particularly interesting situation occurs when the reflection coefficient sequence is finite-order. In that case, the recursion follows a finite raster scan, all of the $H_{n,m}(z_1,z_2)$ are finite-order, and the entire recursion can be performed with a finite number of computations.

Given the filter, $H_{N,M}(z_1,z_2)$, the reflection coefficient sequence and the filter sequence, $\{H_{n,m}(z_1,z_2);\ (0,0)\le(n,m)<(N,M)\}$, can be recovered by running (4.23) "backwards." It is straightforward to show that

$$\rho(n,m) = h(n,m;n,m) \quad , \quad (0,0)<(n,m)\le(N,M) \quad , \tag{4.27}$$

and

$$H_{n,m-1}(z_1,z_2) = \frac{1}{[1-\rho^2(n,m)]} [H_{n,m}(z_1,z_2 + \rho(n,m) \cdot z_1^{-n} z_2^{-m} H_{n,m}(1/z_1,1/z_2)] \quad , \quad (0,0)<(n,m)\le(N,M). \tag{4.28}$$

Using (4.27) and (4.28), $H_{n,m}(z_1,z_2)$ and $\rho(n,m)$ can be recursively computed. Again, the recursion follows an infinite raster scan. In this case, however, the recursion begins with $(n,m)=(N,M)$, and it propagates backwards towards the origin.

We pointed out that if the reflection coefficient sequence is finite-order then $H_{N,M}(z_1,z_2)$ is also finite order. Unfortunately, the converse is not true: if $H_{N,M}(z_1,z_2)$ is finite-order, then the reflection coefficient sequence is almost always infinite-order.

Before we outline the proofs of these theorems it is instructive to consider two numerical examples.

Example 4.1: This example illustrates the important property of the 2-D reflection coefficient representation that if the reflection coefficient sequence is finite-order then the corresponding minimum-phase filter is finite-order, and it can be obtained from the reflection coefficients by a finite number of computations. We begin with the following reflection coefficient sequence:

$$\rho(k,\ell) = \begin{cases} 1/2 \ , & (k,\ell) = (0,1) \\ -1/4 \ , & (k,\ell) = (1,-1) \\ 1/3 \ , & (k,\ell) = (1,0) \\ 0 \ , & \text{elsewhere} \ . \end{cases}$$

According to Theorem 4.3(a) we generate a 2-D sequence of 2-D minimum-phase filters as follows:

1) $$H_{0,0}(z_1,z_2) = 1 \quad ; \qquad (4.29)$$

2) $$H_{0,1}(z_1,z_2) = H_{0,0}(z_1,z_2) - \rho(0,1)z_2^{-1}H_{0,0}(1/z_1,1/z_2)$$

$$= (1 - \tfrac{1}{2}\, z_2^{-1}) \quad ; \qquad (4.30)$$

3) $$H_{n,m}(z_1,z_2) = H_{0,1}(z_1,z_2) = (1 - \tfrac{1}{2}\, z_2^{-1}) \ ,$$

$$(0,2) \le (n,m) \le (1,-2) \quad ; \qquad (4.31)$$

4) $$H_{1,-1}(z_1,z_2) = H_{1,-2}(z_1,z_2) - \rho(1,-1)z_1^{-1}z_2H_{1,-2}(1/z_1,1/z_2)$$
$$= (1 - \frac{1}{2}z_2^{-1} - \frac{1}{8}z_1^{-1}z_2^2 + \frac{1}{4}z_1^{-1}z_2) \quad ; \tag{4.32}$$

5) $$H_{1,0}(z_1,z_2) = H_{1,-1}(z_1,z_2) - \rho(1,0)z_1^{-1}H_{1,-1}(1/z_1,1/z_2)$$
$$= 1 - \frac{7}{12}z_2^{-1} + \frac{1}{24}z_2^{-2} - \frac{1}{8}z_1^{-1}z_2^2 + \frac{5}{12}z_1^{-1}z_2 - \frac{1}{3}z_1^{-1}) \quad . \tag{4.33}$$

The reflection coefficient sequence and $H_{1,0}(z_1,z_2)$ are illustrated in Fig. 4.2.

Example 4.2: The following example illustrates the fact that the reflection coefficient sequence associated with a 2-D FIR minimum-phase filter is almost always infinite-order. We begin with the minimum-phase filter

$$H_{1,0}(z_1,z_2) = (1 + \frac{1}{4}z_2^{-1} + \frac{1}{4}z_1^{-1}) \quad . \tag{4.34}$$

According to Theorem 4.3(b), there is a unique reflection coefficient sequence associated with this filter. We

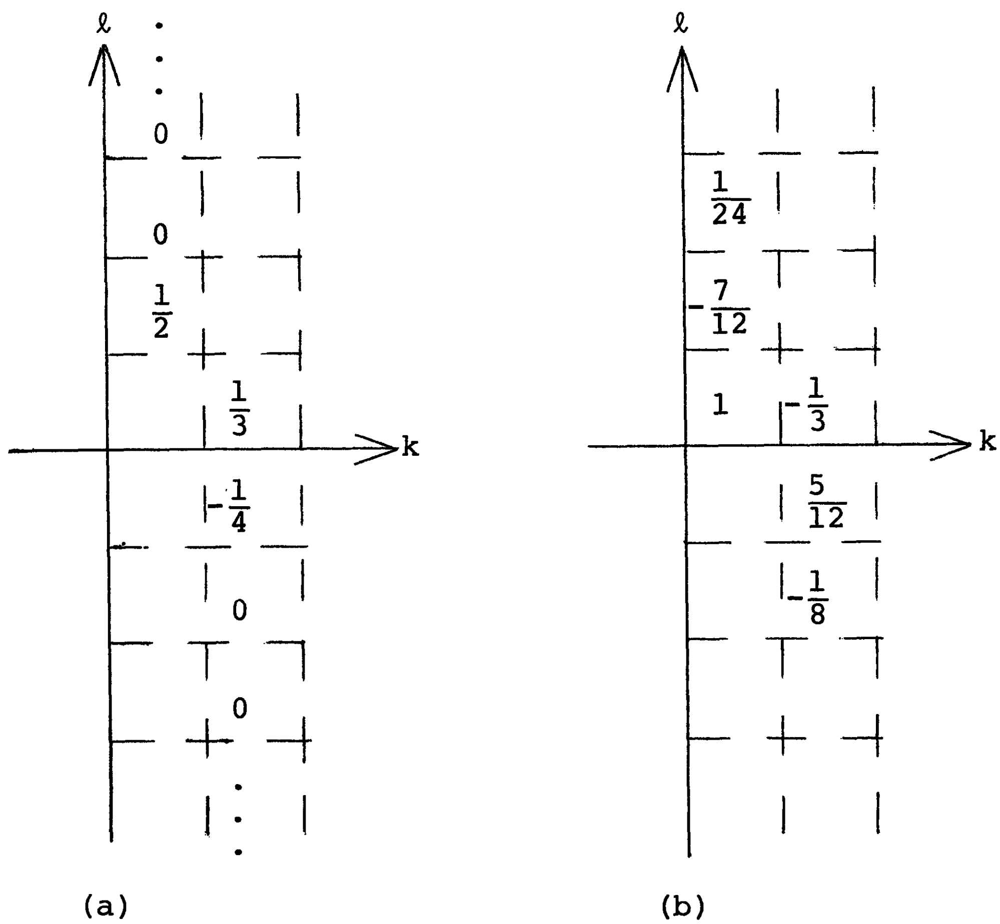


Fig. 4.2 The finite-order reflection coefficient sequence (a) generates the minimum-phase filter (b).

obtain the reflection coefficient sequence by means of (4.27) and (4.28). Applying these formulas to the filter (4.34) we have

1) $$\rho(1,0) = h(1,0;1,0) = -\frac{1}{4} \; ; \tag{4.35}$$

2) $$H_{1,-1}(z_1,z_2) = \frac{1}{[1-\rho^2(1,0)]}\,[H_{1,0}(z_1,z_2) + \rho(1,0)z_1^{-1}H_{1,0}(\frac{1}{z_1}, \frac{1}{z_2})]$$

$$= (1 + \frac{4}{15}z_2^{-1} - \frac{1}{15}z_1^{-1}z_2) \; ; \tag{4.36}$$

$$\rho(1,-1) = h(1,-1;1,-1) = \frac{1}{15} \; ; \tag{4.37}$$

3) $$H_{1,-2}(z_1,z_2) = \frac{1}{[1-\rho^2(1,-1)]}\,[H_{1,-1}(z_1,z_2) + \rho(1,-1)z_1^{-1}z_2H_{1,-1}(\frac{1}{z_1}, \frac{1}{z_2})]$$

$$= (1 + \frac{15}{56}z_2^{-1} + \frac{1}{56}z_1^{-1}z_2^2) \; ; \tag{4.38}$$

$$\rho(1,-2) = -\frac{1}{56} \; ; \tag{4.39}$$

4) We can show that for all $m \leq -2$,

$$\rho(1,m) = \frac{\rho^2(1,m+1)}{\rho(1,m+2)[1-\rho^2(1,m+1)]} \quad , \tag{4.40}$$

so clearly $\rho(1,m)$ is non-zero for all finite values of m less than or equal to zero.

5) In the course of deriving (4.40) we can show that

$$H_{1,m}(z_1,z_2) = [1 - h(1,m;0,1)z_2^{-1} - h(1,m;1,m)z_1^{-1}z_2^{-m}] \quad , \quad m \leq 0 \quad . \tag{4.41}$$

Therefore $\rho(0,\ell) = 0 \quad , \quad \ell \geq 2 \quad .$ (4.42)

6) Finally, employing the constructive argument that we use to prove the existence part of Theorem 4.3(b), we can show that

$$\rho(0,1) = (\sqrt{3} - 2) \quad . \tag{4.43}$$

The filter, and its reflection coefficient sequence are illustrated in Fig. 4.3.

Outline of Proof of Theorem 4.3(a):

Given the reflection coefficient sequence, $\{\rho(k,\ell);\ (0,0)<(k,\ell)\leq(N,M)\}$, the key idea of the proof is to work with a "truncated" reflection coefficient sequence,

$$\{\rho^{(L)}(k,\ell);\ (0,0)<(k,\ell)\leq(N,M)\} \quad , \quad L\geq|M| \quad , \tag{4.44}$$

where

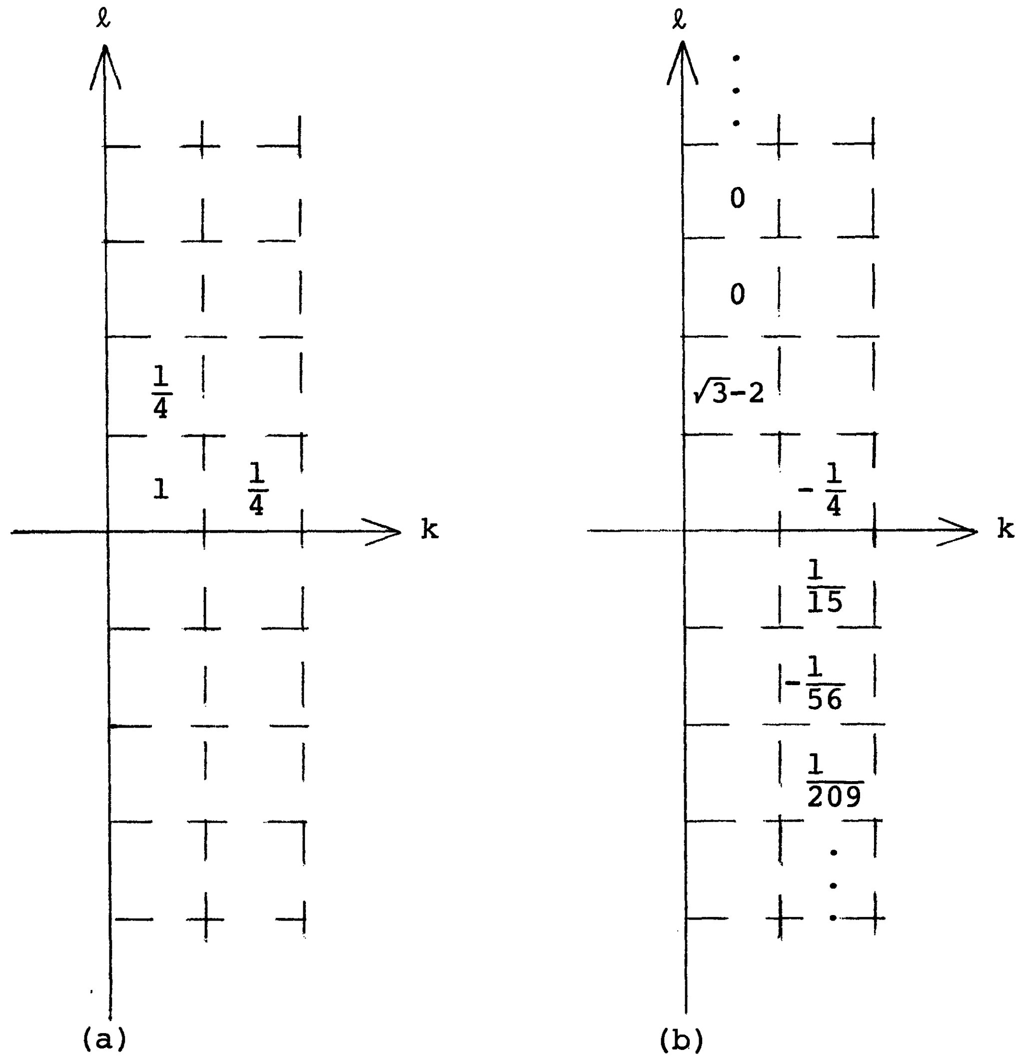


Fig. 4.3 The FIR minimum-phase filter (a), generates the ifninite-order reflection coefficient sequence (b).

$$\rho^{(L)}(k,\ell) = \begin{cases} \rho(k,\ell) & , \quad |\ell| \le L \\ 0 & , \quad \text{elsewhere} \end{cases} \quad . \tag{4.45}$$

The geometry of the truncated reflection coefficient sequence is illustrated in Fig. 4.4. Associated with the reflection coefficient sequence (4.44) is a 2-D sequence of FIR minimum-phase filters, $\{H_{n,m}^{(L)}(z_1,z_2);\ (0,0)\le(n,m) \le(N,M)\}$, which, as in Example 4.1, can be obtained by a finite number of computations. We can prove that as L goes to infinity, the sequence of filters, $H_{n,m}^{(L)}(z_1,z_2)$, converges uniformly in the neighborhood of the unit circles to a limit sequence, $H_{n,m}(z_1,z_2)$. Specifically we prove that for any $\delta>0$ there is a number L such that for all $L_1\ge L$, $L_2\ge L$, and for all $(1-\varepsilon)<|z_1|,|z_2|<(1+\varepsilon)$ where ε is a number independent of δ,

$$|H_{n,m}^{(L_1)}(z_1,z_2) - H_{n,m}^{(L_2)}(z_1,z_2)|<\delta \quad . \tag{4.46}$$

The details of the proof, and the proof for the uniqueness part of the theorem are in Appendix A2.

<u>Outline of Proof of Theorem 4.3(b)</u>: The existence part of Theorem 4.3(b) is almost a direct consequence of Theorem 4.1. Given an analytic minimum-phase filter, $H_{N,M}(z_1,z_2)$, we can choose some arbitrary positive $P_{N,M}$. According to Theorem 4.1(b), there is a unique positive-definite analytic autocorrelation sequence associated with

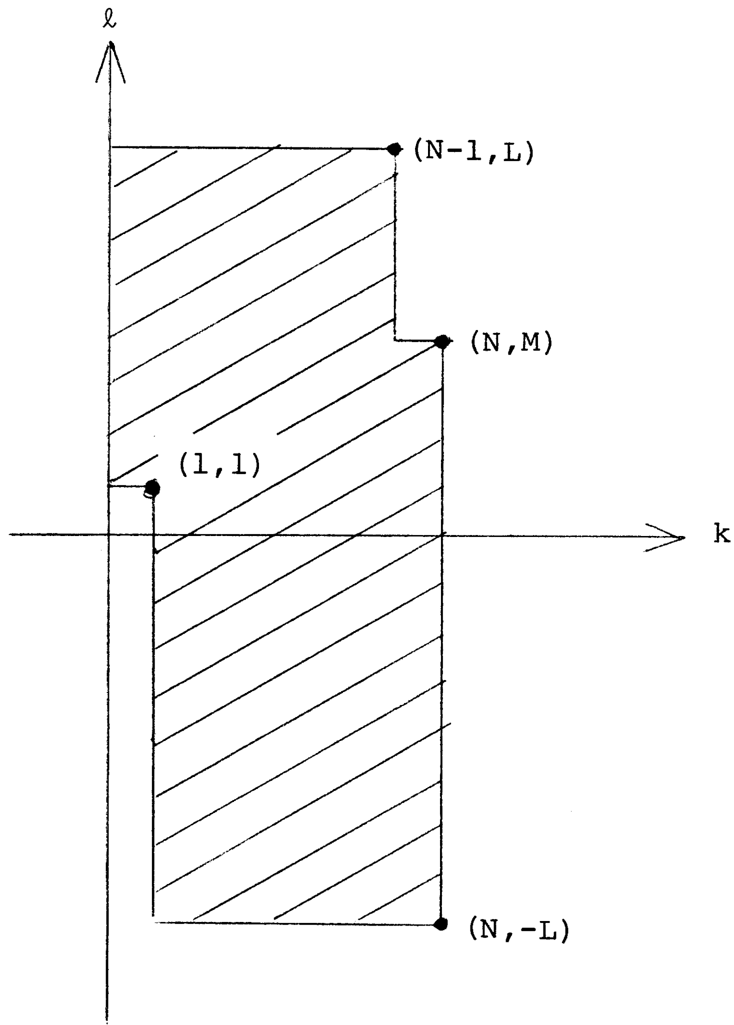


Fig. 4.4 Any analytic minimum-phase filter, $H_{N,M}(z_1,z_2)$, can be approximated arbitrarily closely by an FIR mininum-phase filter generated by the truncated reflection coefficient sequence shown here.

$P_{N,M}$ and $H_{N,M}(z_1,z_2)$. Solving the lower-order normal equations associated with this autocorrelation sequence, we obtain a 2-D sequence of 2-D analytic minimum-phase filters, $H_{n,m}(z_1,z_2)$, and a 2-D reflection coefficient sequence related by the 2-D Levinson algorithm. The remaining details of the proof, and the uniqueness part of the proof are in Appendix A2.

To summarize the results of this section: we have shown that there ia a one-to-one relation between a class of 2-D analytic minimum-phase filters and a class of 2-D reflection coefficient sequences. If the filter is finite-order, then the reflection coefficient sequence is almost always infinite-order. Fortunately if the reflection coefficient sequence is finite-order then the filter is also finite-order. In cases where the reflection coefficient sequence of a particular filter is infinite-order, the filter can be uniformly approximated by a minimum-phase FIR filter corresponding to a finite-order reflection coefficient sequence.

The practical significance of the 2-D reflection coefficient representation is that it provides a new domain in which to design 2-D minimum-phase FIR filters. By designing 2-D FIR filters in the reflection coefficient domain, the difficult minimum-phase requirement is automatically satisfied merely by constraining the reflection coefficient magnitudes to be less than one.

4.3 The Behavior of the PEF $H_{N,M}(z_1,z_2)$ for Large Values of N

So far in this chapter, we have established, for finite values of N, the equivalence of three separate domains: the class of positive-definite analytic autocorrelation sequences, $\{r(k,\ell);\ (0,0)\leq(k,\ell)\leq(N,M)\}$, the class of analytic minimum-phase filters and positive prediction error variances, $\{P_{N,M};\ H_{N,M}(z_1,z_2)\}$, and the class of positive prediction error variances and reflection coefficient sequences, $\{P_{N,M};\ \rho(k,\ell),(0,0)<(k,\ell)\leq(N,M)\}$. The relations among the three domains are illustrated in Fig. 4.5. Now we want to investigate there results for large values of N.

Suppose that we have a positive analytic power density spectrum, $S(z_1,z_2)$. As we saw in section 3.5, the spectrum has the following factorization:

$$S(z_1,z_2) = \frac{P}{H(z_1,z_2)H(1/z_1,1/z_2)}\ , \tag{4.47}$$

where $H(z_1,z_2)$ is the analytic minimum-phase PEF for the present value of the random process given the infinite past, and P is the prediction error variance. We intuitively expect that the sequence of analytic minimum-phase PEFs, $H_{N,M}(z_1,z_2)$, converges to $H(z_1,z_2)$ as N goes to infinity.

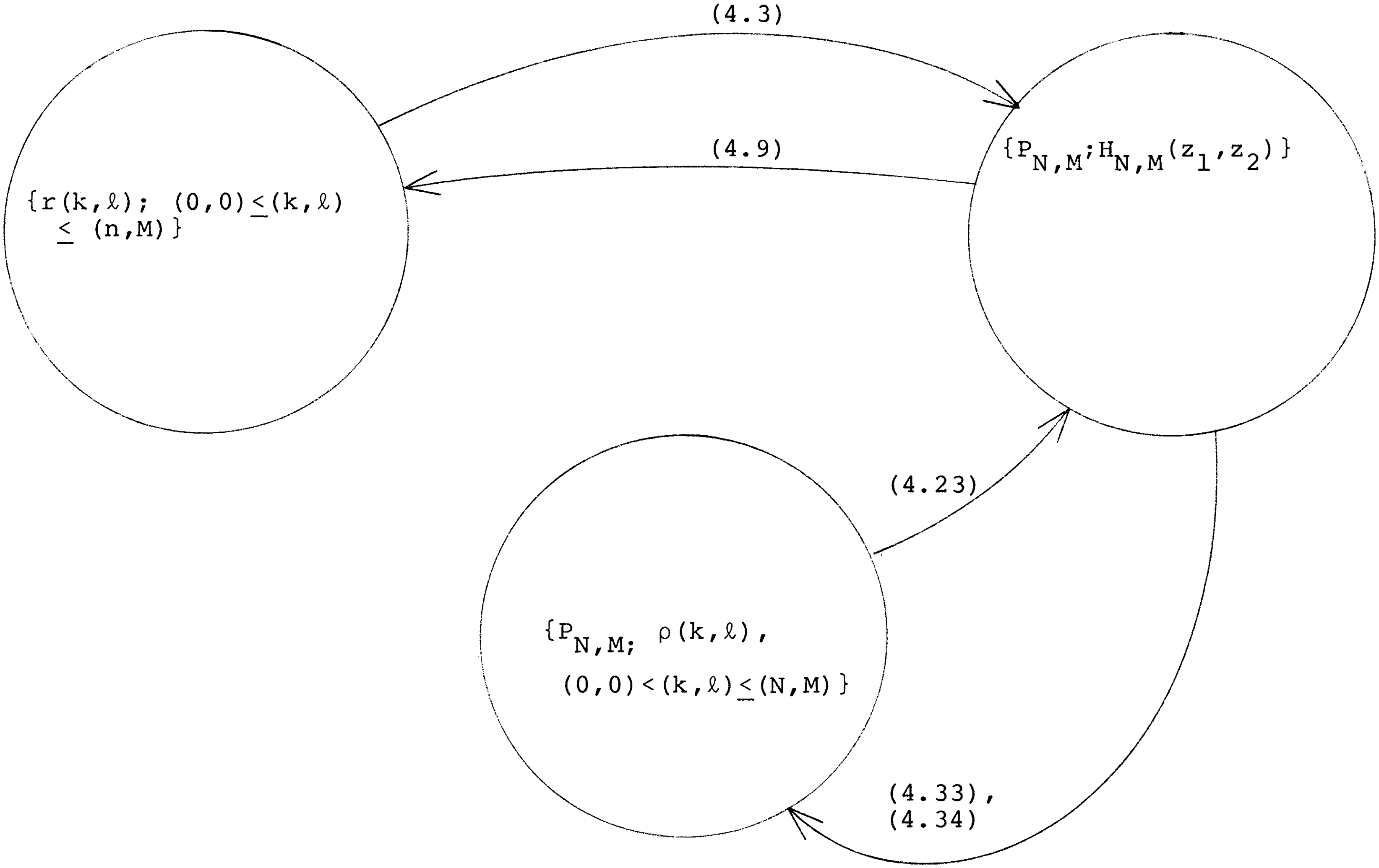


Fig. 4.5 The correspondence among 2-D positive-definite analytic autocorrelation sequences, 2-D analytic minimum-phase filters, and 2-D reflection coefficient sequences.

