AUTOREGRESSIVE SPECTRAL ESTIMATION

IN NOISE WITH APPLICATION

TO SPEECH ANALYSIS

By

ROBERT DAVID PREUSS

Bachelor of Arts
Boston University
Boston, Massachusetts
1975

Master of Science
Boston University
Boston, Massachusetts
1977

AUTOREGRESSIVE SPECTRAL ESTIMATION

IN NOISE WITH APPLICATION

TO SPEECH ANALYSIS

Thesis Approved:

Thesis Adviser

James R. Rowland

C. M. Bacon

Marvin S. Keener

Norman N. Durham
Dean of the Graduate College

1187120

To all those whose love and emotional support

has sustained me these past

thirty years.

Especially and in particular, to my mother

Phyllis Burnap Brooks.

ACKNOWLEDGMENT

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Nearly all sciences are concerned with the analysis of measurement data. The following chapters will present a new tool for the analysis of time series measurements; in particular, a new method of spectral estimation is presented. Many spectral estimation methods already exist and, increasingly, new methods continue to be developed; therefore, it is appropriate to reflect, briefly, upon the reasons for such continued activity in an area already so well researched.

A synergism exists between advances in computer technology and advances in practical methods of time series analysis. As more effective (and complex) methods of time series analysis are developed, the demands for smaller, cheaper, and faster digital circuitry (capable of implementing these methods within the size/cost/power constraints of various applications) are increased. As smaller, cheaper, faster and more reliable digital circuitry becomes available, more complex (and effective) methods of time series analysis become practical. Fundamentally, however, it is the demand for improved solutions to engineering problems that motivates the desire for more effective methods of time series analysis.

1

Motivation

Most information we have about the world around us is received indirectly through time series measurements. In the case of vision, one determines the shape (and other characteristics) of an object by reception (measurement) of light waves scattered by the object. In the case of speech, one determines the intended message of the speaker by reception (measurement) of acoustic pressure waves. Prospecting, manufacturing, astronomy, medicine, and economics are but a few of the areas that can benefit from improved methods of time series analysis.

Spectral estimation is one of the most important areas of time series analysis. In many cases, knowledge of the time series spectrum is adequate to answer all important questions regarding the system producing the time series; in the case of a stable time-invariant linear input-output system, knowledge of the output process spectrum (together with the statistics of the stationary input process) will completely characterize the system.

Noise corruption is among the fundamental problems of time series analysis. All useful analysis techniques for measurement data are at least mildly tolerant of noise since there always exists a small probability of measurement error; some techniques are specifically designed to account for knowledge of the noise statistics in the analysis of noise-corrupted measurement data. Regardless of the analysis technique, the fundamental performance limits are always

reduced by the presence of noise.[1] Consequently, it is always advisable to minimize noise corruption as much as is practical; still, practical constraints imposed by some situations do not permit the reduction of noise corruption to insignificant levels so that sophisticated analysis techniques are required to achieve the best possible performance.

Spectral estimation is of fundamental importance to the various applications of speech analysis and practical constraints imposed by many of these applications do not permit the reduction of noise corruption to insignificant levels. Examples of such applications include low data rate digital voice communications systems and speech recognition/understanding systems among others; often the cost and/or inconvenience of shielding from environmental noise makes significant acoustic noise corruption inevitable.

Autoregressive (AR) spectral models have been successful for various systems involving speech analysis; moreover, numerous speech synthesis systems based upon the AR model have become commercially available in recent years. Because the currently available practical methods for AR parameter estimation yield poor results in common noise

---

[1]In some specialized circumstances the performance limits are unchanged by the presence of noise. Even when this is the case, the complexity of the analysis methods required to achieve these limits is usually increased by the noise presence.

environments but are effective in sufficiently quiet environments, it is reasonable to retain the AR model for the speech process while attempting to develop improved methods for estimating the AR parameters.

The fundamental limit to the performance of any estimation procedure depends upon the available information. In theory, even the most obscure (but not unrelated) additional information may be used to improve a parameter estimate; of course, one should rely first upon information that is both easily available and expected to provide substantial improvement.

Most recent efforts to overcome the poor performance of classical AR estimators in noise, including the present one, have attempted to employ information regarding the noise statistics in addition to the noise corrupted time series observations. This information is often provided simply by deploying additional sensors intended to measure the noise directly; other speech analysis systems employ prior segments of the primary observation signal that are thought to be free from speech activity to predict the current relevant noise statistics.

The present work does not address the problem of obtaining accurate noise statistics. Assuming appropriate noise statistics to be available, the following chapters develop a new and improved method of estimating the AR signal parameters from noise corrupted time series observations.

As might be expected, the method entails increased computational cost over less effective techniques; it is expected that performance requirements of speech analysis (and other) applications - as well as cost reductions that are continually provided by advances in computer technology- shall, in many cases, make the advantages of this method appear relatively inexpensive.