Theorem 4.4: If $S(z_1,z_2)$ is a positive analytic spectrum, then

1) The sequence of analytic minimum-phase PEFs, $H_{N,M}(z_1,z_2)$, converges uniformly in the neighborhood of the unit circles to the analytic minimum-phase limit filter, $H(z_1,z_2)$,

$$\lim_{N\to\infty} H_{N,M}(z_1,z_2) = H(z_1,z_2) \quad ; \tag{4.48}$$

2) The reflection coefficient sequence, $\rho(N,M)$, decays exponentially fast to zero as N goes to infinity

$$|\rho(N,M)| < (1+\varepsilon)^{-|N|}(1+\varepsilon)^{-|M|} \quad , \quad \varepsilon>0 \quad ; \tag{4.49}$$

3) The sequence of prediction error variances converges to the positive limit

$$\lim_{N\to\infty} P_{N,M} = P \quad . \tag{4.50}$$

Since we do not require this theorem for any of our subsequent work, we do not present a proof for it in this thesis. A complete proof can be found in [28].

What we do want to prove is that if a 2-D random process has a positive analytic spectrum, then its minimum-phase whitening filter can be uniformly approximated by an FIR minimum-phase filter corresponding to a finite number of reflection coefficients. While this can be argued by means of Theorem 4.4, there is a much more direct way to show this.

If the minimum-phase whitening filter, $H(z_1,z_2)$, is truncated to obtain the following filter, for a particular (N,M):

$$[1 - \sum_{(0,0)<(k,\ell)\leq(N,M)}\sum h(k,\ell) z_1^{-k} z_2^{-\ell}] \quad , \qquad (4.51)$$

this truncated filter converges uniformly to $H(z_1,z_2)$ in some neighborhood of the unit circles as N goes to plus infinity. Moreover, Theorem 3.1 implies that the truncated filter is minimum-phase for all sufficiently large values of N, so it can be uniformly approximated by a minimum-phase FIR filter corresponding to a finite number of reflection coefficients.

Consequently, any 2-D analytic minimum-phase whitening filter can be uniformly approximated by an FIR minimum-phase filter corresponding to a finite number of reflection coefficients. Therefore we have established that the 2-D reflection coefficient representation is a potentially useful tool in 2-D linear prediction problems.

APPENDIX A1

PROOF OF THEOREM 4.1

A1.1 Proof of Theorem 4.1(a) for $H_{N-1,+\infty}(z_1,z_2)$

Given the positive-definite analytic autocorrelation sequence, $\{r(k,\ell);\ (0,0)\leq(k,\ell)\leq(N-1,+\infty)\}$, we want to prove the existence of an analytic minimum-phase PEF, $H_{N-1,+\infty}(z_1,z_2)$. In this section we derive a constructive procedure for obtaining the PEF.

We begin by solving a related (but different) prediction problem:

$$[\hat{x}(k,\ell)\,|\,\{x(k-s,\ell-t);\ 1\leq s\leq(N-1),-\infty<t<\infty\}]$$

$$= \sum_{s=1}^{N-1}\sum_{t=-\infty}^{\infty} f(N-1;s,t)x(k-s,\ell-t) \quad . \tag{A1.1}$$

The filter coefficients, $f(N-1;s,t)$, are chosen to satisfy the following orthogonality conditions:

$$E\{[x(k,\ell) - \sum_{s=1}^{N-1}\sum_{t=-\infty}^{\infty} f(N-1;s,t)x(k-s,\ell-t)]x(k-u,\ell-v)\} = 0 \quad ,$$

$$1\leq u\leq N-1,-\infty<v<\infty \quad . \tag{A1.2}$$

We claim that we can then solve the original prediction problem as follows:

$$[\hat{x}(k,\ell)\,|\,\{x(k-s,\ell-t);\ (0,0)<(s,t)\leq(N-1,+\infty)\}]$$

$$= \sum\sum_{(0,0)<(s,t)\leq(N-1,+\infty)} h(N-1,+\infty;s,t)x(k-s,\ell-t)$$

$$= \sum_{s=1}^{N-1} \sum_{t=-\infty}^{\infty} f(N-1;s,t)x(k-s,\ell-t)$$

$$+ \sum_{\tau=1}^{\infty} g(N-1;\tau)[x(k,\ell-\tau)$$

$$- \sum_{s=1}^{N-1} \sum_{t=-\infty}^{\infty} f(N-1;s,t)x(k-s,\ell-\tau-t)] \quad . \qquad (A1.3)$$

Because of (A1.2) the orthogonality conditions,

$$E\{[x(k,\ell) - \sum\sum_{(0,0)<(s,t)\leq(N-1,+\infty)} h(N-1,+\infty;s,t)x(k-s,\ell-t)]x(k-u,\ell-v)$$

$$= 0 \quad , \qquad (0,0)<(u,v)\leq(N-1,+\infty) \quad , \qquad (A1.4)$$

are automatically satisfied for $\{1\leq u\leq N-1, -\infty<v<\infty\}$. The remaining orthogonality conditions for $\{u=0, v>0\}$ are satisfied by choosing the $g(N-1;\tau)$ appropriately. If we can find a stable solution for the $f(N-1;s,t)$ and the $g(N-1;\tau)$, then we will have found the unique, stable solution for $H_{N-1,+\infty}(z_1,z_2)$.

Our method of solving for $H_{N-1,+\infty}(z_1,z_2)$ is motivated by a fundamental result from linear estimation theory. Suppose that we are given two zero-mean random variables, y_1 and y_2, and that we wish to find the least-squares, linear estimate for another random variable, x:

$$[\hat{x}|y_1,y_2] = \alpha_1 y_1 + \alpha_2 y_2 \quad .$$

α_1 and α_2 can be found by choosing them to satisfy the orthogonality conditions:

$$E\{(x-\alpha_1 y_1-\alpha_2 y_2)y_1\} = 0 \quad ,$$

and $$E\{(x-\alpha_1 y_1-\alpha_2 y_2)y_2\} = 0 \quad .$$

A different method of solving the same problem begins by finding the least-squares, linear estimates for x and y_2 given y_1:

$$[\hat{x}|y_1] = \beta_1 y_1 \quad ,$$

and $$[\hat{y}_2|y_1] = \beta_2 y_1 \quad ,$$

where β_1 and β_2 are chosen to satisfy the following orthogonality conditions:

$$E\{(x-\beta_1 y_1)y_1\} = 0 \quad ,$$

and $$E\{(y_2-\beta_2 y_1)y_1\} = 0 \quad .$$

It can then be shown that

$$[\hat{x}|y_1,y_2] = [\hat{x}|y_1] + [\hat{x}|(y_2-[\hat{y}_2|y_1])]$$

$$= \beta_1 y_1 + \beta_3(y_2-\beta_2 y_1) \quad ,$$

where β_3 is chosen such that

$$E\{[x-\beta_1 y_1-\beta_3(y_2-\beta_2 y_1)](y_2-\beta_2 y_1)\} = 0 \quad ,$$

or equivalently

$$E\{[x-\beta_1 y_1-\beta_3(y_2-\beta_2 y_1)]y_2\} = 0 \quad .$$

Referring to the error expression, $(y_2-[\hat{y}_2|y_1])$, as the <u>innovation</u> of y_2 with respect to y_1, we have that the least-squares linear estimate for x given y_1 and y_2 is equal to the sum of the least-squares, linear estimate for x given y_1, and the least-squares, linear estimate for x given the innovation of y_2 with respect to y_1. Although we have only stated this result for random variables, it can easily be extended to random vectors and random processes.

Therefore, reconsidering (A1.3) in light of the above discussion, we formally have that the least-squares, linear estimate for $x(k,\ell)$ given $\{x(k-s,\ell-t);\ (0,0)<(s,t) \leq (N-1,+\infty)\}$ is equal to the sum of the least-squares, linear estimate for $x(k,\ell)$ given $\{x(k-s,\ell-t);\ 1\leq s\leq N-1, -\infty<t<\infty\}$, and the least-squares, linear estimate for $x(k,\ell)$ given the innovation of the 1-D sequence, $\{x(k,\ell-\tau);\ \tau\geq 1\}$, with respect to $\{x(k-s,\ell-t);\ 1\leq s\leq N-1, -\infty<t<\infty\}$.

In Z-transform notation, (A1.3) becomes

$$H_{N-1,+\infty}(z_1,z_2) = G_{N-1}(z_2)F_{N-1}(z_1,z_2) \tag{A1.5}$$

where $$G_{N-1}(z_2) = 1 - \sum_{\ell=1}^{\infty} g(N-1;\ell)z_2^{-\ell} \tag{A1.6}$$

and
$$F_{N-1}(z_1,z_2) = 1 - \sum_{k=1}^{N-1} F_{N-1;k}(z_2) z_1^{-k}$$

$$= 1 - \sum_{k=1}^{N-1} \sum_{\ell=-\infty}^{\infty} f(N-1;k,\ell) z_1^{-k} z_2^{-\ell} \quad .$$

(A1.7)

We claim that both $G_{N-1}(z_2)$ and $F_{N-1}(z_1,z_2)$ are analytic minimum-phase, and therefore that $H_{N-1,+\infty}(z_1,z_2)$ is analytic minimum-phase.

We first show how to solve (A1.2). We have

$$[r(u,v) - \sum_{s=1}^{N-1} \sum_{t=-\infty}^{\infty} f(N-1;s,t) r(u-s,v-t)]$$

$$= q(N-1;v)\delta_u \quad , \qquad 0 \le u \le (N-1), -\infty < v < \infty \quad , \qquad \text{(A1.8)}$$

where

$$q(N-1;v) = E\{[x(k,\ell+v) - \sum_{s=1}^{N-1} \sum_{t=-\infty}^{\infty} f(N-1;s,t) x(k-s,\ell+v-t)]$$

$$\cdot \; [x(k,\ell) - \sum_{s=1}^{N-1} \sum_{t=-\infty}^{\infty} f(N-1;s,t) x(k-s,\ell-t)]\}$$

$$= E\{[x(k,\ell+v) - \sum_{s=1}^{N-1} \sum_{t=-\infty}^{\infty} f(N-1;s,t)$$

$$\cdot \; x(k-s,\ell+v-t)] x(k,\ell)\} \quad . \qquad \text{(A1.9)}$$

Taking the transform of both sides of (A1.8) we have

$$[R_u(z_2) - \sum_{s=1}^{N-1} F_{N-1;s}(z_2) R_{u-s}(z_2)]$$

$$= Q_{n-1}(z_2)\delta_u \quad , \qquad 0 \le u \le N-1 \quad , \tag{A1.10}$$

where $$Q_{N-1}(z_2) = \sum_{\ell=-\infty}^{\infty} q(N-1;\ell) z_2^{-\ell} \quad . \tag{A1.11}$$

Writing (A1.10) in matrix form we have

$$\begin{bmatrix} R_0(z_2) & R_1(1/z_2) & \cdots & R_{N-1}(1/z_2) \\ R_1(z_2) & R_0(z_2) & \cdots & R_{N-2}(1/z_2) \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ R_{N-1}(z_2) & R_{N-2}(z_2) & \cdots & R_0(z_2) \end{bmatrix} \begin{bmatrix} 1 \\ -F_{N-1;1}(z_2) \\ \cdot \\ \cdot \\ \cdot \\ -F_{N-1;N-1}(z_2) \end{bmatrix} = \begin{bmatrix} Q_{N-1}(z_2) \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

(A1.12)

We recall that the above Toeplitz matrix is Hermitian positive-definite for all z_2 on the unit circle, and that the $R_k(z_2)$ are analytic functions. Consequently the determinant of the matrix is an analytic function of z_2. Since the determinant is non-zero for z_2 on the unit circle, it must be non-zero for z_2 in some neighborhood of the unit circle. Therefore the matrix is invertible for all z_2 in the neighborhood of the unit circle, so there is a unique solution for $F_{N-1}(z_1,z_2)$ and $Q_{N-1}(z_2)$, and $F_{N-1}(z_1,z_2)$ and $Q_{N-1}(z_2)$ are analytic. Since the matrix is Hermitian positive-definite for all z_2 on the unit circle it follows that $Q_{N-1}(z_2)$ is strictly positive on the unit circle.

To prove that $F_{N-1}(z_1,z_2)$ is minimum-phase we solve (A1.10) by a recursive procedure analogous to the 1-D Levinson algorithm. The recursion is as follows:

1) $$F_0(z_1,z_2) = 1 \quad ; \tag{A1.13}$$

$$Q_0(z_2) = R_0(z_2) \quad ; \tag{A1.14}$$

2) $$F_n(z_1,z_2) = F_{n-1}(z_1,z_2) - \rho_n(z_2)z_1^{-n}F_{n-1}(1/z_1,1/z_2), \tag{A1.15}$$

$$Q_n(z_2) = Q_{n-1}(z_2)[1 - \rho_n(z_2)\rho_n(1/z_2)] \quad , \tag{A1.16}$$

$$1 \leq n \leq N-1 \quad .$$

The "reflection coefficient function," $\rho_n(z_2)$, can be shown to be an analytic function. Since $Q_{n-1}(z_2)$ and $Q_n(z_2)$ are both positive for all z_2 on the unit circle, the magnitude of $\rho_n(z_2)$ must be less than one for all z_2 on the unit circle. Therefore considering (A1.15) we can use Theorem 3.1 to argue inductively that $F_{N-1}(z_1,z_2)$ is minimum-phase.

To find $G_{N-1}(z_2)$ we impose the orthogonality conditions (A1.4) for $\{u=0,v>0\}$. Considering (A1.3) we have

$$[q(N-1;v) - \sum_{\tau=1}^{\infty} g(N-1;\tau)]q(N-1;v-\tau)]$$

$$= P_{N-1,+\infty}\delta_v \quad , \quad v \geq 0 \quad . \qquad (A1.17)$$

But solving (A1.17) is equivalent to performing the following 1-D spectral factorization:

$$Q_{N-1}(z_2) = \frac{P_{N-1,+\infty}}{G_{N-1}(z_2)G_{N-1}(1/z_2)} \quad . \qquad (A1.18)$$

Since $Q_{N-1}(z_2)$ is a positive analytic spectrum, we know that the factorization can always be performed, and that $G_{N-1}(z_2)$ is analytic minimum-phase, and $P_{N-1,+\infty}$ is positive.

A1.2 Proof of Theorem 4.1(a) for $H_{N,M}(z_1,z_2)$

Given the positive-definite analytic autocorrelation sequence, $\{r(k,\ell);\ (0,0)\leq(k,\ell)\leq(N,M)\}$, we want to prove the existence of the analytic minimum-phase PEF, $H_{N,M}(z_1,z_2)$. A key part of the proof involves showing that for sufficiently small values of m ($m\to-\infty$) we can find a solution for $H_{N,m}(z_1,z_2)$ of the form

$$H_{N,m}(z_1,z_2) = A_{N,m}(z_2)H_{N-1,+\infty}(z_1,z_2)$$

$$- B_{N,m}(1/z_2)z_1^{-N}z_2^{-m}H_{N-1,+\infty}(1/z_1,1/z_2) \quad ,$$

$$(A1.19)$$

where $A_{N,m}(z_2)$ and $B_{N,m}(z_2)$ are analytic functions. Furthermore we can show that

$$\lim_{m\to-\infty} A_{N,m}(z_2) = 1 \quad , \tag{A1.20}$$

and
$$\lim_{m\to-\infty} B_{N,m}(z_2) = 0 \quad . \tag{A1.21}$$

(The convergence of both sequences of functions is uniform in some neighborhood of the unit circle.) Therefore $H_{N,m}(z_1,z_2)$ converges uniformly in some neighborhood of the unit circles to the limit filter, $H_{N,-\infty}(z_1,z_2) = H_{N-1,+\infty}(z_1,z_2)$. But since $H_{N-1,+\infty}(z_1,z_2)$ is minimum-phase, we can apply Theorem 3.1 to argue that for all sufficiently small values of m, $H_{N,m}(z_1,z_2)$ is minimum-phase. Finally applying the 2-D Levinson algorithm we can prove inductively that a solution for $H_{N,M}(z_1,z_2)$ exists and is analytic minimum-phase.

In order to obtain (A1.19), we first try a solution for $H_{N,m}(z_1,z_2)$ of the form

$$H_{N,m}(z_1,z_2) = [A'_{N,m}(z_2)F_{N-1}(z_1,z_2) - B'_{N,m}(1/z_2)z_1^{-N}z_2^{-m}F_{N-1}(1/z_1,1/z_2)] \ , \tag{A1.22}$$

where $$A'_{N,m}(z_2) = [1 - \sum_{\ell=1}^{\infty} a'(N,m;\ell) z_2^{-\ell}] \qquad \text{(A1.23)}$$

and $$B'_{N,m}(z_2) = \sum_{\ell=0}^{\infty} b'(N,m;\ell) z_2^{-\ell} \quad . \qquad \text{(A1.24)}$$

($F_{N-1}(z_1,z_2)$ is defined by (A1.7)) Equivalently, we have

$$[\hat{x}(k,\ell) \,|\, \{x(k-s,\ell-t);\ (0,0)<(s,t)\leq(N,m)\}]$$

$$= \sum\sum_{(0,0)<(s,t)\leq(N,m)} h(N,m;s,t)\, x(k-s,\ell-t)$$

$$= \sum_{s=1}^{N-1} \sum_{t=-\infty}^{\infty} f(N-1;s,t)\, x(k-s,\ell-t)$$

$$+ \sum_{\tau=1}^{\infty} a'(N,m;\tau) [x(k,\ell-\tau) - \sum_{s=1}^{N-1} \sum_{t=-\infty}^{\infty} f(N-1;s,t)$$

$$\cdot\ x(k-s,\ell-\tau-t)]$$

$$+ \sum_{\tau=0}^{\infty} b'(N,m;\tau) [x(k-N,\ell-m+\tau)$$

$$- \sum_{s=1}^{(N-1)} \sum_{t=-\infty}^{\infty} f(N-1;s,t)\, x(k-N+s,\ell-m+\tau+t)] \quad .$$

(A1.25)

Recalling (A1.2), we see that the normal equations,

$$E\{[x(k,\ell) - \sum\sum_{(0,0)<(s,t)\leq(N,m)} h(N,m;s,t)\, x(k-s,\ell-t)]\, x(k-u,\ell-v)\}$$

$$= 0 \quad , \quad (0,0)<(u,v)\leq(N,m) \quad , \qquad \text{(A1.26)}$$

are automatically satisfied for $\{1 \leq u \leq N-1, -\infty < v < \infty\}$ if the prediction problem is formulated in the form (A1.25). The sequences $a'(N,m;\tau)$ and $b'(N,m;\tau)$ need only to be chosen to satisfy (A1.26) for $\{u=0, v \geq 1\}$ and $\{u=N, v \leq m\}$. If stable solutions for $a'(N,m;\tau)$ and $b'(N,m;\tau)$ can be found, then we will have found the unique, stable solution for $H_{N,m}(z_1,z_2)$.

The innovations interpretation for (A1.25) is that the least-squares, linear estimate for $x(k,\ell)$ given $\{x(k-s,\ell-t);\ (0,0) < (s,t) \leq (N,m)\}$ is equal to the sum of the least-squares, linear estimate for $x(k,\ell)$ given $\{x(k-s,\ell-t);$ $1 \leq s \leq N-1, -\infty < t < \infty\}$, and the least-squares linear estimate for $x(k,\ell)$ given the innovations of the two sequences, $\{x(k,\ell-\tau);\ \tau \geq 1\}$ and $\{x(k-N,\ell-m+\tau);\ \tau \geq 0\}$, with respect to $\{x(k-s,\ell-t);\ 1 \leq s \leq N-1, -\infty < t < \infty\}$.

Recalling (A1.5),

$$H_{N-1,+\infty}(z_1,z_2) = G_{N-1}(z_2) F_{N-1}(z_1,z_2) , \qquad \text{(A1.27)}$$

we see that (A1.22) reduces to (A1.19) where

$$A_{N,m}(z_2) = A'_{N,m}(z_2) G_{N-1}^{-1}(z_2) = [1 - \sum_{\ell=1}^{\infty} a(N,m;\ell) z_2^{-\ell}] , \qquad \text{(A1.28)}$$

and

$$B_{N,m}(z_2) = B'_{N,m}(z_2) G_{N-1}^{-1}(z_2) = \sum_{\ell=0}^{\infty} b(N,m;\ell) z_2^{-\ell} . \qquad \text{(A1.29)}$$

Using (A1.19) with (A1.28) and (A1.29) we can write the prediction error resulting from $H_{N,m}(z_1,z_2)$ operating on $x(k,\ell)$ in the form:

$$\tilde{x}^{(+)}(N,m;k,\ell) = \tilde{x}^{(+)}(N-1,+\infty;k,\ell) - \sum_{\tau=1}^{\infty} a(N,m;\tau)\tilde{x}^{(+)}(N-1,+\infty;k,\ell-\tau) - \sum_{\tau=0}^{\infty} b(N,m;\tau)\tilde{x}^{(-)}(N-1,+\infty;k-N,\ell-m+\tau) \quad , \tag{A1.30}$$

where, for all $(0,0)\leq(n,m)\leq(N,M)$,

$$\tilde{x}^{(+)}(n,m;k,\ell) = [x(k,\ell) - \sum_{(0,0)<(s,t)\leq(n,m)}\sum h(n,m;s,t)x(k-s,\ell-t)] \quad , \tag{A1.31}$$

and

$$\tilde{x}^{(-)}(n,m;k,\ell) = [x(k,\ell) - \sum_{(0,0)<(s,t)\leq(n,m)}\sum h(n,m;s,t) \cdot x(k+s,\ell+t)] \quad . \tag{A1.32}$$

The advantage of considering the prediction problem in the form (A1.30) is that for fixed values of k, the 1-D stationary random processes, $\{\tilde{x}^{(+)}(N-1,+\infty;\ k,\ell),-\infty<\ell<\infty\}$, and $\{\tilde{x}^{(-)}(N-1,+\infty;\ k-N,\ell),-\infty<\ell<\infty\}$ are both individually white (though they are correlated with each other). Specifically, using (A1.4), it is easy to show that

$$E\{\tilde{x}^{(+)}(N-1,+\infty;k,\ell_1)\tilde{x}^{(+)}(N-1,+\infty;k,\ell_2)\} = P_{N-1,+\infty}\delta_{\ell_1-\ell_2} \quad , \tag{A1.33}$$

and

$$E\{\tilde{x}^{(-)}(N-1,+\infty;k-N,\ell_1)\tilde{x}^{(-)}(N-1,+\infty;k-N,\ell_2)\} = P_{N-1,+\infty}\delta_{\ell_1-\ell_2} \quad . \tag{A1.34}$$

We denote the normalized cross-correlation between the two processes by $\lambda(N;\tau)$:

$$\lambda(N;\tau) = \frac{1}{P_{N-1,+\infty}} E\{\tilde{x}^{(+)}(N-1,+\infty;k,\ell+\tau)\tilde{x}^{(-)}(N-1,+\infty;k-N,\ell)\}$$

$$= \frac{1}{P_{N-1,+\infty}} \frac{1}{(2\pi j)^2} \oint_{|z_1|=1} \oint_{|z_2|=1} z_1^{N-1} z_2^{\tau-1} H^2_{N-1,+\infty}(z_1,z_2)$$

$$\cdot \Big[\sum\sum_{(-N,-m)\le(s,t)\le(N,m)} r(s,t) z_1^{-s} z_2^{-t} \Big] dz_1 dz_2 \quad , \quad -\infty<\tau\le m \quad . \tag{A1.35}$$

Since $\lambda(N;\tau)$ is the inverse Z-transform of a function in z_2, analytic in the neighborhood of the unit circle, it follows that $\lambda(N;\tau)$ decays exponentially fast to zero as τ goes to minus infinity. Applying Schwartz's inequality we see that the magnitude of $\lambda(N;\tau)$ is less than one. Therefore we have the bound

$$|\lambda(N;\tau)| < (1+\varepsilon_0)^{-|\tau|} \quad , \quad \tau\le m, \varepsilon_0>0 \quad . \tag{A1.36}$$

To obtain the optimum values for $a(N,m;\tau)$ and $b(N,m;\tau)$ we apply the following orthogonality conditions:

$$E\{\tilde{x}^{(+)}(N,m;k,\ell)\tilde{x}^{(+)}(N-1,+\infty;k,\ell-v_1)\} = 0 \quad , \quad v_1 \geq 1 \quad , \tag{Al.37}$$

and

$$E\{\tilde{x}^{(+)}(N,m;k,\ell)\tilde{x}^{(-)}(N-1,+\infty;k-N,\ell-m+v_2) = 0 \quad , \quad v_2 \geq 0 \quad . \tag{Al.38}$$

Choosing $a(N,m;\tau)$ and $b(N,m;\tau)$ to satisfy (Al.37) and (Al.38) is equivalent to choosing them to minimize the variance of the right-hand side of (Al.30). But since there is a one-to-one relation between $A_{N,m}(z_2)$ and $A'_{N,m}(z_2)$, and between $B_{N,m}(z_2)$ and $B'_{N,m}(z_2)$, it follows that this is equivalent to optimizing $A_{N,m}(z_2)$ and $B_{N,m}(z_2)$.

Using (Al.30), (Al.33), (Al,34), and (Al.35) we have

$$0 = [-a(N,m;v_1) - \sum_{\tau=0}^{\infty} b(N,m;\tau)\lambda(N;m-v_1-\tau)] \quad , \quad v_1 \geq 1 \quad , \tag{Al.39}$$

and

$$0 = [\lambda(N;m-v_2) - \sum_{\tau=1}^{\infty} a(N,m;\tau)\lambda(N;m-v_2-\tau) - b(N,m;v_2)], \quad v_2 \geq 0 \quad . \tag{Al.40}$$

Substituting (A1.39) into (A1.40) we have

$$b(N,m;v) = [\lambda(N;m-v) + \sum_{\tau_1=0}^{\infty} \sum_{\tau_2=1}^{\infty} b(N,m;\tau_1)$$

$$\cdot \lambda(N;m-v-\tau_2)\lambda(N;m-\tau_1-\tau_2)] \quad , \quad v \geq 0 \quad . \qquad (A1.41)$$

We solve this equation iteratively; the procedure leads to a so-called Neumann series solution. (The same technique is used in the theory of Fredholm integral equations.) We obtain the following <u>formal</u> solution:

$$b(N,m;v) = \sum_{q=1}^{\infty} w_q(N,m;v) \quad , \quad v \geq 0 \quad , \qquad (A1.42)$$

where $w_q(N,m;v)$ satisfies the recursion:

$$w_1(N,m;v) = \lambda(N;m-v) \quad , \quad v \geq 0 \quad , \qquad (A1.43)$$

and

$$w_q(N,m;v) = \sum_{\tau_1=0}^{\infty} \sum_{\tau_2=1}^{\infty} w_{q-1}(N,m;\tau_1)\lambda(N;m-v-\tau_2)\lambda(N;m-\tau_1-\tau_2) \quad ,$$

$$v \geq 0, q \geq 2 \quad . \qquad (A1.44)$$

We need to show that the series (A1.42) converges absolutely and uniformly. Applying the bound (A1.36) to (A1.43) and (A1.44), and assuming that $m<0$ (which involves no loss in generality since we will need to let m approach minus infinity) we have

$$|w_1(N,m;v)| < (1+\varepsilon_0)^{m-v} \quad , \qquad v \geq 0 \quad , \tag{A1.45}$$

and

$$|w_q(N,m;v)| < \sum_{\tau_1=0}^{\infty} \sum_{\tau_2=1}^{\infty} |w_{q-1}(N,m;\tau_1)|(1+\varepsilon_0)^{(2m-v-\tau_1-2\tau_2)}$$

$$= \sum_{\tau_1=0}^{\infty} |w_{q-1}(n,m;\tau_1)|(1+\varepsilon_0)^{(2m-v-\tau_1)} \cdot \frac{(1+\varepsilon_0)^{-2}}{[1-(1+\varepsilon_0)^{-2}]} \quad ,$$

$$q \geq 2, v \geq 0 \quad . \tag{A1.46}$$

It can be verified by successive substitution that

$$|w_q(N,m;v)| < \left[\frac{(1+\varepsilon_0)^{m-1}}{1-(1+\varepsilon_0)^{-2}}\right]^{2(q-1)} \cdot (1+\varepsilon_0)^{(m-v)} \quad ,$$

$$q \geq 1, v \geq 0 \quad . \tag{A1.47}$$

Inspecting the bracketed quantity in (A1.47) we see that the bound can be made exponentially decaying in q by making m small enough. Specifically there is a negative integer, m_0, such that for all $m \leq m_0$,

$$\frac{(1+\varepsilon_0)^{m-1}}{1-(1+\varepsilon_0)^{-2}} < 1 \quad . \tag{A1.48}$$

Therefore, for all $m \leq m_0$, the Weierstrass M-test can be applied to show that the series solution (A1.42) converges absolutely and uniformly.

Applying the bound (A1.47) to (A1.42), we have

$$|b(N,m;v)| \leq \sum_{q=1}^{\infty} |w_q(N,m;v)|$$

$$< \frac{(1+\varepsilon_0)^{m-v}}{1 - \left[\frac{(1+\varepsilon_0)^{m-1}}{1-(1+\varepsilon_0)^{-2}}\right]^2} \,, \qquad m\leq m_0, v\geq 0 \quad . \tag{A1.49}$$

Applying the above bound to (A1.39) we have

$$|a(N,m;v)| \leq \sum_{\tau=0}^{\infty} |b(N,m;\tau)|\,|\lambda(N;m-v-\tau)|$$

$$< \frac{(1+\varepsilon_0)^{(2m-v)}}{\left(1 - \left[\frac{(1+\varepsilon_0)^{m-1}}{1-(1+\varepsilon_0)^{-2}}\right]^2\right)\left[1-(1+\varepsilon_0)^{-2}\right]} \,, \qquad m\leq m_0, v\geq 1 \quad . \tag{A1.50}$$

Consequently, for all $m\leq m_0$, $A_{N,m}(z_2)$ and $B_{N,m}(z_2)$ (defined by (A1.28) and (A1.29)) are analytic in the neighborhood at the unit circle (in fact, in the region $|z_2|>(1+\varepsilon_0)^{-1}$). Moreover, (A1.49) and (A1.50) imply that as m goes to minus infinity, $A_{N,m}(z_2)$ converges uniformly to one and $B_{N,m}(z_2)$ converges uniformly to zero in the neighborhood of the unit circle.

Therefore in this section we have proved that there is some number m_0 such that for all $m\leq m_0$ we can find a solution for the PEF, $H_{N,m}(z_1,z_2)$, analytic in the neighborhood of the unit circles. $H_{N,m}(z_1,z_2)$ converges uniformly to the analytic minimum-phase filter, $H_{N,-\infty}(z_1,z_2) = H_{N-1,+\infty}(z_1,z_2)$, as m goes to minus infinity. The remainder of the proof of Theorem 4.1(a) has already been discussed.

A1.3 Proof of Theorem 4.1(b) for $H_{N-1,+\infty}(z_1,z_2)$

Given an analytic minimum-phase filter

$$H_{N-1,+\infty}(z_1,z_2) = [1 - \sum\sum_{(0,0)<(k,\ell)\leq(N-1,+\infty)} h(N-1,+\infty;k,\ell)\, z_1^{-k} z_2^{-\ell}] \quad , \tag{A1.51}$$

and a positive number $P_{N-1,+\infty}$, we want to prove that there is a unique positive-definite analytic autocorrelation sequence, $\{r(k,\ell);\ (0,0)\leq(k,\ell)\leq(N-1,+\infty)\}$, such that the normal equations (A1.4) are satisfied. The existence part of the proof was discussed in section 4.1.

To prove the uniqueness part of the theorem we assume the existence of two positive-definite and analytic autocorrelation sequences, $\{r(k,\ell);\ (0,0)\leq(k,\ell)\leq(N-1,+\infty)\}$ and $\{r'(k,\ell);\ (0,0)\leq(k,\ell)\leq(N-1,+\infty)\}$, both of which generate $H_{N-1,+\infty}(z_1,z_2)$ and $P_{N-1,+\infty}$ via the normal equations. In each case the normal equations can be solved by means of the constructive procedure derived in section A1.1. Therefore working with $\{r(k,\ell)\}$ we first solve (A1.10) to obtain $F_{N-1}(z_1,z_2)$ and $Q_{N-1}(z_2)$. We then perform the 1-D spectral factorization (A1.18) to obtain $P_{N-1,+\infty}$ and $G_{N-1}(z_2)$. The PEF is then given by the formula

$$H_{N-1,+\infty}(z_1,z_2) = G_{N-1}(z_2)F_{N-1}(z_1,z_2) \quad . \tag{A1.52}$$

Likewise, working with $\{r'(k,\ell)\}$ we solve (A1.10) to obtain $F'_{N-1}(z_1,z_2)$ and $Q'_{N-1}(z_2)$, and we solve (A1.18) to obtain $P_{N-1,+\infty}$ and $G'_{N-1}(z_2)$. We obtain the PEF by the formula

$$H_{N-1,+\infty}(z_1,z_2) = G'_{N-1}(z_2)F'_{N-1}(z_1,z_2) \quad . \tag{A1.53}$$

Comparing (A1.52) and (A1.53) we have

$$G_{N-1}(z_2)F_{N-1}(z_1,z_2) = G'_{N-1}(z_2)F'_{N-1}(z_1,z_2) \quad , \tag{A1.54}$$

or

$$G_{N-1}(z_2)[1 - \sum_{k=1}^{N-1} F_{N-1;k}(z_2)z_1^{-k}]$$

$$= G'_{N-1}(z_2)[1 - \sum_{k=1}^{N-1} F'_{N-1;k}(z_2)z_1^{-k}] \quad . \tag{A1.55}$$

Comparing similar coefficients of z_1 in (A1.55) we see that

$$G_{N-1}(z_2) = G'_{N-1}(z_2) \quad , \tag{A1.56}$$

and $$F_{N-1}(z_1,z_2) = F'_{N-1}(z_1,z_2) \quad . \tag{A1.57}$$

Considering (A1.18) we have

$$Q_{N-1}(z_2) = \frac{P_{N-1,+\infty}}{G_{N-1}(z_2)\,G_{N-1}(1/z_2)} \quad , \tag{A1.58}$$

and
$$Q'_{N-1}(z_2) = \frac{P_{N-1,+\infty}}{G'_{N-1}(z_2)\,G'_{N-1}(1/z_2)} \quad . \tag{A1.59}$$

Using (A1.56) we see that

$$Q_{N-1}(z_2) = Q'_{N-1}(z_2) \quad . \tag{A1.60}$$

Therefore we see that $\{r(k,\ell)\}$ and $\{r'(k,\ell)\}$ both generate the same $F_{N-1}(z_1,z_2)$ and $Q_{N-1}(z_2)$ via (A1.10). Writing (A1.10) for both cases we have

$$[R_u(z_2) - \sum_{s=1}^{N-1} F_{N-1;s}(z_2)R_{u-s}(z_2)] = Q_{N-1}(z_2)\delta_u \quad , \qquad 0 \le u \le (N-1) \quad , \tag{A1.61}$$

and

$$[R'_u(z_2) - \sum_{s=1}^{N-1} F_{N-1;s}(z_2)R'_{u-s}(z_2)] = Q_{N-1}(z_2)\delta_u \quad , \qquad 0 \le u \le (N-1) \quad . \tag{A1.62}$$

But as we observed in section A1.1, for all values of z_2 on the unit circle, the (A1.10) are merely the normal equations for a 1-D predictor. Therefore since $F_{N-1}(z_1,z_2)$ is a 1-D minimum-phase filter in z_1 for all fixed z_2 on the unit circle, and since $Q_{N-1}(z_2)$ is strictly positive for all z_2 on the unit circle, the complex version of

Theorem 2.1(b) implies that

$$R_k(z_2) = R'_k(z_2) \quad , \quad 0 \le k \le (N-1) \quad , \quad |z_2|=1 \quad . \tag{A1.63}$$

Taking the inverse Z-transform of both sides of (A1.63) (with the integration contour on the unit circle) we have that

$$r(k,\ell) = r'(k,\ell) \quad , \quad (0,0) \le (k,\ell) \le (N-1,+\infty) \quad . \tag{A1.64}$$

A1.4 Proof of Theorem 4.1(b) for $H_{N,M}(z_1,z_2)$

Given an analytic minimum-phase filter,

$$H_{N,M}(z_1,z_2) = [1 - \sum\sum_{(0,0)<(k,\ell)\le(N,M)} h(N,M;k,\ell) z_1^{-k} z_2^{-\ell}] \quad , \tag{A1.65}$$

and a positive number $P_{N,M}$, we want to prove that there is a unique positive-definite analytic autocorrelation sequence, $\{r(k,\ell);\ (0,0)\le(k,\ell)\le(N,M)\}$, such that the normal equations, (4.3), are satisfied. We concentrate here only on the uniqueness part of the proof, since the existence part was discussed in section 4.1.

We assume the existence of two positive-definite analytic autocorrelation sequences, $\{r(k,\ell);\ (0,0)\le(k,\ell)\le(N,M)\}$ and $\{r'(k,\ell);\ (0,0)\le(k,\ell)\le(N,M)$ both of which generate $H_{N,M}(z_1,z_2)$ and $P_{N,M}$ via the normal equations. In both cases the normal equations can be solved by the method derived in sections A1.1 and A1.2. Therefore working with $\{r(k,\ell)\}$ we generate a sequence of analytic minimum-phase PEFs $\{H_{N,m}(z_1,z_2);\ -\infty<m\leq M\}$ and positive prediction error variances, $\{P_{N,m};\ -\infty<m\leq M\}$, related by the 2-D Levinson recursion:

$$H_{N,m}(z_1,z_2) = [H_{N,m-1}(z_1,z_2) - \rho(N,m)z_1^{-N}z_2^{-m}H_{N,m-1}(1/z_1,1/z_2] \ , \quad m\leq M \ , \qquad \text{(A1.66)}$$

and $$P_{N,m} = P_{N,m-1}[1 - \rho^2(N,m)] \ , \quad m\leq M \ . \qquad \text{(A1.67)}$$

Likewise working with $\{r'(k,\ell)\}$ we generate a sequence at analytic minimum-phase PEFs $\{H'_{N,m}(z_1,z_2);\ m\leq M\}$ and positive prediction error variances $\{P'_{N,m};\ m\leq M\}$, related by the 2-D Levinson algorithm:

$$H'_{N,m}(z_1,z_2) = [H'_{N,m-1}(z_1,z_2) - \rho'(N,m)z_1^{-N}z_2^{-m}H'_{N,m-1}(1/z_1,1/z_2)], \quad m\leq M \ , \qquad \text{(A1.68)}$$

and $$P'_{N,m} = P'_{N,m-1}[1 - \rho'^2(N,m)] \ , \quad m\leq M \ , \qquad \text{(A1.69)}$$

where

$$H'_{N,M}(z_1,z_2) = H_{N,M}(z_1,z_2) \ , \qquad \text{(A1.70)}$$

and

$$P'_{N,M} = P_{N,M} \ . \qquad \text{(A1.71)}$$

Using (A1.70) we have that

$$\rho'(N,M) = h'(N,M;N,M)$$
$$= h(N,M;N,M) = \rho(N,M) \quad . \tag{A1.72}$$

Therefore using (A1.69), (A1.71), and (A1.72) we see that

$$P'_{N,M-1} = P_{N,M-1} \quad . \tag{A1.73}$$

Since the reflection coefficient magnitudes are less than one, the recursions (A1.66) and (A1.68) can be "run backwards" as follows:

$$H_{N,m-1}(z_1,z_2) = \frac{1}{[1-\rho^2(N,m)]} [H_{N,m}(z_1,z_2) + \rho(N,m) z_1^{-N} z_2^{-m} H_{N,m}(1/z_1,1/z_2)] \quad , \quad m \leq M \quad , \tag{A1.74}$$

and

$$H'_{N,m-1}(z_1,z_2) = \frac{1}{[1-\rho^2(N,m)]} [H'_{N,m}(z_1,z_2) + \rho'(N,m) z_1^{-N} z_2^{-m} H'_{N,m}(1/z_1,1/z_2)] \quad , \quad m \leq M \quad . \tag{A1.75}$$

Using (A1.70), (A1.72), (A1.74), and (A1.75) we see that

$$H'_{N,M-1}(z_1,z_2) = H_{N,M-1}(z_1,z_2) \quad . \tag{A1.76}$$

Therefore we can argue inductively that

$$H'_{N,m}(z_1,z_2) = H_{N,m}(z_1,z_2) \quad , \quad m \le M \quad , \tag{A1.77}$$

and $$P'_{N,m} = P_{N,m} \quad , \quad m \le M \quad . \tag{A1.78}$$

Taking the limit as m goes to minus infinity, we have that

$$H'_{N-1,+\infty}(z_1,z_2) = H_{N-1,+\infty}(z_1,z_2) \quad , \tag{A1.79}$$

and $$P'_{N-1,+\infty} = P_{N-1,+\infty} \quad . \tag{A1.80}$$

But as we proved in the previous section, this implies that

$$r'(k,\ell) = r(k,\ell) \quad , \quad (0,0) \le (k,\ell) \le (N-1,+\infty) \quad . \tag{A1.81}$$

To obtain the remainder of the proof we write the normal equations for $H_{N,M}(z_1,z_2)$:

$$[r(s,t) - \sum\sum_{(0,0)<(k,\ell)\le(N,M)} h(N,M;k,\ell)r(s-k,t-\ell)] = P_{N,M}\delta_s\delta_t \quad ,$$

$$(0,0) \le (s,t) \le (N,M) \quad , \tag{A1.82}$$

and

$$[r'(s,t) - \sum\sum_{(0,0)<(k,\ell)\le(N,M)} h(N,M;k,\ell)r'(s-k,t-\ell)] = P_{N,M}\delta_s\delta_t \quad ,$$

$$(0,0) \le (s,t) \le (N,M) \quad . \tag{A1.83}$$

Subtracting the two equations, and using (A1.81) we have for $s = N$:

$$[\Delta r(N,t) - \sum_{\ell=1}^{\infty} h(N,M;0,\ell)\Delta r(N,t-\ell)] = 0 \quad , \quad t \leq M \quad , \tag{A1.84}$$

where $\Delta r(N,t) = r(N,t) - r'(N,t)$. (A1.85)

We claim that the filter,

$$[1 - \sum_{\ell=1}^{\infty} h(N,M;0,\ell) z_2^{-1}] = H_{N,M}(z_1,z_2)\Big|_{z_1=\infty} \quad , \tag{A1.86}$$

is minimum-phase. This is easily seen since we have

$$H_{N,M}(z_1,z_2)\Big|_{z_1=\infty} \cdot H_{N,M}^{-1}(z_1,z_2)\Big|_{z_1=\infty} = 1 \tag{A1.87}$$

Considering (A1.84), if we interpret the sequence $\Delta r(N,t)$ as the input to the filter, $[1 - \sum_{\ell=1}^{\infty} h(N,M;0,\ell) z_2^{-\ell}]$, then the output of the filter is zero for all $t \leq M$. But since the filter is minimum-phase, this can only mean that

$$\Delta r(N,t) = 0 \quad , \quad t \leq M \quad , \tag{A1.88}$$

which completes the proof.

APPENDIX A2

PROOF OF THEOREM 4.3

A2.1 Proof of Existence Part of Theorem 4.3(a)

We are given a 2-D reflection coefficient sequence, $\{\rho(k,\ell);\ (0,0)<(k,\ell)\leq(N,M)\}$, where

$$|\rho(k,\ell)| < (1+\varepsilon_0)^{-|\ell|} \tag{A2.1}$$

and ε_0 is some positive constant, and we want to prove the existence of a 2-D sequence of 2-D analytic minimum-phase filters, $\{H_{n,m}(z_1,z_2);\ (0,0)\leq(n,m)\leq(N,M)\}$, satisfying equations (4.21)-(4.26). As we indicated in section 4.2, this is trivial to prove if the reflection coefficient sequence is finite-order. To prove the existence part of the theorem for the general case where the reflection coefficient sequence is infinite-order, we work with a finite-order truncated reflection coefficient sequence, $\{\rho^{(L)}(k,\ell);\ (0,0)<(k,\ell)\leq(N,M)\}$, where

$$\rho^{(L)}(k,\ell) = \begin{cases} \rho(k,\ell) & , \quad |\ell|\leq L \\ 0 & , \quad |\ell|>L \ . \end{cases}$$

Associated with this truncated reflection coefficient sequence is a 2-D sequence of 2-D FIR minimum-phase filters, $\{H^{(L)}_{n,m}(z_1,z_2);\ (0,0)\leq(n,m)\leq(N,M)\}$. We then prove that as L goes to infinity, the sequence of filters

$\{H_{n,m}^{(L)}(z_1,z_2)\}$ converges uniformly, in the Cauchy sense, to a limit sequence of analytic minimum-phase filters, $\{H_{n,m}(z_1,z_2)\}$.

We first write the equations needed to obtain the filter sequence, $\{H_{n,m}^{(L)}(z_1,z_2)\}$. The filter sequence is computed recursively, the ordering of the recursion being a finite raster scan. The recursion begins as follows (we assume that $L>|M|$):

$$H_{0,0}^{(L)}(z_1,z_2) = 1 \quad , \tag{A2.2}$$

$$H_{0,m}^{(L)}(z_1,z_2) = H_{0,m-1}^{(L)}(z_1,z_2) - \rho(0,m)z_2^{-m}H_{0,m-1}^{(L)}(1/z_1,1/z_2) \quad , \quad 1\leq m\leq L \quad . \tag{A2.3}$$

The next column of the recursion begins as follows:

$$H_{1,-(L+1)}^{(L)}(z_1,z_2) = H_{0,L}^{(L)}(z_1,z_2) \quad . \tag{A2.4}$$

For the remainder of the column we have:

$$H_{1,m}^{(L)}(z_1,z_2) = H_{1,m-1}^{(L)}(z_1,z_2) - \rho(1,m)z_1^{-1}z_2^{-m}H_{1,m-1}^{(L)}(1/z_1,1/z_2) \quad , \quad -L\leq m\leq L \quad . \tag{A2.5}$$

In general, within each column of the recursion we have

$$H_{n,m}^{(L)}(z_1,z_2) = H_{n,m-1}^{(L)}(z_1,z_2) - \rho(N,m)z_1^{-n}z_2^{-m}H_{n,m-1}^{(L)}(1/z_1,1/z_2) \quad , \quad \{1\leq n\leq N-1,-L\leq m\leq L\},\{n=N,-L\leq m\leq M\} \quad . \tag{A2.6}$$

The equation for the transition between adjacent columns of the recursion is

$$H^{(L)}_{n,-(L+1)}(z_1,z_2) = H^{(L)}_{n-1,L}(z_1,z_2) \quad , \quad 1\leq n\leq N \quad . \tag{A2.7}$$

Finally, for $|m|>L$ we have

$$H^{(L)}_{k,m}(z_1,z_2) = H^{(L)}_{n,L}(z_1,z_2) \quad , \quad (n,L)<(k,m)<(n+1,-L) \ , \ 0\leq n\leq N-1 \ . \tag{A2.8}$$

Theorem 3.1 can be applied to (A2.3) and (A2.6) to prove inductively that all of the filters are minimum-phase.

We now prove that as L goes to infinity, the filter sequence $\{H^{(L)}_{n,m}(z_1,z_2)\}$ is uniformly bounded in some neighborhood of the unit circles, and for all $\{(0,0\leq(n,m)\leq(N,M)\}$. We confine z_1 and z_2 to a particular neighborhood of the unit circles, $\{(1+\varepsilon_1)^{-1}<|z_1|,|z_2|<(1+\varepsilon_1)\}$, and we denote the least upper bound for the magnitude of the filter $H^{(L)}_{n,m}(z_1,z_2)$ by $\|H^{(L)}_{n,m}\|$ where

$$\|H^{(L)}_{n,m}\| \geq |H^{(L)}_{n,m}(z_1,z_2)|, (1+\varepsilon_1)^{-1}<|z_1|,|z_2|<(1+\varepsilon_1) \quad . \tag{A2.9}$$

Considering (A2.3) we have

$$\|H_{0,m}^{(L)}\| \leq \|H_{0,m-1}^{(L)}\| \, [1 + |\rho(0,m)|(1+\varepsilon_1)^{|m|}]$$

$$< \|H_{0,m-1}^{(L)}\| \, [1 + \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|m|}] \quad , \quad 1 \leq m \leq L. \qquad \text{(A2.10)}$$

Using successive substitution, it can be verified that

$$\|H_{0,m}^{(L)}\| < \prod_{\ell=1}^{m} [1 + \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}]$$

$$< \prod_{\ell=1}^{L} [1 + \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}] \quad , \quad 1 \leq m \leq L \quad . \qquad \text{(A2.11)}$$

Considering (A2.8) for n=0, we have that

$$\|H_{n,m}^{(L)}\| < \prod_{\ell=1}^{L} [1 + \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}] \quad , \quad (0,0) \leq (n,m) \leq (1,-L-1) \quad .$$

(A2.12)

Similarly, using (A2.6) for n=1, it can be shown that

$$\|H_{1,m}^{(L)}\| < \|H_{1,-(L+1)}^{(L)}\| \prod_{\ell=-L}^{m} [1 + (1+\varepsilon_1)\left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}]$$

$$< \|H_{1,-(L+1)}^{(L)}\| \prod_{\ell=-L}^{L} [1 + (1+\varepsilon_1)\left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}] \quad ,$$

$$-L \leq m \leq L \quad . \qquad \text{(A2.13)}$$

Combining (A2.12) and (A2.13) we have that

$$\|H_{n,m}^{(L)}\| < \{\prod_{\ell=1}^{L} [1 + \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}]\}$$

$$\cdot \{\prod_{\ell=-L}^{L} [1 + (1+\varepsilon_1)\left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}]\} ,$$

$$(0,0) \leq (n,m) \leq (1,L) . \qquad \text{(A2.14)}$$

Using the same type of arguments, it can be shown inductively that

$$\|H_{n,m}^{(L)}\| < \{\prod_{\ell=1}^{L} [1 + \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}]\}$$

$$\cdot \{\prod_{k=1}^{N-1} \prod_{\ell=-L}^{L} [1 + (1+\varepsilon_1)^k \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}]\}$$

$$\cdot \{\prod_{\ell=-L}^{M} [1+(1+\varepsilon_1)^N \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}]\} ,$$

$$(0,0) \leq (n,m) \leq (N,M) . \qquad \text{(A2.15)}$$

If we now choose ε so that

$$0<\varepsilon_1<\varepsilon_0 , \qquad \text{(A2.16)}$$

we claim that the right-hand side of (A2.15) is bounded as L goes to infinity. We need only to show that the infinite products converge. We demonstrate this for the first term. We have

$$\log \prod_{\ell=1}^{\infty} [1 + \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}] = \sum_{\ell=1}^{\infty} \log[1 + \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}]$$

$$< \sum_{\ell=1}^{\infty} \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|} < \infty \quad , \qquad \text{(A2.17)}$$

where we have used the fact that for nonnegative x, $\log(1+x)\leq x$. Therefore we have the following uniform bound:

$$||H_{n,m}^{(L)}|| < K_0 < \infty \quad , \quad (0,0)\leq(n,m)\leq(N,M) \quad , \qquad \text{(A2.18)}$$

where K_0 is a constant independent of L, n, and m.

We now want to prove that the filter sequence, $\{H_{n,m}^{(L)}(z_1,z_2)\}$, converges uniformly in the Cauchy sense to a limit filter sequence. We consider the filter sequence for two values of L, L_1 and L_2, where $L_1>L_2$. We denote the least upper bound for $|H_{n,m}^{(L_1)}(z_1,z_2)-H_{n,m}^{(L_2)}(z_1,z_2)|$ in the neighborhood $\{(1+\varepsilon_1)^{-1}<|z_1|,|z_2|<(1+\varepsilon_1)\}$ by $||H_{n,m}^{(L_1)}-H_{n,m}^{(L_2)}||$, where

$$|H_{n,m}^{(L_1)}(z_1,z_2)-H_{n,m}^{(L_2)}(z_1,z_2)| \leq ||H_{n,m}^{(L_1)}-H_{n,m}^{(L_2)}|| ,$$

$$(1+\varepsilon_1)^{-1}<|z_1|,|z_2|<(1+\varepsilon_1) \quad . \qquad \text{(A2.19)}$$

We want to prove that the least upper bound (A2.19) converges uniformly to zero, for all $(0,0)\leq(n,m)\leq(N,M)$, as L_1 and L_2 go to infinity (with $L_1>L_2$). To prove this, we need to consider the propagation of $\| H_{n,m}^{(L_1)} - H_{n,m}^{(L_2)} \|$ in two regions: for $|m| \leq L_2$, and for $L_2+1 \leq |m| \leq L_1$.

First considering (A2.2) and (A2.3) we have that

$$\| H_{0,m}^{(L_1)} - H_{0,m}^{(L_2)} \| = 0 \quad , \quad 0 \leq m \leq L_2 \quad . \qquad \text{(A2.20)}$$

Next, using (A2.6) we have that

$$[H_{n,m}^{(L_1)}(z_1,z_2) - H_{n,m}^{(L_2)}(z_1,z_2)] = [H_{n,m-1}^{(L_1)}(z_1,z_2) - H_{n,m-1}^{(L_2)}(z_1,z_2)]$$

$$-\rho(n,m) z_1^{-n} z_2^{-m} [H_{n,m-1}^{(L_1)}(1/z_1,1/z_2) - H_{n,m-1}^{(L_2)}(1/z_1,1/z_2)] \; ,$$

$$\{1 \leq n \leq N-1, -L_2 \leq m \leq L_2\}, \; \{n=N, -L_2 \leq m \leq M\} \quad . \qquad \text{(A2.21)}$$

Or

$$\| H_{n,m}^{(L_1)} - H_{n,m}^{(L_2)} \| < \| H_{n,m-1}^{(L_1)} - H_{n,m-1}^{(L_2)} \| \, [1 + (1+\varepsilon_1)^n \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|m|}] \quad ,$$

$$\{1 \leq n \leq N-1, -L_2 \leq m \leq L_2\}, \; \{n=N, -L_2 \leq m \leq M\} \quad . \qquad \text{(A2.22)}$$

By successive substitution, we have that

$$\| H_{n,m}^{(L_1)} - H_{n,m}^{(L_2)} \| < \| H_{n,-(L_2+1)}^{(L_1)} - H_{n,-(L_2+1)}^{(L_2)} \|$$

$$\cdot \prod_{\ell=-L_2}^{m} [1 + (1+\varepsilon_1)^n \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}] \quad ,$$

$$\{1 \leq n \leq N-1, -L_2 \leq m \leq L_2\}, \; \{n=N, -L_2 \leq m \leq M\} \quad . \qquad \text{(A2.23)}$$

Finally using the worst-case values for n, m, and L_2 in (A2.23), we have that

$$||H_{n,m}^{(L_1)} - H_{n,m}^{(L_2)}|| < K_1 \, ||H_{n,-(L_2+1)}^{(L_1)} - H_{n,-(L_2+1)}^{(L_2)}|| \quad ,$$

$$\{1 \le n \le N-1, -L_2 \le m \le L_2\}, \ \{n=N, -L_2 \le m \le M\} \quad , \qquad \text{(A2.24)}$$

where K_1 is a constant independent of n, m, L_1, and L_2:

$$K_1 = \prod_{\ell=-\infty}^{\infty} [1+(1+\varepsilon_1)^N \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}] < \infty \quad . \qquad \text{(A2.25)}$$

We now consider the propagation of $||H_{n,m}^{(L_1)} - H_{n,m}^{(L_2)}||$ in the region $(L_2+1) \le |m| \le L_1$. Considering (A2.3) and (A2.6), we have that

$$H_{n,(L_2+1)}^{(L_1)}(z_1,z_2) = H_{n,L_2}^{(L_1)}(z_1,z_2) - \rho(n,L_2+1) z_1^{-n} z_2^{-(L_2+1)}$$

$$\cdot \, H_{n,L_2}^{(L_1)}(1/z_1,1/z_2) \quad , \quad 0 \le n \le (N-1) \quad ,$$

(A2.26)

and

$$H_{n,(L_2+2)}^{(L_1)}(z_1,z_2) = H_{n,(L_2+1)}^{(L_1)}(z_1,z_2) - \rho(n,L_2+2) z_1^{-n} z_2^{-(L_2+2)}$$

$$\cdot \, H_{n,(L_2+1)}^{(L_1)}(1/z_1,1/z_2) \quad , \quad 0 \le n \le (N-1) \quad .$$

(A2.27)

Adding the two equations, we have that

$$H^{(L_1)}_{n,(L_2+2)}(z_1,z_2) = H^{(L_1)}_{n,L_2}(z_1,z_2) - \sum_{\ell=(L_2+1)}^{(L_2+2)} \rho(n,\ell) \cdot z_1^{-n} z_2^{-\ell} H^{(L_1)}_{n,(\ell-1)}(1/z_1,1/z_2) \quad , \quad 0 \le n \le (N-1). \tag{A2.28}$$

It can be shown inductively that

$$H^{(L_1)}_{n,m}(z_1,z_2) = H^{(L_1)}_{n,L_2}(z_1,z_2) - \sum_{\ell=(L_2+1)}^{m} \rho(n,\ell) z_1^{-n} z_2^{-\ell} \cdot H^{(L_1)}_{n,(\ell-1)}(1/z_1,1/z_2) \quad ,$$

$$0 \le n \le (N-1) \quad , \quad (L_2+1) \le m \le L_1 \quad . \tag{A2.29}$$

Similarly, it can be shown that

$$H^{(L_1)}_{n+1,m}(z_1,z_2) = H^{(L_1)}_{n+1,-(L_1+1)}(z_1,z_2) - \sum_{\ell=-L_1}^{m} \rho(n+1,\ell) z_1^{-(n+1)} z_2^{-\ell} H^{(L_1)}_{n+1,\ell-1}(1/z_1,1/z_2) \quad ,$$

$$0 \le n \le (N-1), \quad (L_1 \le m \le -(L_2+1) \quad . \tag{A2.30}$$

Using (A2.8) we have that

$$H_{n,m}^{(L_2)}(z_1,z_2) = H_{n,L_2}^{(L_2)}(z_1,z_2) \quad ,$$

$$0 \le n \le (N-1), \; (L_2+1) \le m \le L_1 \quad , \qquad \text{(A2.31)}$$

and

$$H_{n+1,m}^{(L_2)}(z_1,z_2) = H_{n+1,-(L_1+1)}^{(L_2)}(z_1,z_2) \quad ,$$

$$0 \le n \le (N-1), \; -L_1 \le m \le -(L_2+1) \quad . \qquad \text{(A2.32)}$$

Using (A2.29) and (A2.31), we have

$$\|H_{n,m}^{(L_1)} - H_{n,m}^{(L_2)}\| < \|H_{n,L_2}^{(L_1)} - H_{n,L_2}^{(L_2)}\|$$

$$+ K_0 \sum_{\ell=(L_2+1)}^{m} (1+\varepsilon_1)^n \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}$$

$$< \|H_{n,L_2}^{(L_1)} - H_{n,L_2}^{(L_2)}\|$$

$$+ K_0 \sum_{\ell=(L_2+1)}^{L_1} (1+\varepsilon_1)^N \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|} \quad ,$$

$$0 \le n \le (N-1), \; (L_2+1) \le m \le L_1 \quad . \qquad \text{(A2.33)}$$

Using (A2.30) and (A2.32), we have

$$\|H_{n+1,m}^{(L_1)} - H_{n+1,m}^{(L_2)}\| < \|H_{n+1,-(L_1+1)}^{(L_1)} - H_{n+1,-(L_1+1)}^{(L_2)}\|$$

$$+ K_0 \sum_{\ell=-L_1}^{m} (1+\varepsilon_1)^{(n+1)} \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}$$

$$< \;\| H^{(L_1)}_{n+1,-(L_1+1)} - H^{(L_2)}_{n+1,-(L_1+1)} \|$$

$$+ \; K_0 \sum_{\ell=-L_1}^{-(L_2+1)} (1+\varepsilon_1)^N \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|} \quad ,$$

$$0 \le n \le (N-1), \; -L_1 \le m \le -(L_2+1) \quad . \qquad \text{(A2.34)}$$

Using (A2.8), we have that

$$\| H^{(L_1)}_{k,m} - H^{(L_2)}_{k,m} \| \;=\; \| H^{(L_1)}_{n,L_1} - H^{(L_2)}_{n,L_1} \| \quad ,$$

$$(n,L_1) \le (k,m) \le (n+1,-L_1-1), \; 0 \le n \le (N-1) \quad . \qquad \text{(A2.35)}$$

Combining (A2.33), (A2.34), and (A2.35) we have

$$\| H^{(L_1)}_{k,m} - H^{(L_2)}_{k,m} \| \;<\; \| H^{(L_1)}_{n,L_2} - H^{(L_2)}_{n,L_2} \|$$

$$+ \; 2K_0 \sum_{\ell=(L_2+1)}^{L_1} (1+\varepsilon_1)^N \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|} \quad ,$$

$$0 \le n \le N-1, \; (n,L_2) < (k,m) < (n+1,-L_2) \quad . \qquad \text{(A2.36)}$$

The behavior of $\| H^{(L_1)}_{n,m} - H^{(L_2)}_{n,m} \|$ is completely described by (A2.20), (A2.24), and (A2.36). Combining (A2.20) with (A2.36), for n=0, we have

$$||H_{k,m}^{(L_1)} - H_{k,m}^{(L_2)}|| < 2K_0(1+\varepsilon_1)^N \sum_{\ell=(L_2+1)}^{L_1} \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|} ,$$

$$(0,0) \leq (k,m) \leq (1,-L_2) \quad . \qquad \text{(A2.37)}$$

Combining (A2.37) with (A2.24) for n=1, we have

$$||H_{k,m}^{(L_1)} - H_{k,m}^{(L_2)}|| < K_1 [2K_0(1+\varepsilon_1)^N \sum_{\ell=L_2+1}^{L_1} \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|}] ,$$

$$(0,0) \leq (k,m) \leq (1,L_2) \quad . \qquad \text{(A2.38)}$$

Successively using (A2.24) and (A2.36), it can b shown that

$$||H_{n,m}^{(L_1)} - H_{n,m}^{(L_2)}|| < (\sum_{s=1}^{N} K_1^s) 2K_0(1+\varepsilon_1)^N \sum_{\ell=L_2+1}^{L_1} \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|} ,$$

$$(0,0) \leq (n,m) \leq (N,M), \; L_1 > L_2 > |M| \quad . \qquad \text{(A2.39)}$$

The bound goes to zero as L_1 and L_2 go to infinity, so we have proved the uniform Cauchy convergence of the filter sequence, $\{H_{n,m}^{(L)}(z_1,z_2)\}$.

All that remains to be proved is that $H_{n,m}(z_1,z_2)$ converges uniformly to limit filters as m goes to plus and minus infinity, and that $H_{n,m}(z_1,z_2)$ is minimum-phase. To prove that $H_{n,m}(z_1,z_2)$ converges uniformly to $H_{n,+\infty}(z_1,z_2)$ as m goes to plus infinity we proceed as follows:

$$\|H_{n,+\infty}-H_{n,m}\|$$

$$\leq \|H_{n,+\infty}-H^{(L)}_{n,+\infty}\| + \|H_{n,m}-H^{(L)}_{n,+\infty}\|$$

$$\leq \|H_{n,+\infty}-H^{(L)}_{n,+\infty}\| + \|H_{n,m}-H^{(L)}_{n,m}\|$$

$$+ \|H^{(L)}_{n,+\infty}-H^{(L)}_{n,m}\| \quad , \quad 0\leq n\leq N-1 \quad . \tag{A2.40}$$

Using (A2.3), (A2.6) and (A2.8) we can show that, for all L greater than m,

$$H^{(L)}_{n,+\infty}(z_1,z_2) = H^{(L)}_{n,L}(z_1,z_2)$$

$$= H^{(L)}_{n,m}(z_1,z_2) - \sum_{\ell=m+1}^{L} \rho(n,\ell)$$

$$\cdot\; z_1^{-n} z_2^{-\ell} H^{(L)}_{n,\ell-1}(1/z_1,1/z_2) \quad . \tag{A2.41}$$

Substituting (A2.41) into (A2.40) we have

$$\|H_{n,+\infty}-H_{n,m}\|$$

$$< \|H_{n,+\infty}-H^{(L)}_{n,+\infty}\| + \|H_{n,m}-H^{(L)}_{n,m}\|$$

$$+ \sum_{\ell=m+1}^{L} (1+\varepsilon_1)^n \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|} \|H^{(L)}_{n,\ell-1}\| \quad , \quad L>m \quad .$$

(A2.42)

Letting L go to infinity, we have

$$||H_{n,+\infty}-H_{n,m}|| < K_0(1+\varepsilon_1)^n \sum_{\ell=m+1}^{\infty} \left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|\ell|} .$$

(A2.43)

Therefore,

$$\lim_{m\to\infty} ||H_{n,+\infty}-H_{n,m}|| = 0, \quad 0\le n\le N-1$$

(A2.44)

In precisely the same way we can prove that $H_{n,m}(z_1,z_2)$ converges uniformly to $H_{n,-\infty}(z_1,z_2)$ as m goes to minus infinity, for $1\le n\le N$.

We now prove that the $H_{n,m}(z_1,z_2)$ are minimum-phase. Although $H_{n,m}(z_1,z_2)$ is the uniform limit of a sequence of analytic minimum-phase filters we cannot directly infer from this that $H_{n,m}(z_1,z_2)$ is minimum-phase (though we can infer that $H_{n,m}(z_1,z_2)$ is analytic). We first prove that $H_{n,m}(z_1,z_2)$ is non-zero on the unit circles. In the same way that $|H^{(L)}_{n,m}(z_1,z_2)|$ was upper-bounded, it can be shown that:

$$|H^{(L)}_{n,m}(z_1,z_2)| \ge \left(\prod_{\ell=1}^{L} [1-|\rho(0,\ell)|]\right)\left(\prod_{k=1}^{N-1}\prod_{\ell=-L}^{L} [1-|\rho(k,\ell)|]\right)$$

$$\cdot \left(\prod_{\ell=-L}^{M} [1-|\rho(N,\ell)|]\right) ,$$

$$(0,0)\le(n,m)\le(N,M), \quad |z_1|=|z_2|=1 \quad . \qquad \text{(A2.45)}$$

We claim that as L goes to infinity, the right-hand side of (A2.45) converges to a positive number. We need to show that the infinite products are non-zero. We demonstrate this for the first term. We have

$$\log \prod_{\ell=1}^{\infty} [1-|\rho(0,\ell)|] = \sum_{\ell=1}^{\infty} \log[1-|\rho(0,\ell)|]$$

$$= - \sum_{\ell=1}^{\infty} \log\left(1 + \frac{|\rho(0,\ell)|}{(1-|\rho(0,\ell)|)}\right)$$

$$\geq - \sum_{\ell=1}^{\infty} \frac{|\rho(0,\ell)|}{(1-|\rho(0,\ell)|)} \quad . \tag{A2.46}$$

The denominator of the series is lower-bounded, and the numerator is exponentially decaying; therefore

$$\log \prod_{\ell=1}^{\infty} [1-|\rho(0,\ell)|] > -\infty \quad . \tag{A2.47}$$

Therefore $H_{n,m}(z_1,z_2)$ is non-zero on the unit circles, and because of continuity it must be non-zero in some neighborhood of the unit circles. Consequently, $H^{-1}_{n,m}(z_1,z_2)$ exists and is analytic in some neighborhood of the unit circles.

All that remains to be shown is that $H^{-1}_{n,m}(z_1,z_2)$ is causal. We do this by showing that <u>on the unit circles</u>, $[H^{(L)}_{n,m}(z_1,z_2)]^{-1}$ converges uniformly to $H^{-1}_{n,m}(z_1,z_2)$. We have

$$[H^{-1}_{n,m}(z_1,z_2)-H^{(L)^{-1}}_{n,m}(z_1,z_2)]$$

$$= \frac{[H^{(L)}_{n,m}(z_1,z_2)-H_{n,m}(z_1,z_2)]}{H^{(L)}_{n,m}(z_1,z_2)H_{n,m}(z_1,z_2)} \quad . \qquad \text{(A2.48)}$$

The denominator of (A2.48) is lower-bounded, and the numerator converges uniformly to zero, for all z_1 and z_2 on the unit circles, as L goes to infinity. Therefore, $H^{(L)-1}_{n,m}(z_1,z_2)$ converges uniformly to $H^{-1}_{n,m}(z_1,z_2)$ on the unit circles. Since $H^{(L)-1}_{n,m}(z_1,z_2)$ is causal, we therefore have that $H^{-1}_{n,m}(z_1,z_2)$ is causal.

A2.2 Proof of Uniqueness Part of Theorem 4.3(a)

Given a 2-D reflection coefficient sequence, $\{\rho(k,\ell);\ (0,0)<(k,\ell)\leq(N,M)\}$, where

$$|\rho(k,\ell)| < (1+\varepsilon_0)^{-|\ell|} \quad , \quad \varepsilon_0>0 \quad , \qquad \text{(A2.49)}$$

we want to prove that there is at most one 2-D sequence of 2-D analytic minimum-phase filters, $\{H_{n,m}(z_1,z_2);$ $(0,0)\leq(n,m)\leq(N,M)\}$, satisfying equations (4.21)-(4.26). (The existence of such a filter sequence was proved in the previous section.) Therefore we assume the existence of two such filter sequences, $\{H_{n,m}(z_1,z_2)\}$ and $\{H'_{n,m}(z_1,z_2)\}$, and we want to show that $H_{n,m}(z_1,z_2)$ $= H'_{n,m}(z_1,z_2)$. Beginning with n=0, it is trivial to prove that

$$H_{0,m}(z_1,z_2) = H'_{0,m}(z_1,z_2) \quad , \quad 0 \le m \le \infty \quad . \qquad \text{(A2.50)}$$

For n=1, we have

$$[H_{1,m}(z_1,z_2)-H'_{1,m}(z_1,z_2] = [H_{1,m-1}(z_1,z_2)-H'_{1,m-1}(z_1,z_2)]$$

$$-\rho(1,m)z_1^{-1}z_2^{-m}[H_{1,m-1}(1/z_1,1/z_2)-H'_{1,m-1}(1/z_1,1/z_2)] \quad ,$$

$$-\infty<m<\infty \quad . \qquad \text{(A2.51)}$$

Denoting the least upper bound for $|H_{1,m}(z_1,z_2)-H'_{1,m}(z_1,z_2)|$, in some neighborhood of the unit circles, by $||H_{1,m}-H'_{1,m}||$, we can show (using A2.51) that for all $\ell<m$:

$$||H_{1,m}-H'_{1,m}|| \le ||H_{1,\ell}-H_{1,\ell}||$$

$$\cdot \prod_{t=\ell+1}^{m} [1+|\rho(1,t)|(1+\varepsilon_1)(1+\varepsilon_1)^{|t|}]$$

$$< ||H_{1,\ell}-H'_{1,\ell}|| \prod_{t=-\infty}^{m} [1+(1+\varepsilon_1)\left(\frac{1+\varepsilon_0}{1+\varepsilon_1}\right)^{-|t|}] \quad .$$

$$\text{(A2.52)}$$

The infinite product in (A2.52) is upper-bounded for all m. Furthermore we recall that $H_{1,m}(z_1,z_2)$ and $H'_{1,m}(z_1,z_2)$ converge uniformly to the same limit function as m goes to minus infinity. Consequently, $||H_{1,\ell}-H'_{1,\ell}||$ goes to zero as ℓ goes to minus infinity. Letting ℓ go to minus infinity, (A2.52) becomes

$$||H_{1,m}-H'_{1,m}|| \leq 0 \quad . \tag{A2.53}$$

Therefore,

$$H_{1,m}(z_1,z_2) = H'_{1,m}(z_1,z_2) \quad , \quad -\infty \leq m \leq \infty \quad . \tag{A2.54}$$

Using similar arguments we can inductively prove that

$$H_{n,m}(z_1,z_2) = H'_{n,m}(z_1,z_2) \quad , \quad (0,0) \leq (n,m) \leq (N,M) \quad . \tag{A2.55}$$

A2.3 Proof of Existence Part of Theorem 4.3(b)

We are given a 2-D analytic minimum-phase filter,

$$H_{N,M}(z_1,z_2) = [1 - \sum_{(0,0)<(k,\ell)\leq(N,M)} \sum h(N,M;k,\ell) z_1^{-k} z_2^{-\ell}] \quad , \tag{A2.56}$$

and we want to prove the existence of a 2-D reflection coefficient sequence, $\{\rho(k,\ell);\ (0,0)<(k,\ell)\leq(N,M)\}$, where

$$|\rho(k,\ell)| < (1+\varepsilon)^{-|\ell|} \quad , \quad \varepsilon>0 \quad , \tag{A2.57}$$

and a 2-D sequence of 2-D analytic minimum-phase filters, $\{H_{n,m}(z_1,z_2);\ (0,0)\leq(n,m)<(N,M)\}$, such that equations (4.21)-(4.26) are satisfied.

Our proof uses Theorem 4.1. Arbitrarily letting $P_{N,M}$ equal 1, we define a 2-D positive-definite analytic autocorrelation sequence by the formula:

$$r(k,\ell) = \frac{1}{(2\pi j)^2} \oint_{|z_1|=1} \oint_{|z_2|=1} \frac{dz_1 dz_2}{H_{N,M}(z_1,z_2)H_{N,M}(1/z_1,1/z_2)} ,$$

$$(0,0) \leq (k,\ell) \leq (N,M) \quad . \qquad \text{(A2.58)}$$

According to Theorem 4.1(a) we can generate a 2-D sequence of 2-D analytic minimum-phase filters, $\{H_{n,m}(z_1,z_2);\ (0,0) \leq (n,m) \leq (N,M)\}$, by solving the appropriate sets of normal equations, using the autocorrelation sequence (A2.58). (Theorem 4.1(b) guarantees that the original filter, $H_{N,M}(z_1,z_2)$, can be recovered from the autocorrelation sequence.) According to Theorem 4.2, the filters are related by the formula

$$H_{n,m}(z_1,z_2) = H_{n,m-1}(z_1,z_2) - \rho(n,m) z_1^{-n} z_2^{-m} H_{n,m-1}(1/z_1,1/z_2) ,$$

$$(0,0) < (n,m) \leq (N,M) \quad , \qquad \text{(A2.59)}$$

where $$|\rho(n,m)| < 1 \quad . \qquad \text{(A2.60)}$$

Consequently, the only remaining non-trivial part of the proof is to show that the reflection coefficient sequence $\rho(n,m)$, decays exponentially fast to zero as m goes to plus or minus infinity.

As m goes to <u>minus</u> infinity, we can show this directly, using our Neumann series solution for $H_{n,m}(z_1,z_2)$. Using (A1.19) and (A1.49) we have:

$$\rho(n,m) = h(n,m;n,m)$$

$$= b(n,m;0) \quad , \quad 1 \leq n \leq N \quad , \tag{A2.61}$$

and

$$|b(n,m;0)| < \frac{(1+\varepsilon_0)^m}{1 - \left[\dfrac{(1+\varepsilon_0)^{m-1}}{1-(1+\varepsilon_0)^{-2}}\right]^2} \quad , \quad \varepsilon_0 > 0 \quad , \quad 1 \leq n \leq N. \tag{A2.62}$$

To prove the exponential decay of $\rho(n,m)$ as m goes to plus infinity, we use an argument similar to one used by Grenander and Szego for the 1-D case [21]. We observe that

$$P_{n,m} = E\{[x(k,\ell) - \sum\sum_{(0,0)<(s,t)\leq(n,m)} h(n,m;s,t)x(k-s,\ell-t)]^2\}$$

$$\leq E\{[x(k,\ell) - \sum\sum_{(0,0)<(s,t)\leq(n,m)} h(n,+\infty;s,t)x(k-s,\ell-t)]^2\}$$

$$= E\{([x(k,\ell) - \sum\sum_{(0,0)<(s,t)\leq(n,+\infty)} h(n,+\infty;s,t)x(k-s,\ell-t)]$$

$$+ [\sum_{t=m+1}^{\infty} h(n,+\infty;n,t)x(k-n,\ell-t)])^2\}$$

$$= P_{n,+\infty} + E\{[\sum_{t=m+1}^{\infty} h(n,+\infty;n,t)x(k-n,\ell-t)]^2\} \quad . \tag{A2.63}$$

Therefore

$$[P_{n,m}-P_{n,+\infty}] \leq E\{[\sum_{t=m+1}^{\infty} h(n,+\infty;n,t)x(k-n,\ell-t)]^2\} \quad .$$

(A2.64)

But since $H_{n,+\infty}(z_1,z_2)$ is analytic, it follows that $h(n,+\infty;n,t)$ decays exponentially to zero as t goes to plus infinity. Therefore we have

$$[P_{n,m}-P_{n,+\infty}] \leq c(1+\varepsilon_1)^{-m} \quad , \qquad \text{(A2.65)}$$

where c and ε_1 are positive constants. Recalling that

$$P_{n,m+1} = P_{n,m}[1-\rho^2(n,m+1)] \quad , \qquad \text{(A2.66)}$$

we have that

$$\frac{1}{P_{n,m+1}} = \frac{1}{P_{n,m}} \frac{1}{[1-\rho^2(n,m+1)]} = \frac{1}{P_{n,m}} \left[1 + \frac{\rho^2(n,m+1)}{1-\rho^2(n,m+1)}\right]$$

$$= \frac{1}{P_{n,m}} + \frac{\rho^2(n,m+1)}{P_{n,m+1}} \quad . \qquad \text{(A2.67)}$$

Using (A2.67) it is easy to show that

$$\frac{1}{P_{n,\ell}} = \frac{1}{P_{n,m}} + \sum_{t=m+1}^{\ell} \frac{\rho^2(n,t)}{P_{n,t}} \quad , \qquad \ell>m. \qquad \text{(A2.68)}$$

Since (A2.65) implies that $P_{n,m}$ converges to $P_{n,+\infty}$, we can let ℓ go to plus infinity in (A2.68):

$$\frac{1}{P_{n,+\infty}} = \frac{1}{P_{n,m}} + \sum_{t=m+1}^{\infty} \frac{\rho^2(n,t)}{P_{n,t}} . \tag{A2.69}$$

Substituting (A2.69) into (A2.65) we have

$$[P_{n,m} - P_{n,+\infty}] = P_{n,m} - \frac{1}{\dfrac{1}{P_{n,m}} + \displaystyle\sum_{t=m+1}^{\infty} \frac{\rho^2(n,t)}{P_{n,t}}}$$

$$= \frac{\displaystyle\sum_{t=m+1}^{\infty} \frac{\rho^2(n,t)}{P_{n,t}}}{\dfrac{1}{P_{n,m}} \left[\dfrac{1}{P_{n,m}} + \displaystyle\sum_{t=m+1}^{\infty} \frac{\rho^2(n,t)}{P_{n,t}}\right]}$$

$$= \frac{\displaystyle\sum_{t=m+1}^{\infty} \frac{\rho^2(n,t)}{P_{n,t}}}{\dfrac{1}{P_{n,m}} \dfrac{1}{P_{n,+\infty}}} \leq c(1+\varepsilon_1)^{-m} . \tag{A2.70}$$

Or

$$\sum_{t=m+1}^{\infty} \frac{\rho^2(n,t)}{P_{n,t}} \leq \frac{c(1+\varepsilon_1)^{-m}}{P_{n,m} P_{n,+\infty}} \tag{A2.71}$$

Using the fact that $P_{n,t}$ is non-increasing for increasing values of t, we have

$$\sum_{t=m+1}^{\infty} \rho^2(n,t) \leq \frac{c}{P_{n,+\infty}} (1+\varepsilon_1)^{-m} . \tag{A2.72}$$

Therefore

$$|\rho(n,m+1)| \leq \sqrt{\frac{c}{P_{n,+\infty}}}\,(1+\varepsilon_1)^{-m/2} \quad . \qquad (A2.73)$$

In the process of proving Theorem 4.1(a) we proved that $H_{n,m}(z_1,z_2)$ converges uniformly to $H_{n,-\infty}(z_1,z_2)$ as m goes to minus infinity. It is trivial to prove that $H_{n,m}(z_1,z_2)$ converges uniformly to $H_{n,+\infty}(z_1,z_2)$ as m goes to plus infinity. We have

$$H_{n,m}(z_1,z_2) = H_{n,0}(z_1,z_2) - \sum_{\ell=1}^{m} \rho(n,\ell)\, z_1^{-n} z_2^{-\ell} H_{n,\ell-1}(1/z_1,1/z_2) \quad , \qquad m>0 \quad . \qquad (A2.74)$$

Using the fact that the $H_{n,\ell}(z_1,z_2)$ are uniformly bounded, and that the reflection coefficients are exponentially decaying, we can apply the Weierstrass M-test to show that $H_{n,m}(z_1,z_2)$ converges uniformly to $H_{n,+\infty}(z_1,z_2)$ in some neighborhood of the unit circles

A2.4 Proof of Uniqueness Part of Theorem 4.3(b)

We are given a 2-D analytic minimum-phase filter, $H_{N,M}(z_1,z_2)$. In section A2.3, we proved the existence of a 2-D reflection coefficient sequence, $\{\rho(k,\ell);\ (0,0) < (k,\ell) \leq (N,M)\}$, where

$$|\rho(k,\ell)| < (1+\varepsilon)^{-|\ell|} \quad , \quad \varepsilon>0 \quad , \tag{A2.75}$$

and a 2-D sequence of 2-D analytic minimum-phase filters, $\{H_{n,m}(z_1,z_2);\ (0,0)\leq(n,m)<(N,M)\}$, satisfying equations (4.21)-(4.26). In this section we prove that the reflection coefficient sequence and the filter sequence are unique.

Therefore we assume the existence of some other reflection coefficient sequence, $\{\rho'(k,\ell);\ (0,0)<(k,\ell)\leq(N,M)\}$, where

$$|\rho'(k,\ell)| < (1+\varepsilon')^{-|\ell|} \quad , \quad \varepsilon'>0, \tag{A2.76}$$

and some other 2-D sequence of 2-D analytic minimum-phase filters, $\{H'_{n,m}(z_1,z_2);\ (0,0)\leq(n,m)<(N,M)\}$, such that equations (4.21)-(4.26) are satisfied. (We assume that $\{H_{n,m}(z_1,z_2)\}$ and $\{\rho(n,m)\}$ are obtained as in section A2.3.) For n=N we have

$$H_{N,m}(z_1,z_2) = H_{N,m-1}(z_1,z_2) - \rho(N,m)z_1^{-N}z_2^{-m}H_{N,m-1}(1/z_1,1/z_2) \quad , \quad m\leq M \quad , \tag{A2.77}$$

and

$$H'_{N,m}(z_1,z_2) = H'_{N,m-1}(z_1,z_2) - \rho'(N,m)z_1^{-N}z_2^{-m}H'_{N,m-1}(1/z_1,1/z_2) \quad , \quad m\leq M \quad , \tag{A2.78}$$

where $H_{N,M}(z_1,z_2) = H'_{N,M}(z_1,z_2)$. (A2.79)

As in section A1.4 we can argue that the recursions (A2.77) and (A2.78) can be "run backwards," thereby proving that

$$H_{N,m}(z_1,z_2) = H'_{N,m}(z_1,z_2) \quad , \quad m \leq M \qquad \text{(A.280)}$$

and $$\rho(N,m) = \rho'(N,m) \quad , \quad m \leq M \quad . \qquad \text{(A2.81)}$$

Next, we consider the filter sequence $H'_{N-1,m}(z_1,z_2)$, $-\infty \leq m \leq \infty$. We intend to prove that this filter sequence can be generated, via the normal equations, from the same autocorrelation sequence used to generate the sequence $H_{N-1,m}(z_1,z_2)$. Using Theorem 4.1(a) we will then be able to argue that $H'_{N-1,m}(z_1,z_2) = H_{N-1,m}(z_1,z_2)$ and $\rho'(N,m) = \rho(N,m)$.

According to (A2.80) we have

$$H'_{N-1,+\infty}(z_1,z_2) = H_{N-1,+\infty}(z_1,z_2) \quad . \qquad \text{(A2.82)}$$

We define the sequence, $P'_{N-1,m}$, by the recursion

$$P'_{N-1,m} = P'_{N-1,m-1}[1-\rho'^2(N-1,m)] \quad , \quad -\infty < m < \infty \quad , \qquad \text{(A2.83)}$$

where $$P'_{N-1,+\infty} = P_{N-1,+\infty} \quad . \qquad \text{(A2.84)}$$

According to Theorem 4.1, associated with each $\{H'_{N-1,m}(z_1,z_2), P'_{N-1,m}\}$ is a unique positive-definite analytic autocorrelation sequence which we denote $\{r'(N-1,m;k,\ell);\ (0,0) \leq (k,\ell) \leq (N-1,m)\}$. The autocorrelation sequence is given by the formula

$$r'(N-1,m;k,\ell) = \frac{1}{(2\pi j)^2} \oint_{|z_1|=1} \oint_{|z_2|=1}$$

$$\cdot \frac{z_1^{k-1} z_2^{\ell-1} P'_{N-1,m} dz_1 dz_2}{H'_{N-1,m}(z_1,z_2) H'_{N-1,m}(1/z_1,1/z_2)} ,$$

$$(0,0) \leq (k,\ell) \leq (N-1,m) \quad . \qquad \text{(A2.85)}$$

We claim that

$$r'(N-1,m;k,\ell) = r'(N-1,m+1;k,\ell) \quad , \quad (0,0) \leq (k,\ell) \leq (N-1,m) \quad .$$

(A2.86)

To see this we write the normal equations for $H'_{N,m+1}(z_1,z_2)$:

$$[r'(N-1,m+1;s,t) - \sum_{(k,\ell)}\sum h'(N-1,m+1;k,\ell) r'(N-1,m+1;s-k,t-\ell)]$$

$$= P'_{N-1,m+1} \delta_s \delta_t \quad , \quad (0,0) \leq (s,t) \leq (N-1,m+1) \quad .$$

(A2.87)

Recalling that

$$H'_{N-1,m+1}(z_1,z_2) = H'_{N-1,m}(z_1,z_2) - \rho(N-1,m+1) z_1^{-(N-1)} z_2^{-(m-1)}$$

$$\cdot H'_{N-1,m}(1/z_1,1/z_2) \quad , \qquad \text{(A2.88)}$$

we have that

$$[r'(N-1,m+1;s,t) - \sum_{(k,\ell)}\sum h'(N-1,m;k,\ell)r'(N-1,m+1;s-k,t-\ell)]$$

$$- \rho(N-1,m+1)[r'(N-1,m+1;s-N+1,t-m-1)$$

$$- \sum_{(k,\ell)}\sum h'(N-1,m;k,\ell)r'(N-1,m+1;s-N+1+k,t-m-1+\ell)]$$

$$= P'_{N-1,m+1}\delta_s\delta_t \quad , \quad (0,0)\le(s,t)\le(N-1,m+1) \quad . \qquad \text{(A2.89)}$$

Making the substitution

$$s' = N-1-s \; , \quad t' = m+1-t \quad , \qquad \text{(A2.90)}$$

we have

$$[r'(N-1,m+1;s'-N+1,t'-m-1) - \sum_{(k,\ell)}\sum h'(N-1,m;k,\ell)$$

$$r'(N-1,m+1;s'-N+1+k,t'-m-1+\ell)]$$

$$- \rho(N-1,m+1)[r'(N-1,m+1;s',t')$$

$$- \sum_{(k,\ell)}\sum h'(N-1,m;k,\ell)r'(N-1,m+1;s'-k,t'-\ell)]$$

$$= P'_{N-1,m+1}\delta_{N-1-s'}\delta_{m+1-t'} \quad , \quad (0,0)\le(s',t')\le(N-1,m-1) \quad .$$

(A2.91)

Comparing (A2.89) and (A2.91) it can be shown that

$$[r'(N-1,m+1;s,t) - \sum_{(k,\ell)}\sum h'(N-1,m;k,\ell)r'(N-1,m+1;s-k,t-\ell)]$$

$$= P'_{N-1,m}\delta_s\delta_t \quad , \quad (0,0)\le(s,t)\le(N-1,m) \quad . \qquad \text{(A2.92)}$$

Given (A2.92), Theorem 4.1(b) implies that (A2.86) is satisfied.

Using (A2.86) we can argue inductively that

$$r'(N-1,m;k,\ell) = r'(N-1,v;k,\ell) \quad ,$$

$$(0,0) \leq (k,\ell) \leq (N-1,m) \quad , \quad v>m \quad . \tag{A2.93}$$

Substituting (A2.85) into the right-hand side of (A2.93) we have that

$$r'(N-1,m;k,\ell)$$

$$= \frac{1}{(2\pi j)^2} \oint_{|z_1|=1} \oint_{|z_2|=1} \frac{z_1^{k-1} z_2^{\ell-1} P'_{N-1,v} dz_1 dz_2}{H'_{N-1,v}(z_1,z_2) H'_{N-1,v}(1/z_1,1/z_2)} \quad ,$$

$$(0,0) \leq (k,\ell) \leq (N-1,m) \quad , \quad v>m \quad . \tag{A2.94}$$

Letting v go to plus infinity (which we are permitted to do since $H'_{N-1,v}(z_1,z_2)$ converges uniformly to $H_{N-1,+\infty}(z_1,z_2)$ and $P'_{N-1,v}$ converges to $P_{N-1,+\infty}$) we have

$$r'(N-1,m;k,\ell)$$

$$= \frac{1}{(2\pi j)^2} \oint_{|z_1|=1} \oint_{|z_2|=1} \frac{z_1^{k-1} z_2^{\ell-1} P_{N-1,+\infty} dz_1 dz_2}{H_{N-1,+\infty}(z_1,z_2) H_{N-1,+\infty}(1/z_1,1/z_2)}$$

$$= r(k,\ell) \quad , \quad (0,0) \leq (k,\ell) \leq (N-1,m) \quad . \tag{A2.95}$$

Therefore $H'_{N-1,m}(z_1,z_2)$ and $P'_{N-1,m}$ can be obtained from the same autocorrelation sequence used to generate $H_{N-1,m}(z_1,z_2)$ and $P_{N-1,m}$, so $H'_{N-1,m}(z_1,z_2) = H_{N-1,m}(z_1,z_2)$, $P'_{N-1,m} = P_{N-1,m}$, and $\rho'(N-1,m) = \rho(N,m)$.

At this point it is clear that the remainder of the proof can be obtained inductively using the same type of arguments.

CHAPTER 5

THE DESIGN OF 2-D MINIMUM-PHASE WHITENING FILTERS IN THE REFLECTION COEFFICIENT DOMAIN

In this chapter we use the 2-D reflection coefficient representation as a tool for designing 2-D minimum-phase whitening filters. By designing 2-D filters in the reflection coefficient domain, we automatically satisfy the minimum-phase condition merely by restricting the reflection coefficient magnitudes to be less than one.

We consider the two general classes of 2-D linear prediction problems: spectral factorization and autoregressive model fitting. We recall that the spectral factorization problem is a deterministic problem; we are given the exact autocorrelation function of the random process (or an autocorrelation function assumed to be exact), and we wish to find a minimum-phase approximation to the minimum-phase whitening filter. In contrast, the autoregressive model fitting problem is a stochastic estimation problem; we have a finite set of samples from the random process itself, and we wish to estimate the minimum-phase whitening filter by modeling it as an FIR, minimum-phase filter. Because of its stochastic nature, autoregressive model fitting is inherently more difficult than spectral factorization.

Our approach to both problems is to represent the approximate whitening filter in terms of a finite number of reflection coefficients, and to optimize over the reflection coefficients subject to the constraint that their magnitudes are less than one. Clearly the utility of the 2-D reflection coefficient representation as a practical tool depends on our finding effective and computationally tractable algorithms for choosing the reflection coefficients.

The problem of optimally choosing the reflection coefficients has not been exhaustively studied in this thesis research. Instaed we have developed two convenient, but generally suboptimal methods of spectral factorization and autoregressive model fitting. In both algorithms, the reflection coefficients are chosen sequentially (in a finite raster scan fashion), each new reflection coefficient being chosen according to a least-squares criterion. For the 1-D case, the spectral factorization algorithm reduces to the 1-D Levinson algorithm, and the autoregressive model fitting algorithm reduces to the Burg algorithm. A computer program was written to implement the spectral factorization algorithm, and numerical results are presented for two examples.

5.1 Equations Relating the Filter to the Reflection Coefficients

As indicated earlier, our approach to designing 2-D minimum-phase filters is based on representing the filter in terms of a finite-order 2-D reflection coefficient sequence. A convenient geometry for the reflection coefficient sequence is the rectangular geometry illustrated in Fig. 5.1(a). (The rectangular geometry is used merely for the sake of convenience. There is no reason why some other geometry could not be used.) We denote the reflection coefficient sequence by $\{\hat{\rho}(n,m);\ (n=0,1\leq m\leq M),(1\leq n\leq N,$ $-M\leq m\leq N)\}$, where N and M are positive integers. As we saw in the previous chapter, we can obtain an FIR filter from the reflection coefficients, denoted $\hat{H}_{N,M}(z_1,z_2)$, by recursively computing a 2-D finite-order sequence of FIR filters, $\{\hat{H}_{n,m}(z_1,z_2);\ (n=0,0\leq m\leq M),\ (1\leq n\leq N,-M\leq m\leq M)\}$. The order in which the $\hat{H}_{n,m}(z_1,z_2)$ are computed follows a finite raster scan.

The recursion proceeds as follows (the equations are nearly identical to equations (A2.2)-(A2.8)): We begin with

$$\hat{H}_{0,0}(z_1,z_2) = 1 \quad ; \tag{5.1}$$

we then recursively compute $\hat{H}_{0,m}(z_1,z_2)$ as follows:

$$\hat{H}_{0,m}(z_1,z_2) = \hat{H}_{0,m-1}(z_1,z_2) - \hat{\rho}(0,m)\,z_2^{-m}\hat{H}_{0,m-1}(1/z_1,1/z_2) \quad , \quad 1\leq m\leq M \quad . \tag{5.2}$$

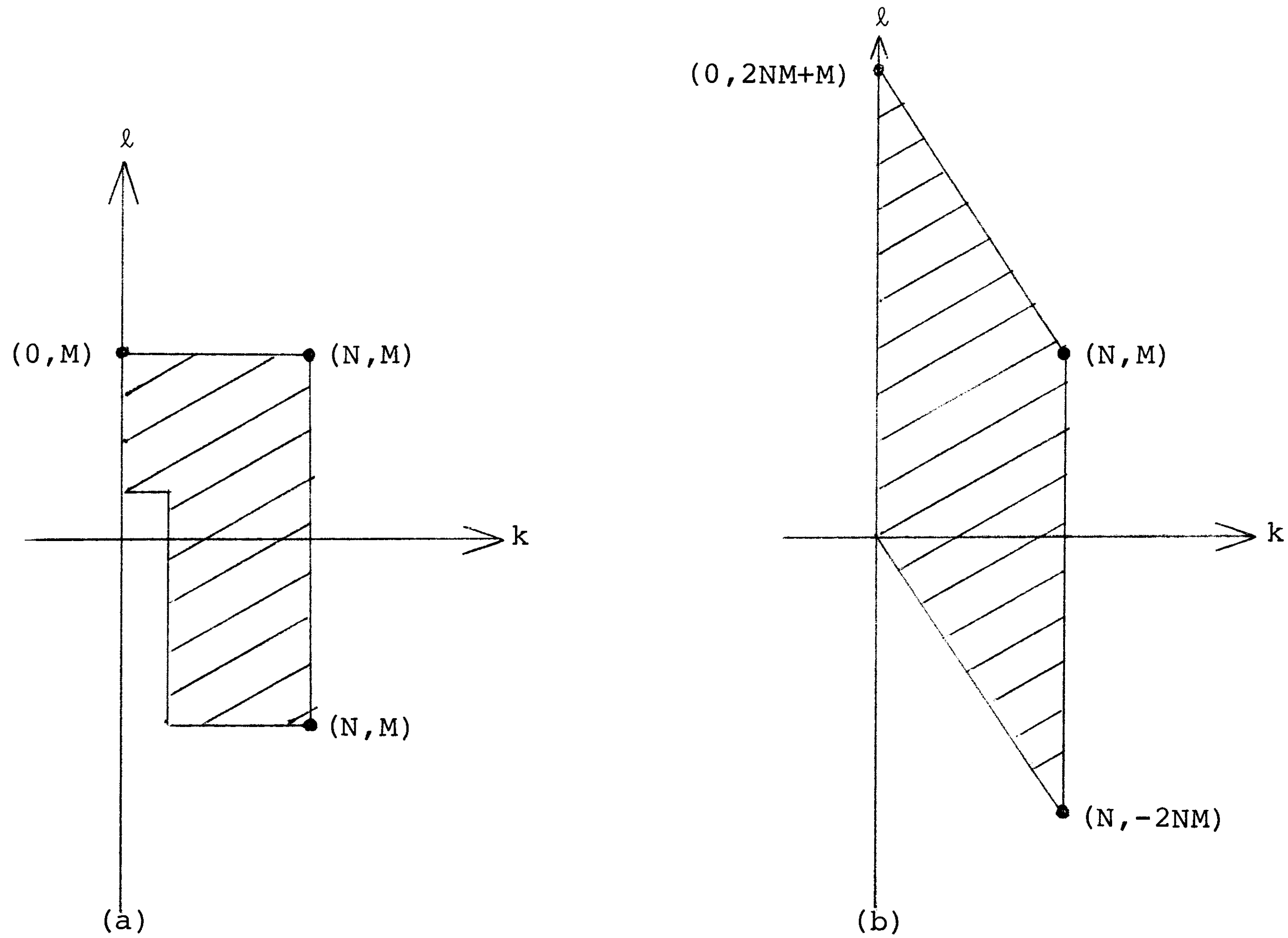


Fig. 5.1 The 2-D reflection coefficient sequence (a) generates the filter (b).

The next column of the recursion begins with the boundary condition,

$$\hat{H}_{1,-(M+1)}(z_1,z_2) = \hat{H}_{0,M}(z_1,z_2) \quad ; \tag{5.3}$$

the remainder of the column is recursively computed by the formula:

$$\hat{H}_{1,m}(z_1,z_2) = \hat{H}_{1,m-1}(z_1,z_2) - \hat{\rho}(1,m)z_1^{-1}z_2^{-m}\hat{H}_{1,m-1}(1/z_1,1/z_2) \; , \quad -M\leq m\leq M \; . \tag{5.4}$$

In general, within each column of the recursion we have

$$\hat{H}_{n,m}(z_1,z_2) = \hat{H}_{n,m-1}(z_1,z_2) - \hat{\rho}(n,m)z_1^{-n}z_2^{-m}\hat{H}_{n,m-1}(1/z_1,1/z_2) \; , \quad \{n=0,1\leq m\leq M\}, \; \{1\leq n\leq N,-M\leq m\leq M\} \; . \tag{5.5}$$

The transition between adjacent columns of the recursion is

$$\hat{H}_{n,-(M+1)}(z_1,z_2) = \hat{H}_{n-1,M}(z_1,z_2) \quad , \quad 1\leq n\leq N \; . \tag{5.6}$$

One possibly serious disadvantage of the reflection coefficient representation is that the filter $\hat{H}_{N,M}(z_1,z_2)$ has a considerably greater number of non-zero coefficients than its reflection coefficient sequence has. It can be shown that $\hat{H}_{n,m}(z_1,z_2)$ is of the form

$$\hat{H}_{n,m}(z_1,z_2) = 1 - \sum_{\ell=1}^{(2nM+m)} \hat{h}(n,m;0,\ell)\, z_2^{-\ell}$$

$$- \sum_{k=1}^{n} \sum_{\ell=-2kM}^{[2(n-k)M+m]} \hat{h}(n,m;k,\ell)\, z_1^{-k} z_2^{-\ell} \quad ,$$

$$\{n=0, 0\le m\le M\},\ \{1\le n\le N, -M\le m\le M\} \quad . \qquad (5.7)$$

The support for $\hat{H}_{N,M}(z_1,z_2)$ is illustrated in Fig. 5.1(b).

Equation (5.7) can be verified by direct substitution. It is more easily understood by studying Figures 5.2, 5.3, and 5.4. Fig. 5.2 illustrates the geometry of the recursion for $(n,m) = (1,-M)$, Fig. 5.3 illustrates the geometry for $\{1\le n\le N, m=M\}$, and Fig. 5.4 illustrates the geometry for $\{1\le n\le N, 1-M\le m\le M\}$.

The fact that the filter, $\hat{H}_{N,M}(z_1,z_2)$, has more coefficients than the reflection coefficient sequence is an unavoidable property of the 2-D reflection coefficient representation. This effect occurs even if some other geometry is used for the reflection coefficient sequence. In some cases, the "tails" of the filter may be small enough to truncate; Theorem 3.1 implies that if the tails are small enough, the truncated filter will still be minimum-phase.

The reflection coefficient sequence contains approximately 2NM parameters, $\hat{H}_{N,M}(z_1,z_2)$ consists of approximately $2N^2M$ coefficients, and approximately N^3M^2 additions and multiplications are needed to go from the reflection coefficients to the filter.

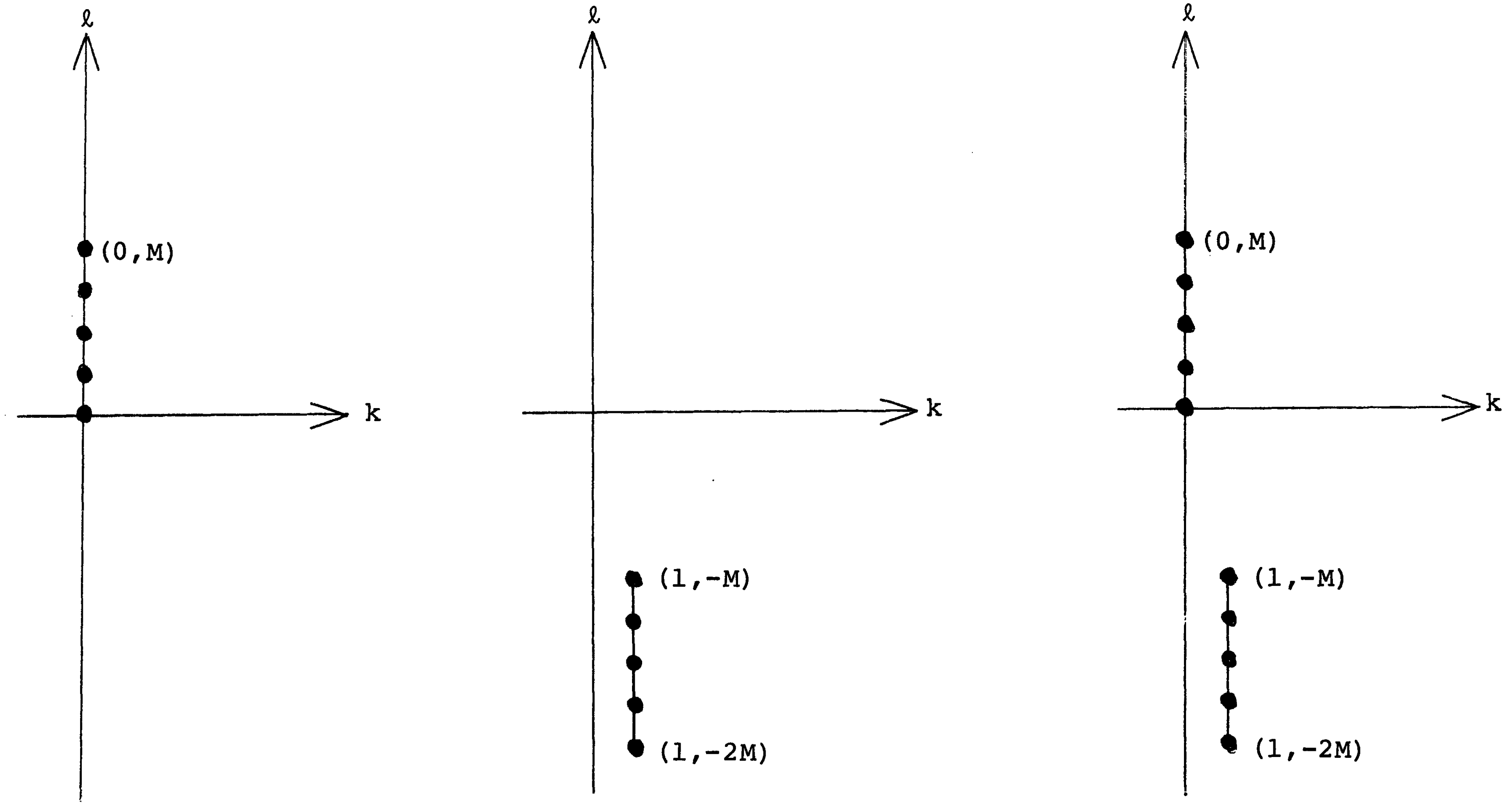
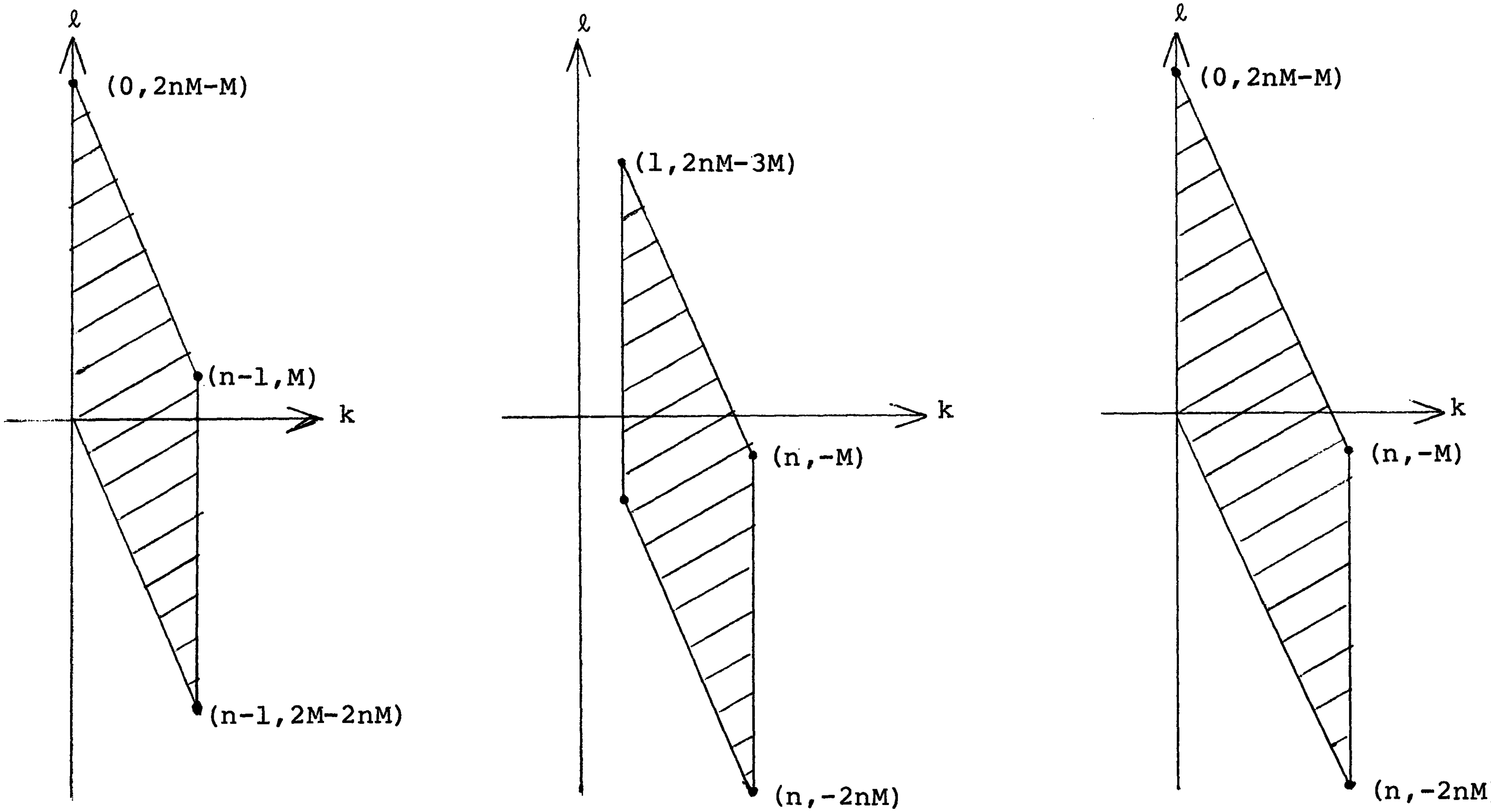


Fig. 5.2 (a) $\hat{H}_{0,M}(z_1,z_2)$; (b) $[z_1^{-1}z_2^{M}\hat{H}_{0,M}(1/z_1,1/z_2)]$;

(c) $\hat{H}_{1,-M}(z_1,z_2)$.



<u>Fig. 5.3</u> (a) $\hat{H}_{n-1,M}(z_1,z_2)$; (b) $[z_1^{-n}z_2^{M}\hat{H}_{n-1,M}(z_1,z_2)$;

(c) $\hat{H}_{n,-M}(z_1,z_2)$

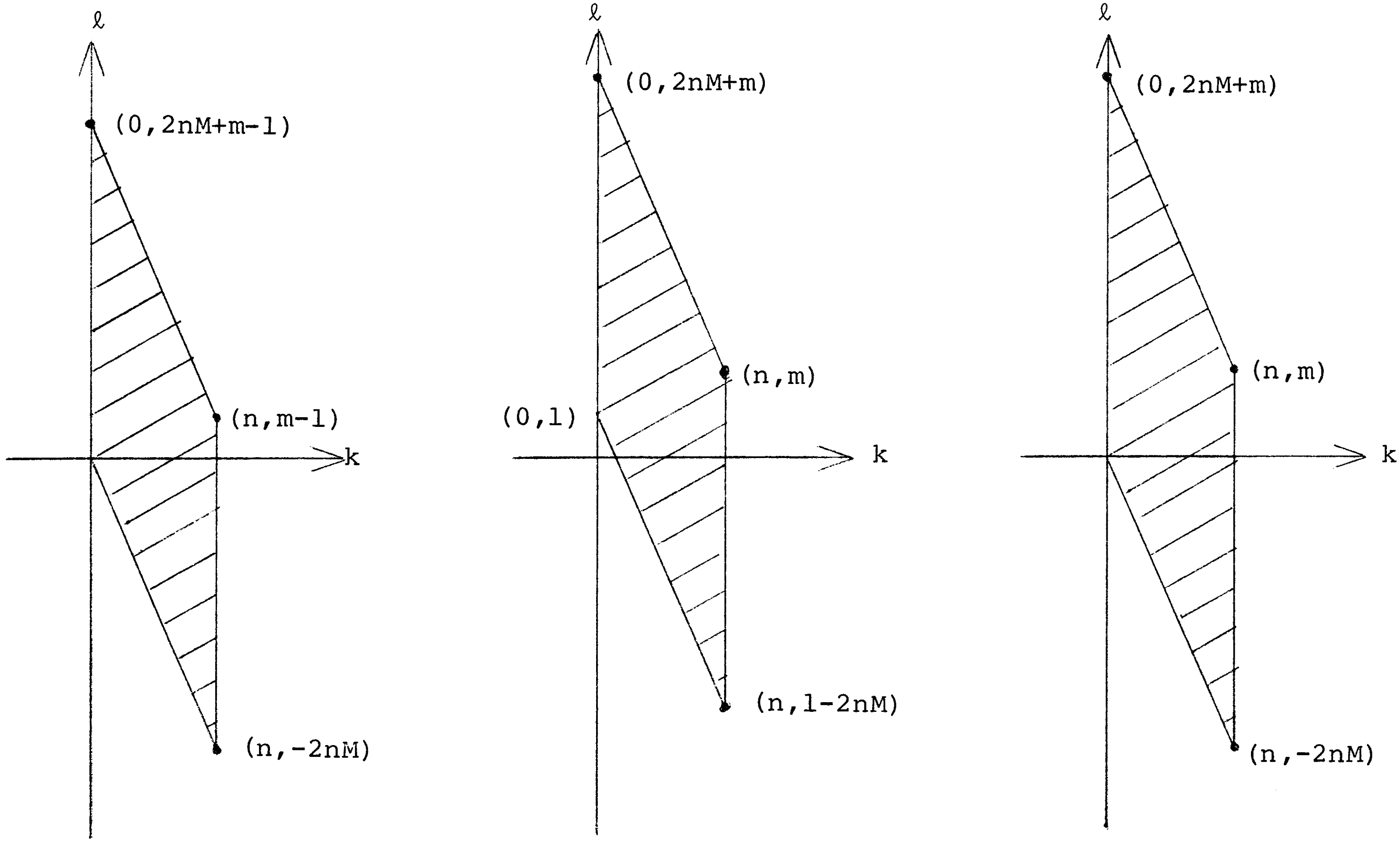


Fig. 5.4 (a) $\hat{H}_{n,m-1}(z_1,z_2)$; (b) $[z_1^{-n}z_2^{-m}\hat{H}_{n,m-1}(1/z_1,1/z_2)$;
(c) $\hat{H}_{n,m}(z_1,z_2)$.

5.2 A 2-D Spectral Factorization Algorithm

Given a 2-D power density spectrum, the problem is to choose the reflection coefficient sequence, $\{\hat{\rho}(n,m);\ (n=0, 1\leq n\leq M), (1\leq n\leq N, -M\leq m\leq M)\}$, for some (N,M), so that the filter $\hat{H}_{N,M}(z_1,z_2)$ is a good approximation to a minimum-phase whitening filter, subject to the constraint that the reflection coefficient magnitudes are less than one. An obvious way to choose the reflection coefficients is to choose them sequentially in a finite raster scan fashion, each new reflection coefficient being chosen to minimize the mean-square prediction error of the new filter, $\hat{H}_{n,m}(z_1,z_2)$. The motivation for this particular approach is based on two observations: 1) in the 1-D case $(N=0)$ the algorithm is simply the 1-D Levinson algorithm; and 2) if the <u>true</u> reflection coefficient sequence, $\rho(n,m)$, is equal to zero for $\{|m|>M, 0\leq n\leq N-1\}$ and for $\{n=N, m<-M\}$, then the procedure will yield the optimal values for the reflection coefficients (optimal in the sense that $\hat{H}_{N,M}(z_1,z_2)$ will equal $H_{N,M}(z_1,z_2)$, or equivalently that $\hat{H}_{N,M}(z_1,z_2)$ will satisfy the normal equations.) In other words, this algorithm is simply the 2-D Levinson algorithm, used under the assumption that the true reflection coefficient sequence, $\rho(n,m)$, vanishes for $\{|m|>M, 0\leq n\leq N-1\}$ and $\{n=N, m<-M\}$. Of course this situation will never occur in practice. Nevertheless, if the reflection coefficients that have been "skipped over" are approximately equal to zero, we intuitively expect that the procedure will yield nearly optimal values for the reflection coefficients.

Considering the algorithm itself, at the beginning of a particular stage of the procedure we have $\hat{H}_{n,m-1}(z_1,z_2)$ and $\hat{P}_{n,m-1}$, $(n=0, 0\leq m\leq M)$ or $(1\leq n\leq N, -M\leq m\leq M)$, where

$$\hat{P}_{n,m-1} = E\{[x(k,\ell) - \Sigma\Sigma\hat{h}(n,m-1;s,t)x(k-s,\ell-t)]^2\} \quad . \qquad (5.8)$$

We then choose the new reflection coefficient, $\hat{\rho}(n,m)$, to minimize the mean-square prediction error associated with the new filter, $\hat{H}_{n,m}(z_1,z_2)$. We have

$$\begin{aligned}\hat{P}_{n,m} &= E\{[x(k,\ell) - \sum_{(s,t)}\sum \hat{h}(n,m;s,t)x(k-s,\ell-t)]^2\} \\ &= E\{([x(k,\ell) - \sum_{(s,t)}\sum \hat{h}(n,m-1;s,t)x(k-s,\ell-t)] \\ &\quad - \hat{\rho}(n,m)[x(k-n,\ell-m) - \sum_{(s,t)}\sum \hat{h}(n,m-1;s,t) \\ &\quad \cdot x(k-n+s,\ell-m+t)])^2\} \\ &= \hat{P}_{n,m-1}[1+\hat{\rho}^2(n,m)] - 2\hat{\rho}(n,m)E\{[x(k,\ell) \\ &\quad - \sum_{(s,t)}\sum \hat{h}(n,m-1;s,t)x(k-s,\ell-t)][x(k-n,\ell-m) \\ &\quad - \sum_{(s,t)}\sum \hat{h}(n,m-1;s,t)x(k-n+s,\ell-m+t)]\} \quad . \qquad (5.9)\end{aligned}$$

Taking the derivative of (5.9) with respect to $\hat{\rho}(n,m)$, and setting it equal to zero, we have that

$$\hat{\rho}(n,m) = \frac{1}{\hat{P}_{n,m-1}} E\{[x(k,\ell) - \sum\sum_{(s,t)} \hat{h}(n,m-1;s,t)x(k-s,\ell-t)]$$

$$\cdot [x(k-n,\ell-m) - \sum\sum_{(s,t)} \hat{h}(n,m-1;s,t)x(k-n+s,\ell-m+t)]\}$$

$$= \frac{1}{\hat{P}_{n,m-1}} \frac{1}{(2\pi j)^2} \oint_{|z_1|=1} \oint_{|z_2|=1} z_1^{n-1} z_2^{m-1} \hat{H}^2_{n,m-1}(z_1,z_2)$$

$$\cdot S(z_1,z_2)dz_1dz_2$$

$$= \frac{1}{\hat{P}_{n,m-1}} [r(n,m) - 2 \sum\sum_{(s,t)} \hat{h}(n,m-1;s,t)r(n-s,m-t)$$

$$+ \sum\sum_{(s_1,t_1)} \sum\sum_{(s_2,t_2)} \hat{h}(n,m-1;s_1,t_1)\hat{h}(n,m-1;s_2,t_2)$$

$$\cdot r(n-s_1-s_2,m-t_1-t_2)] \quad . \tag{5.10}$$

Using Schwartz's inequality, we have that the magnitude of $\hat{\rho}(n,m)$ is less than one. Substituting (5.10) into (5.9), we have that

$$\hat{P}_{n,m} = \hat{P}_{n,m-1}[1 - \hat{\rho}^2(n,m] \quad . \tag{5.11}$$

We note that since the filter $\hat{H}_{n,m-1}(z_1,z_2)$ does not generally satisfy the normal equations, the expression (5.10) does not simplify any further. But computing $\hat{\rho}(n,m)$ directly according to this formula requires an excessive number of computations. A much faster way to compute the reflection coefficient is to compute recursively the autoconvolution, $\hat{H}^2_{n,m-1}(z_1,z_2)$, updating it at each stage of the algorithm, and then to compute $\hat{\rho}(n,m)$ by computing the inner product between $z_1^n z_2^m \hat{H}^2_{n,m-1}(z_1,z_2)$ and $S(z_1,z_2)$. In order to recursively compute the filter autoconvolution, we must also recursively compute the filter autocorrelation. We have

$$\hat{H}_{n,m}(z_1,z_2) = \hat{H}_{n,m-1}(z_1,z_2) - \hat{\rho}(n,m) z_1^{-n} z_2^{-m} \hat{H}_{n,m-1}(1/z_1,1/z_2) \quad . \tag{5.12}$$

It is easy to show that

$$\hat{H}^2_{n,m}(z_1,z_2) = \hat{H}^2_{n,m-1}(z_1,z_2) + \hat{\rho}^2(n,m) z_1^{-2n} z_2^{-2m} \cdot \hat{H}^2_{n,m-1}(1/z_1,1/z_2) - 2\hat{\rho}(n,m) z_1^{-n} z_2^{-m} \cdot \hat{H}_{n,m-1}(z_1,z_2) \hat{H}_{n,m-1}(1/z_1,1/z_2) \quad , \tag{5.13}$$

and

$$\hat{H}_{n,m}(z_1,z_2)\hat{H}_{n,m}(1/z_1,1/z_2) = [1+\hat{\rho}^2(n,m)]\hat{H}_{n,m-1}(z_1,z_2) \cdot \hat{H}_{n,m-1}(1/z_1,1/z_2) - \hat{\rho}(n,m) z_1^{-n} z_2^{-m} \hat{H}^2_{n,m-1}(1/z_1,1/z_2) - \hat{\rho}(n,m) z_1^{n} z_2^{m} \hat{H}^2_{n,m-1}(z_1,z_2) \quad . \tag{5.14}$$

Collecting the various formulas, the entire spectral factorization algorithm can be summarized as follows (the algorithm is expressed here in Z-transform notation, but it is implemented algebraically):

1) Initially,

$$\hat{H}_{0,0}(z_1,z_2) = 1 \quad , \tag{5.15}$$

$$\hat{P}_{0,0} = r(0,0) \quad , \tag{5.16}$$

$$\hat{H}^2_{0,0}(z_1,z_2) = 1 \quad , \tag{5.17}$$

$$\hat{H}_{0,0}(z_1,z_2)\hat{H}_{0,0}(1/z_1,1/z_2) = 1 \quad ; \tag{5.18}$$

2) At the beginning of the $(n,m)^{th}$ stage of the algorithm, $(n=0, 1\leq m\leq M)$ or $(1\leq n\leq N, -M\leq m\leq M)$, we have

a) $\hat{H}_{n,m-1}(z_1,z_2)$,

b) $\hat{P}_{n,m-1}$,

c) $\hat{H}^2_{n,m-1}(z_1,z_2)$,

d) $\hat{H}_{n,m-1}(z_1,z_2)\hat{H}_{n,m-1}(1/z_1,1/z_2)$;

We first compute the new reflection coefficient:

$$\hat{\rho}(n,m) = \frac{1}{\hat{P}_{n,m-1}} [r(n,m) - \sum_{(k,\ell)}\sum \hat{h}^{*2}(n,m-1;k,\ell)r(n-k,m-\ell)] \quad , \tag{5.19}$$

(here we have used the notation:

$$\hat{H}^2_{n,m-1}(z_1,z_2) = [1 - \sum\sum_{(k,\ell)} \hat{h}^{*2}(n,m-1;k,\ell) z_1^{-k} z_2^{-\ell}] \qquad (5.20));$$

We then have the following update equations:

$$\hat{H}_{n,m}(z_1,z_2) = \hat{H}_{n,m-1}(z_1,z_2) - \hat{\rho}(n,m) z_1^{-n} z_2^{-m} \hat{H}_{n,m-1}(1/z_1,1/z_2) , \qquad (5.21)$$

$$\hat{P}_{n,m} = \hat{P}_{n,m-1}[1-\hat{\rho}^2(n,m)] , \qquad (5.22)$$

$$\hat{H}^2_{n,m}(z_1,z_2) = \hat{H}^2_{n,m-1}(z_1,z_2) + \hat{\rho}^2(n,m) z_1^{-2n} z_2^{-2m} \cdot \hat{H}^2_{n,m-1}(1/z_1,1/z_2) - 2\hat{\rho}(n,m) z_1^{-n} z_2^{-m} \cdot \hat{H}_{n,m-1}(z_1,z_2)\hat{H}_{n,m-1}(1/z_1,1/z_2) , \qquad (5.23)$$

$$\hat{H}_{n,m}(z_1,z_2)\hat{H}_{n,m}(1/z_1,1/z_2) = [1 + \hat{\rho}^2(n,m)]\hat{H}_{n,m-1}(z_1,z_2) \cdot \hat{H}_{n,m-1}(1/z_1,1/z_2) - \hat{\rho}(n,m) z_1^{-n} z_2^{-m} \hat{H}^2_{n,m-1}(1/z_1,1/z_2) - \hat{\rho}(n,m) z_1^{n} z_2^{m} \hat{H}^2_{n,m-1}(z_1,z_2) ; \qquad (5.24)$$

3) For the transition between adjacent columns of the recursion, we have (trivially), for $1 \leq n \leq N$:

$$\hat{H}_{n,-M-1}(z_1,z_2) = \hat{H}_{n-1,M}(z_1,z_2) , \qquad (5.25)$$

$$\hat{P}_{n,-M-1} = \hat{P}_{n-1,M} , \qquad (5.26)$$

$$\hat{H}^2_{n,-M-1}(z_1,z_2) = \hat{H}^2_{n-1,M}(z_1,z_2) , \qquad (5.27)$$

$$\hat{H}_{n,-M-1}(z_1,z_2)\hat{H}_{n,-M-1}(1/z_1,1/z_2)$$
$$= \hat{H}_{n-1,M}(z_1,z_2)\hat{H}_{n-1,M}(1/z_1,1/z_2) \quad . \qquad (5.28)$$

The entire algorithm requires approximately $20N^3M^2$ additions and multiplications. All of the computations can be performed "in place" (i.e., separate storage is not required for the "old" and the "new" parameters.)

We now illustrate this algorithm with a simple example.

Example 5.1: We have a power density spectrum,

$$S(z_1,z_2) = (5+z_1^{-1}+z_1+z_2^{-1}+z_2) \quad ,$$

or

$$r(k,\ell) = \begin{cases} 5 \;, & (k,\ell) = (0,0) \;; \\ 1 \;, & (k,\ell) = (1,0),(0,1),(-1,0),(0,-1); \\ 0 \;, & \text{otherwise} \;. \end{cases}$$

We implement our spectral factorization algorithm for N=M=1. The algorithm proceeds as follows:

1) Using (5.15)-(5.18) we have

$$\hat{H}_{0,0}(z_1,z_2) = 1 \quad ,$$

$$\hat{P}_{0,0} = 5 \quad ,$$

$$\hat{H}^2_{0,0}(z_1,z_2) = 1 \quad ,$$

$$\hat{H}_{0,0}(z_1,z_2)\hat{H}_{0,0}(1/z_1,1/z_2) = 1 \quad ;$$

2) Using (5.19)-(5.24), for (n,m)=(0,1), we have

$$\hat{\rho}(0,1) = \frac{1}{5} \quad ,$$

$$\hat{H}_{0,1}(z_1,z_2) = (1 - \frac{1}{5} z_2^{-1}) \quad ,$$

$$\hat{P}_{0,1} = \frac{24}{5} \quad ,$$

$$\hat{H}^2_{0,1}(z_1,z_2) = (1 - \frac{2}{5} z_2^{-1} + \frac{1}{25} z_2^{-2}) \quad ,$$

$$\hat{H}_{0,1}(z_1,z_2)\hat{H}_{0,1}(1/z_1,1/z_2) = (\frac{26}{25} - \frac{1}{5} z_2^{-1} - \frac{1}{5} z_2) \quad ;$$

3) Combining (5.25)-(5.28) for n=1, with (5.19)-(5.24) for (n,m)=(1,-1), we have

$$\hat{\rho}(1,-1) = 0 \quad ,$$

$$\hat{H}_{1,-1}(z_1,z_2) = (1 - \frac{1}{5} z_2^{-1}) \quad ,$$

$$\hat{P}_{1,-1} = \frac{24}{5} \quad ,$$

$$\hat{H}^2_{1,-1}(z_1,z_2) = (1 - \frac{2}{5} z_2^{-1} + \frac{1}{25} z_2^{-2}) \quad ,$$

$$\hat{H}_{1,-1}(z_1,z_2)\hat{H}_{1,-1}(1/z_1,1/z_2) = (\frac{26}{25} - \frac{1}{5} z_2^{-1} - \frac{1}{5} z_2) \quad ;$$

4) Using (5.19)-(5.23) for (n,m)=(1,0), (we do not update the filter autocorrelation in this case, because it is not needed for the remaining computations) we have:

$$\hat{\rho}(1,0) = \frac{5}{24} \quad ,$$

$$\hat{H}_{1,0}(z_1,z_2) = (1 - \frac{1}{5} z_2^{-1} + \frac{1}{24} z_1^{-1} z_2 - \frac{5}{24} z_1^{-1}) \quad ,$$

$$\hat{P}_{1,0} = \frac{551}{120} \quad ,$$

$$\hat{H}^2_{1,0}(z_1,z_2) = [(1 - \frac{2}{5} z_2^{-1} + \frac{1}{25} z_2^{-2})$$

$$- \frac{5}{12} (\frac{26}{25} - \frac{1}{5} z_2^{-1} - \frac{1}{5} z_2) z_1^{-1}$$

$$+ \frac{5}{24}^{2} (1 - \frac{2}{5} z_2 + \frac{1}{25} z_2^2) z_1^{-2}] \quad ;$$

5) Using (5.19)-(5.22) for (n,m)=(1,1), we have:

$$\hat{\rho}(1,1) = \frac{-50}{551} \quad ,$$

$$\hat{H}_{1,1}(z_1,z_2) = [1 - (\frac{7237}{33060}) z_2^{-1} + (\frac{25}{6612}) z_2^{-2} + (\frac{1}{24}) z_1^{-1} z_2$$

$$- (\frac{2995}{13224}) z_1^{-1} + (\frac{50}{551}) z_1^{-1} z_2^{-1}] \quad ,$$

$$\hat{P}_{1,1} = \frac{301101}{66120} \quad .$$

The next two numerical examples were implemented on a computer, using double-precision arithmetic.

Example 5.2: We begin with the spectrum

$$S(z_1,z_2) = (1 + \frac{1}{4} z_1^{-1} + \frac{1}{4} z_2^{-1})(1 + \frac{1}{4} z_1 + \frac{1}{4} z_2) \quad .$$

It can be seen that the spectrum is already factored; the minimum-phase whitening filter is

$$H(z_1,z_2) = (1 + \frac{1}{4} z_1^{-1} + \frac{1}{4} z_2^{-1})^{-1}$$

$$= \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \frac{(k+\ell)!}{k!\ell!} (-1/4)^{(k+\ell)} z_1^{-k} z_2^{-\ell} \quad ,$$

and $P = 1$. The spectral factorization algorithm was implemented for N=M=3. As can be seen from Fig. 5.1, the algorithm produces a parallelogram-shaped minimum-phase filter, $\hat{H}_{3,3}(z_1,z_2)$, 22 points in height and 4 points in width. A portion of the unit sample response of the filter is shown in Fig. 5.5. (The region enclosed by the dotted line indicates the support for the reflection coefficient sequence.) We note that for $|\ell| \leq 3$, the filter coefficients closely match the unit sample response of $(1 + \frac{1}{4} z_1^{-1} + \frac{1}{4} z_2^{-1})^{-1}$. The filter coefficients for $|\ell| > 3$(which we call the "tails" of the filter) decay very rapidly to zero. The mean-square prediction error associated with $\hat{H}_{3,3}(z_1,z_2)$ is

$$\hat{P}_{3,3} = 1.00007$$

(compared with P=1 for the exact whitening filter).

Therefore, we have that

$$\hat{H}_{3,3}(z_1,z_2) \simeq (1 + \frac{1}{4} z_1^{-1} + \frac{1}{4} z_2^{-1})^{-1} \quad .$$

ℓ				
	⋮	⋮	⋮	
	1.326×10^{-4}	-5.883×10^{-5}	1.852×10^{-5}	
	-3.556×10^{-4}	9.725×10^{-5}	-2.327×10^{-5}	
	-1.427×10^{-2}	1.409×10^{-2}	-8.673×10^{-3}	3.814×10^{-3}
	6.207×10^{-2}	-4.635×10^{-2}	2.284×10^{-2}	-8.580×10^{-3}
	-2.499×10^{-1}	1.247×10^{-1}	-4.634×10^{-2}	1.410×10^{-2}
	1.0	-2.498×10^{-1}	6.200×10^{-2}	-1.426×10^{-2}
k →				
		-2.534×10^{-4}	3.044×10^{-4}	-3.716×10^{-4}
		4.312×10^{-4}	-2.733×10^{-4}	1.406×10^{-4}
		-7.555×10^{-4}	3.261×10^{-4}	-9.356×10^{-5}
		9.292×10^{-4}	-2.886×10^{-4}	6.107×10^{-5}
		3.766×10^{-7}	1.999×10^{-6}	-5.754×10^{-7}
		⋮	⋮	⋮

<u>Fig. 5.5</u> Unit sample response of filter $\hat{H}_{3,3}(z_1,z_2) \simeq (1 + \frac{1}{4} z_1^{-1} + \frac{1}{4} z_2^{-1})^{-1}$.

If another spectrum is formed, $\hat{H}_{3,3}(z_1,z_2)\hat{H}_{3,3}(1/z_1,1/z_2)$, and if this new spectrum is approximately factored, we should obtain a filter, denoted $\hat{\hat{H}}_{3,3}(z_1,z_2)$, approximately equal to $(1 + \frac{1}{4} z_1^{-1} + \frac{1}{4} z_2^{-1})$. This spectrum was factored by means of our algorithm for N=M=3, to produce a filter $\hat{\hat{H}}_{3,3}(z_1,z_2)$. The filter is illustrated in Fig. 5.6. It can be seen that the filter closely matches $(1 + \frac{1}{4} z_1^{-1} + \frac{1}{4} z_2^{-1})$.

In the previous example, the spectrum to be factored was very smooth, and under these nearly optimum conditions, our spectral factorization algorithm performed satisfactorily. The following example demonstrates some serious difficulties associated with this sequential method of choosing the reflection coefficients.

<u>Example 5.3</u>: We wish to design a recursive 2-D fan filter. The desired magnitude-squared frequency response is

$$S(e^{j\omega_1},e^{j\omega_2}) = \begin{cases} 1 \ , & |\omega_2|<|\omega_1| \\ .02 \ , & \text{otherwise} \end{cases} .$$

The desired frequency response is illustrated in Fig. 5.7. (Fan filters are used in array processing to discriminate against signals arriving from certain directions.) The autocorrelation function is

⋮	⋮	⋮	
-1.491×10^{-5}	5.221×10^{-7}	-1.070×10^{-6}	
-2.051×10^{-4}	-8.409×10^{-7}	1.892×10^{-6}	
-2.382×10^{-3}	1.239×10^{-3}	-5.579×10^{-4}	2.260×10^{-4}
3.777×10^{-4}	-2.137×10^{-4}	1.575×10^{-4}	-5.632×10^{-4}
2.499×10^{-1}	1.097×10^{-4}	-2.098×10^{-4}	1.264×10^{-3}
1.0	2.499×10^{-1}	3.003×10^{-4}	-1.900×10^{-3}
	8.492×10^{-5}	-1.151×10^{-4}	3.762×10^{-4}
	-1.049×10^{-4}	5.369×10^{-6}	-5.452×10^{-5}
	-2.294×10^{-5}	8.442×10^{-5}	-2.255×10^{-5}
	-8.405×10^{-4}	8.405×10^{-6}	8.880×10^{-6}
	-5.408×10^{-5}	-7.817×10^{-7}	5.410×10^{-7}
	⋮	⋮	⋮

(vertical axis ℓ, horizontal axis k)

Fig. 5.6 Unit sample response of filter $\hat{\hat{H}}_{3,3}(z_1,z_2) \simeq (1 + \frac{1}{4} z_1^{-1} + \frac{1}{4} z_2^{-1})$.

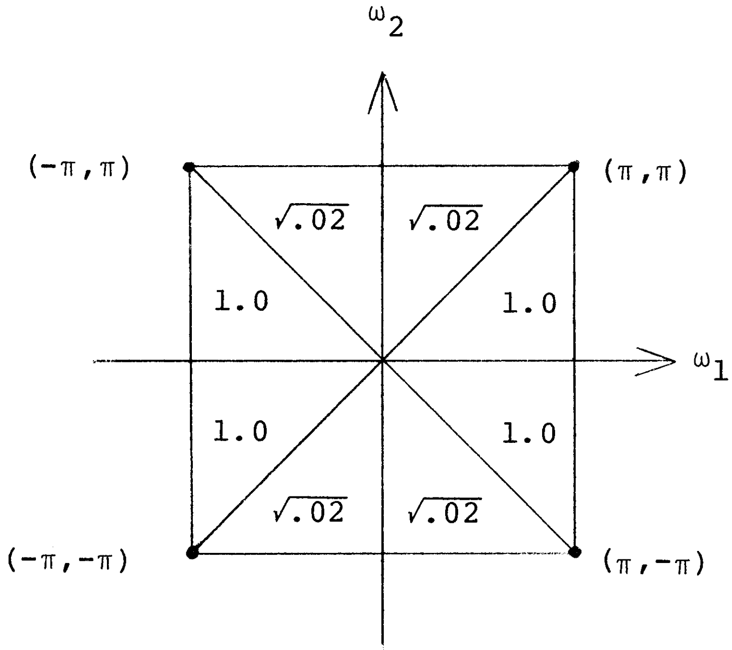


Fig. 5.7 The desired magnitude frequency response of the 2-D recursive fan filter.

$$
r(k,\ell) = \begin{cases} .51 , & (k,\ell)=(0,0) \\ \dfrac{1.96}{\pi^2(\ell^2-k^2)} , & (k+\ell) \text{ odd} \\ 0 , & \text{otherwise} \end{cases}
$$

Given the slow rate of decay of the autocorrelation function, we anticipate that this is a difficult spectrum to factor. In fact, since the spectrum is discontinuous, no sequence of approximate whitening filters can converge uniformly to a limit whitening filter; a Gibbs-type phenomenon occurs in the neighborhood of the discontinuities of the spectrum.

Our spectral factorization algorithm was implemented for N=M=4. A projection plot of the frequency response of the recursive filter is shown in Fig. 5.8. A contour plot of the frequency response is shown in Fig. 5.9. It can be seen that there are very large ripples in the transition region and in the passband. For most purposes, this could be an unacceptable design.

Another problem with this design is that the tails of the filter decay very slowly; only for $|\ell|>15$ are the magnitudes of the tails less than 10^{-3}.

The rather poor performance of the spectral factorization algorithm in the above example is to be expected, since the conditions under which the algorithm would yield optimal values for the reflection coefficients are not met.

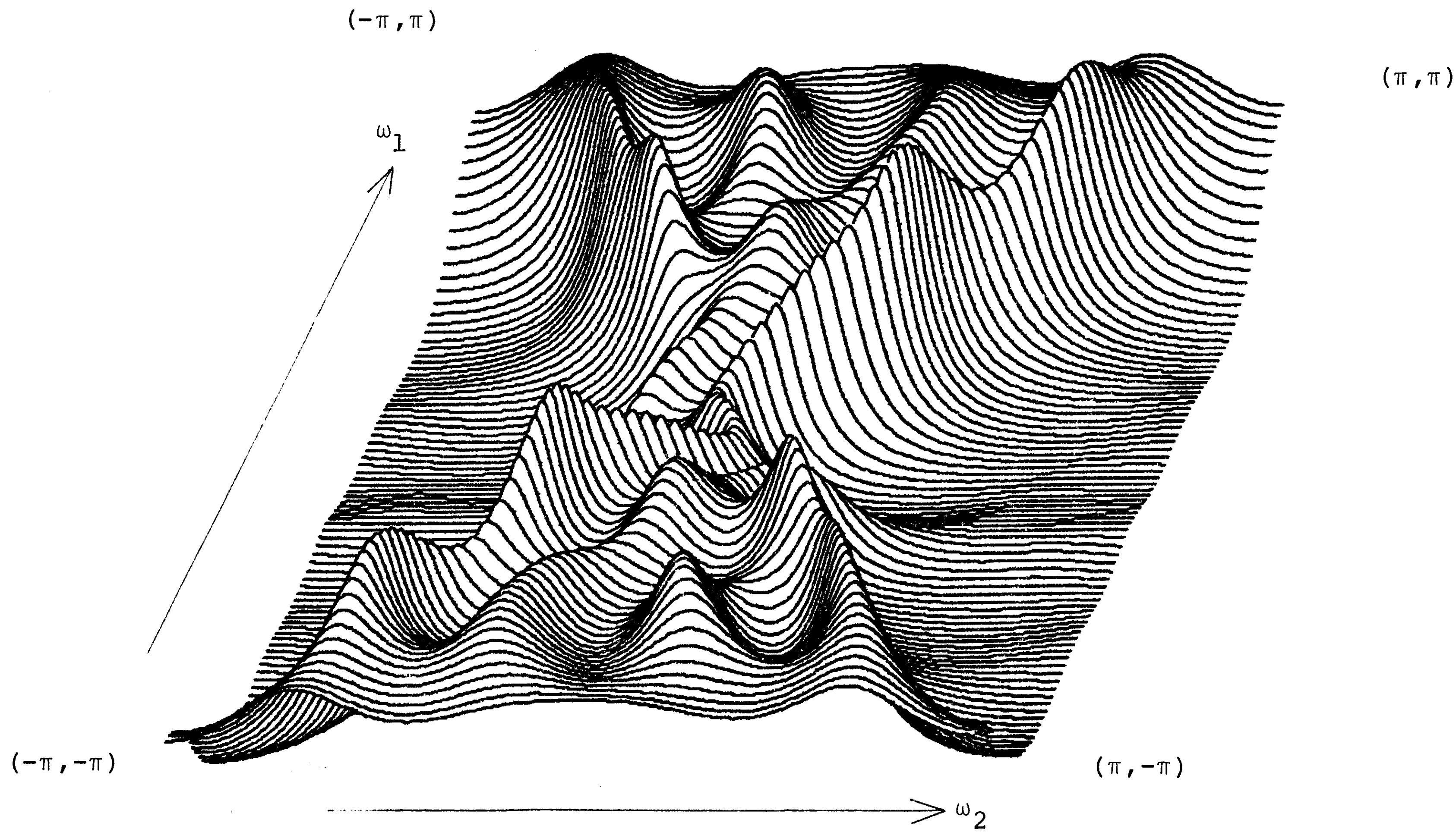


Fig. 5.8 Projection plot of magnitude frequency response of 2-D recursive fan filter.

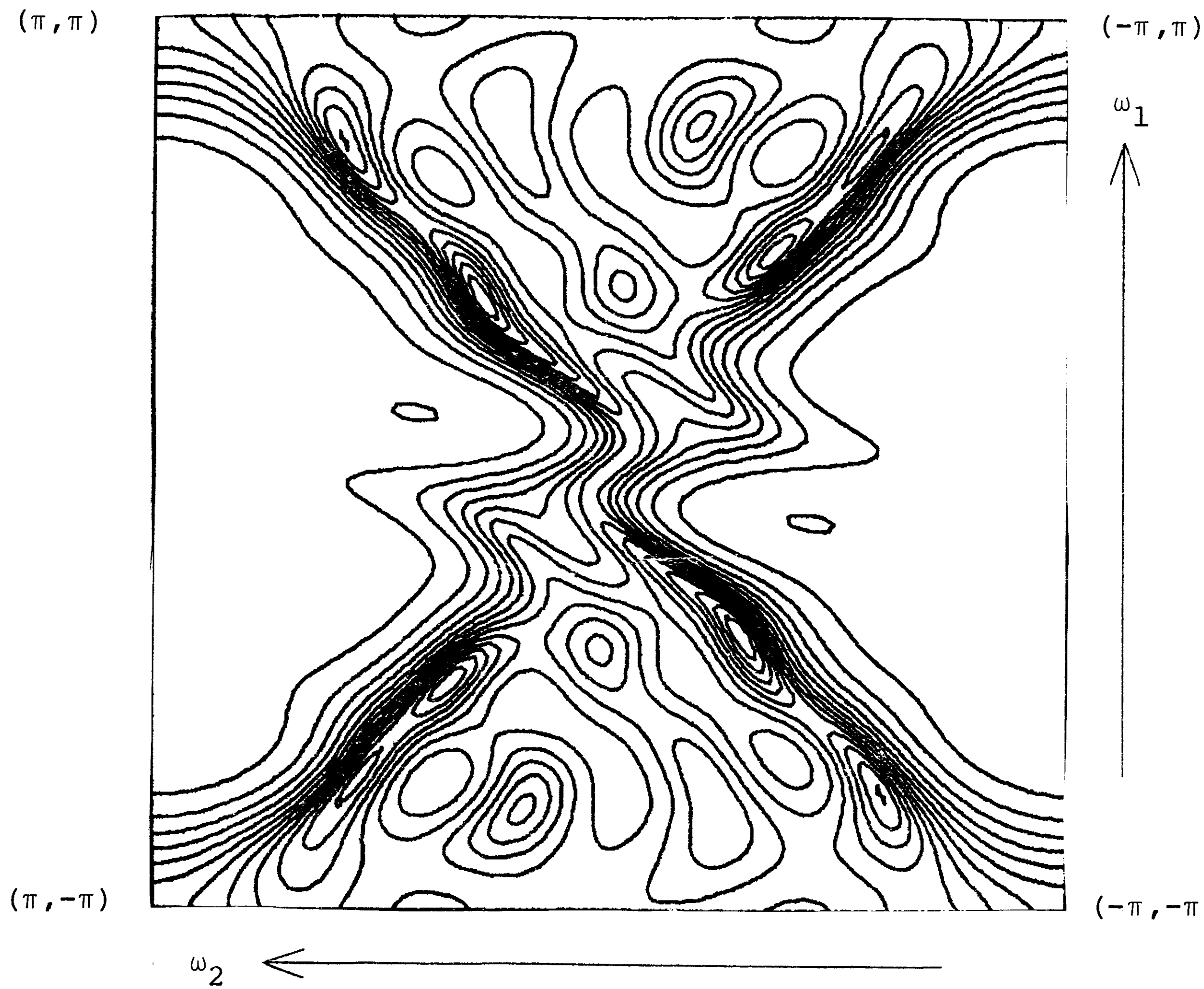


Fig. 5.9 Contour plot of frequency response of 2-D recursive fan filter.

One possible approach to improving the design would be to window the autocorrelation function (equivalently, to smooth the spectrum) prior to applying the spectral factorization algorithm. Given a relatively smooth spectrum instead of the original discontinuous spectrum, the algorithm would probably yield more optimal values for the reflection coefficients.

It is apparent that, in general, the full potential of the 2-D reflection coefficient representation can only be realized by the development of an algorithm that would simultaneously choose the reflection coefficients to maximize some index of performance.

5.3 A 2-D Autoregressive Model Fitting Algorithm

We are given a finite set of samples from a 2-D random process, and the object is to estimate the minimum-phase whitening filter by modeling it as an FIR minimum-phase filter, $\hat{H}_{N,M}(z_1,z_2)$. Our approach to this problem is to represent $\hat{H}_{N,M}(z_1,z_2)$ in terms of a finite number of reflection coefficients, $\{\hat{\rho}(n,m);\ (n=0, 1\le m\le M), (1\le n\le N, -M\le m\le M)\}$, and to choose the reflection coefficients to obtain a good fit between the whitening filter model and the data.

One approach to choosing the reflection coefficients is analogous to the 1-D autocorrelation method: the available data is used to estimate the 2-D autocorrelation function to a finite lag, and then the estimated autocorrelation function is used in the spectral factorization algorithm of the previous section to compute the whitening filter estimate, $\hat{H}_{N,M}(z_1,z_2)$.

Another approach to choosing the reflection coefficients is analogous to the Burg algorithm. Instead of first forming an autocorrelation estimate, the filter is obtained directly from the data. The reflection coefficients are chosen sequentially, each new reflection coefficient, $\hat{\rho}(n,m)$, being chosen to achieve the best fit between the data and the new filter, $\hat{H}_{n,m}(z_1,z_2)$.

At the beginning of the $(n,m)^{th}$ stage of the algorithm, for $\{n=0, 1\le m\le M\}$, or $\{n=N, -M\le m\le M\}$, we have $\hat{H}_{n,m-1}(z_1,z_2)$. The new reflection coefficient, $\hat{\rho}(n,m)$ is chosen to minimize the sum of the squares of the new forward and backward prediction errors:

$$\sum_{(k,\ell)}\sum \{[\varepsilon^{(+)}(n,m;k,\ell)]^2 + [\varepsilon^{(-)}(n,m;k-n,\ell-m)]^2\} \quad , \qquad (5.29)$$

where $\varepsilon^{(+)}(n,m;k,\ell)$ is a forward prediction error,

$$\varepsilon^{(+)}(n,m;k,\ell) = [x(k,\ell) - \sum_{(s,t)}\sum \hat{h}(n,m;s,t)x(k-s,\ell-t)] \quad , \qquad (5.30)$$

and $\varepsilon^{(-)}(n,m;k,\ell)$ is a backward prediction error,

$$\varepsilon^{(-)}(n,m;k,\ell) = [x(k,\ell) - \sum_{(s,t)}\sum \hat{h}(n,m;s,t)x(k+s,\ell+t)] \quad . \tag{5.31}$$

The indices of the summation in (5.29) depend on both the extent of the data, and on the extent of the filter. In general, (k,ℓ) should cover as many points as possible without running the filter off the data anywhere, since that would tend to prejudice the estimate for the reflection coefficient. This can be shown to imply that the support for the data should be at least as great as the support for the final filter, $\hat{H}_{N,M}(z_1,z_2)$. However, in some cases, the tails of the filter may be so insignificant that they can be run off the edge of the data without adversely affecting the reflection coefficient estimate.

Using (5.21) it can be shown that

$$\varepsilon^{(+)}(n,m;k,\ell) = [\varepsilon^{(+)}(n,m-1;k,\ell) - \hat{\rho}(n,m)\varepsilon^{(-)}(n,m-1;k-n,\ell-m)], \tag{5.32}$$

and

$$\varepsilon^{(-)}(n,m;k,\ell) = [\varepsilon^{(-)}(n,m-1;k,\ell) - \hat{\rho}(n,m)\varepsilon^{(+)}(n,m-1;k+n,\ell+m)]. \tag{5.33}$$

Substituting (5.32) and (5.33) into (5.29) we want to choose $\hat{\rho}(n,m)$ to minimize the following expression:

$$\sum_{(k,\ell)}\sum \{[1+\hat{\rho}^2(n,m)]([\varepsilon^{(+)}(n,m-1;k,\ell)]^2+[\varepsilon^{(-)}(n,m-1;k-n,\ell-m)]^2)$$

$$- 4\hat{\rho}(n,m)[\varepsilon^{(+)}(n,m-1;k,\ell)\varepsilon^{(-)}(n,m-1;k-n,\ell-m)]\} \quad . \tag{5.34}$$

Taking the derivative of (5.34) with respect to $\hat{\rho}(n,m)$, setting the derivative equal to zero, and solving for $\hat{\rho}(n,m)$, we have

$$\hat{\rho}(n,m) = \frac{2 \sum_{(k,\ell)}\sum \{[\varepsilon^{(+)}(n,m-1;k,\ell)][\varepsilon^{(-)}(n,m-1;k-n,\ell-m)]\}}{\sum_{(k,\ell)}\sum [\varepsilon^{(+)}(n,m-1;k,\ell)]^2 + [\varepsilon^{(-)}(n,m-1;k-n,\ell-m)]^2\}} \quad . \tag{5.35}$$

Using Schwartz's inequality, it can be shown that the magnitude of the reflection coefficient is less than one. The forward and backward prediction errors do not have to be directly computed; instead, they can be recursively updated at each stage of the algorithm. The complete algorithm is as follows:

1) $$\hat{H}_{0,0}(z_1,z_2) = 1 \quad , \tag{5.36}$$

$$\hat{P}_{0,0} = \text{const} \sum_{(k,\ell)}\sum x^2(k,\ell) \quad , \tag{5.37}$$

$$\varepsilon^{(+)}(0,0;k,\ell) = x(k,\ell) \quad , \tag{5.38}$$

$$\varepsilon^{(-)}(0,0;k,\ell) = x(k,\ell) \quad ; \tag{5.39}$$

2) At the beginning of the $(n,m)^{th}$ stage of the algorithm, for $\{n=0, 1\leq m\leq M\}$ or $\{1\leq n\leq N, -M\leq m\leq M\}$, we have

$$\hat{H}_{n,m-1}(z_1,z_2) \quad , \tag{5.40}$$

$$\hat{P}_{n,m-1} \quad , \tag{5.41}$$

$$\varepsilon^{(+)}(n,m-1;k,\ell) \quad , \tag{5.42}$$

$$\varepsilon^{(-)}(n,m-1;k,\ell) \quad ; \tag{5.43}$$

We first compute the reflection coefficient estimate:

$$\hat{\rho}(n,m) = \frac{2 \sum\sum_{(k,\ell)} \{[\varepsilon^{(+)}(n,m-1;k,\ell)][\varepsilon^{(-)}(n,m-1;k-n,\ell-m)]\}}{\sum\sum_{(k,\ell)} \{[\varepsilon^{(+)}(n,m-1;k,\ell)]^2 + [\varepsilon^{(-)}(n,m-1;k-n,\ell-m]^2\}} \quad ; \tag{5.44}$$

we then perform the following updates:

$$\hat{H}_{n,m}(z_1,z_2) = \hat{H}_{n,m-1}(z_1,z_2) - \hat{\rho}(n,m) z_1^{-n} z_2^{-m} \hat{H}_{n,m-1}(1/z_1,1/z_2) \quad , \tag{5.45}$$

$$\hat{P}_{n,m} = \hat{P}_{n,m-1}[1-\hat{\rho}^2(n,m)] \quad , \tag{5.46}$$

$$\varepsilon^{(+)}(n,m;k,\ell) = \varepsilon^{(+)}(n,m-1;k,\ell) - \hat{\rho}(n,m)\varepsilon^{(-)}(n,m-1;k-n,\ell-m) \quad , \tag{5.47}$$

$$\varepsilon^{(-)}(n,m;k,\ell) = \varepsilon^{(-)}(n,m-1;k,\ell) - \hat{\rho}(n,m)\varepsilon^{(+)}(n,m-1;k+n,\ell+m) \quad ; \tag{5.48}$$

3) For the transition between adjacent columns of the recursion, for $1 \leq n \leq N$, we have

$$\hat{H}_{n,-(M+1)}(z_1,z_2) = \hat{H}_{n-1,M}(z_1,z_2) \quad , \tag{5.49}$$

$$\hat{P}_{n,-(M+1)} = \hat{P}_{n-1,M} \quad , \tag{5.50}$$

$$\varepsilon^{(+)}(n,-M-1;k,\ell) = \varepsilon^{(+)}(n-1,M;k,\ell) \quad , \tag{5.51}$$

$$\varepsilon^{(-)}(n,-M-1;k,\ell) = \varepsilon^{(-)}(n-1,M;k,\ell) \quad . \tag{5.52}$$

The expression for the reflection coefficient in this 2-D Burg algorithm, (5.44), is very similar to the expression for the reflection coefficient in the spectral factorization algorithm of the previous section, (5.10). If the extent of the data is much greater than the extent of the filter, $\hat{H}_{n,m}(z_1,z_2)$, then we expect the two expressions to give nearly the same values for the reflection coefficient.

Although our 2-D Burg algorithm has not been implemented for any examples, we can anticipate some of the difficulties that would be encountered in using it. As in the case of our spectral factorization algorithm, the sequential choosing of the reflection coefficients generally is suboptimal, and the extent of the tails of the filters may be unacceptable.

Once again, the only way to take full advantage of the reflection coefficient representation would be to develop an algorithm for simultaneously optimizing the reflection coefficients. It is interesting to note that a 1-D algorithm of this type has been proposed as an alternative to the 1-D Burg algorithm [29].

CHAPTER 6

CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

It has been shown that by adopting a particular notion of 2-D causality, virtually all of the major results from 1-D linear prediction theory can be extended to the 2-D case. Having obtained these results, we can claim to understand the theoretical aspects of 2-D linear prediction.

From a practical point of view, the most important result in this thesis is the reflection coefficient representation for 2-D minimum-phase filters. The significance of this representation is that by designing 2-D filters in the reflection coefficient domain, the minimum-phase constraint is made an integral part of the design procedure. Future research efforts need to be directed towards the development of effective algorithms for choosing the reflection coefficients. The sequential least-squares approaches to choosing the reflection coefficients, that were discussed in this thesis, may be useful in some cases, but they do not generally realize the full potential of the reflection coefficient representation.

Regarding theoretical extensions of the new results in this thesis, it should be possible to find similar results for

1) complex-valued 2-D random processes (the equations should be the same except for complex-conjugate symbols at various places);

2) higher-dimensional random processes; and

3) vector valued multi-dimensional random processes.

REFERENCES

1. Wiener, N., The Extrapolation, Interpolation, and Smoothing of Stationary Time Series, Wiley, New York: 1949.

2. Robinson, E., Statistical Communication and Detection with special reference to digital data processing of radar and seismic signals, Hafner, New York: 1967.

3. Burg, J., "Maximum Entropy Spectral Analysis," Ph.D. dissertation, Dept. of Geophysics, Stanford Univ., 1975.

4. Makhoul, J., "Linear Prediction: A Tutorial Review," IEEE Proceedings, v. 63, no. 4, pp. 561-580 (April, 1975).

5. Lindquist, A., "A New Algorithm for Optimal Filtering of Discrete-time Stationary Processes," SIAM J. Control, v. 12, no. 4, pp. 736-746 (November, 1974).

6. Claerbout, J., "Synthesis of a Layered Medium from Its Acoustic Transmission Response," Geophysics, v. 33, no. 2, pp. 264-269 (April, 1968).

7. Pusey, L, "An Innovations Approach to Spectral Estimation and Wave Propagation," Sc.D. Thesis, Dept. of Electrical Engineering, M.I.T. (June, 1975).

8. Whittle, P., "On Stationary Processes in the Plane," Biometrika, v. 41, pp. 434-449 (December, 1954).

9. Woods, J., "Two-Dimensional Discrete Markovian Fields," IEEE Trans. Inf. Th., v. IT-18, no. 2, pp. 232-240 (March, 1972).

10. Shanks, J., Treitel, S., and Justice, J., "Stability and Synthesis of Two-Dimensional Recursive Filters," IEEE Trans. Aud. Electr., v. AU-20, no. 2, pp. 115-128 (June, 1972).

11. Genin, Y., and Kamp, Y., "Counterexample in the Least Square Inverse Stabilization of 2D-Recursive Filters," Electronics Letters, v. 11, pp. 330-331 (1975).

12. Pistor, P., "Stability Criterion for Recursive Filters," IBM J. Res. Devel., v. 18, no. 1, pp. 59-71 (Jan., 1974).

13. Dudgeon, D., "Two-Dimensional Recursive Filtering," Sc.D. Dissertation, Dept. of Electrical Engineering, M.I.T. (May, 1974).

14. Ekstrom, M., and Woods, J., "Two-Dimensional Spectral Factorization with Applications in Recursive Digital Filtering," IEEE Trans. Acoust., Speech, Signal Processing, v. ASSP-24, no. 2, pp. 115-127 (April, 1976).

15. Levinson, N., "The Wiener RMS Error in Filter Design and Prediction," Appendix B in N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series, Wiley, New York: 1949.

16. Van Trees, H., Detection, Estimation, and Modulation Theory, Part I, Wiley, New York: 1968.

17. Robinson, E., and Wold., H., "Minimum-delay Structure of Least-Squares/Eo-ipso Predicting Systems for Stationary Stochastic Processes," Proceedings of the Symposium on Time Series Analysis, edited by M. Rosenblatt, Wiley, New York: 1962.

18. Markel, J., and Gray, H., "On Autocorrelation Equations as Applied to Speech Analysis," IEEE Transactions on Audio and Electracoustics, v. Au-21, no. 2, pp. 60-79 (April, 1973).

19. Dubroff, R., "The Effective Autocorrelation Function of Maximum Entropy Spectra," Proc. IEEE, pp. 1622-1623 (Nov., 1975).

20. Durbin, J., "The Fitting of Time-series Models," Rev. Int. Inst. Stat. 28, pp. 233-244 (1960).

21. Grenander, U., and Szego, G., Toeplitz Forms and Their Applications, Univ. of California Press, Berkeley: 1958.

22. Claerbout, J., Fundamentals of Geophysical Data Processing with Applications to Petroleum Prospecting, McGraw Hill: 1976.

23. Oppenheim, A., and Schafer, R., Digital Signal Processing, Prentice-Hall: 1975.

24. Huang, T., "Stability of Two-Dimensional Recursive Filters," IEEE Trans. Aud. Electr., v. AU-20, no. 2, pp. 158-163 (June, 1972).

25. Anderson, B., and Jury, E., "Stability of Multidimensional Digital Filters," IEEE Trans. Circ. and Sys., v. CAS-21, no. 2, pp. 300-304 (March, 1974).

26. Maria, G., and Fahmy, M., "On the Stability of Two-Dimensional Digital Filters," IEEE Trans. Audio Electroa. (Corresp.) v. AU-21 (Oct., 1973).

27. Marzetta, T., "Stability Tests for Non-symmetric Half-plane Recursive Filters," to appear.

28. Marzetta, T., "The Behavior of the PEF $H_{N,M}(z_1,z_2)$ for Large Values of N", Internal Memo, Digital Filtering Group, M.I.T., R.L.E., Feb. 15, 1978.

29. Fougere, P., "A Solution to the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis," J. of Geophysical Research, v. 82, no. 7 (March, 1977).