Overview

Chapter II provides a general discussion of the various issues and techniques of spectral estimation; particular attention is given to the problems of AR spectral estimation. In addition, this discussion introduces basic formulae and provides an historical perspective for the subsequent chapters.

Chapter III presents the theoretical foundations of the new (weighted information) estimation procedure. After some additional motivational discussion, the method is formulated as an approximation to an ideal (but intractable) formulation and a generalization of a commonly employed (noise filtering) estimation procedure. In addition to the general formulation, significant contributions of this chapter include the analogy leading to Equation (3.20) and the properties developed in the fifth section.

Chapter IV discusses a variety of computational methods relevant to AR estimation based upon the weighted information formulation. The author considers the area of

computational procedures as requiring the greatest attention for further extension and refinement of this work. Only the formulae for vector quantization, in particular Equations (4.58a) and (4.81), appear ready for detailed cost/performance analyses.

Chapter V demonstrates clearly that the weighted information formulation leads to reduced estimation error as compared to the more common noise filtering formulation. Examples from both simulated and real speech are provided. The demonstration relies upon the reader's visual assessment of scatter plots; thus it is somewhat qualitative. A more quantitative assessment (e.g. a comparison of empirical variance to theoretical performance bounds) would be interesting; however, one would still have difficulty evaluating the significance of a reduction in empirical variance to the performance of a particular system. Without a full implementation one must rely upon experience and judgement as well as the available experimental evidence.

Finally, Chapter VI summarizes the results of this effort and provides suggestions as to how this work may be effectively extended and refined.

CHAPTER II

GENERAL DISCUSSION

Spectral estimation is a problem of statistical inference with a long history due to its pervasive importance in scientific applications [1]. Modern empirical spectral analysis began to take shape as an organized discipline with the introduction in 1893 of the periodogram by Schuster [2].

Given N observations $\{x_n; n=0,1,\ldots,N-1\}$ of a time series at unit time intervals the periodogram, $f(\theta)$, is defined as

$$f(\theta) = X_N(e^{i\theta}) \, X_N(e^{-i\theta})/N \tag{2.1}$$

where

$$X_N(z) = \sum_{n=0}^{N-1} x_n \, z^{-n} \; ; \; z = e^{i\theta} \tag{2.2}$$

Still in use today, the periodogram was practically the sole computational tool of empirical spectral analysis until Yule introduced in 1927 his method of autoregressive (AR) spectral analysis [3].

An $AR(P)$, or P^{th} order autoregressive, model spectrum, $g(\theta)$, is characterized by a model gain, σ, and a monic P^{th}

7

order polynomial, $z^P A_P(z)$, and is defined by them as

$$g(\theta) = \sigma^2 / \left| A_P(e^{i\theta}) \right|^2 \tag{2.3}$$

The polynomial may be characterized by a variety of parameter sets. One parameter set, known as predictor coefficients $\{a_n; n=1,2,\ldots,P\}$, defines the polynomial according to

$$A_P(z) = \sum_{n=0}^{P} a_n z^{-n} \; ; \; a_0 = 1 \tag{2.4}$$

In contrast to Schuster's nonparametric method of spectral analysis, Yule's parametric method first introduces the above mathematical model, justified by physical arguments, and then uses the available data to estimate the model parameters. These estimates are provided by the solution to the Yule-Walker [4] equations

$$\sum_{m=0}^{P} r_{|n-m|} a_m = \sigma^2 \delta_n \; ; \; n=0,1,\ldots,P \tag{2.5}$$

where

$$r_n = \sum_{m=0}^{N-n-1} x_m x_{m+n}/N \; ; \; n=0,1,\ldots,P \tag{2.6}$$

are the biased sample autocorrelation lag estimates.

Model Selection

A variety of other parametric spectral models have been introduced and studied during the past half century; several of them are worth noting. The moving-average (MA) model, like the AR model, is characterized by a polynomial but differs in that the polynomial appears in the numerator; the Schuster periodogram may be viewed as an MA model spectrum.[1] Similarly, ARMA models are described by both numerator and denominator polynomials; these spectra are of particular importance in engineering applications since they characterize all stable linear systems with a finite dimensional state vector. The Blackman-Tukey [5] model spectrum consists of a finite sum of cosine terms; it is obtained by Fourier [6] transformation of the product of the autocorrelation sequence and a finite support window. The Pisarenko [7] model consists of a constant plus a finite number of delta functions. Various combinations of these models are also occasionally employed.

Most often a new model is introduced (together with a procedure for estimating its parameters) simply because it seems reasonable relative to the phenomenon being studied and due to deficiencies in the currently popular

---

[1]Facts such as these tend to blur the distinction between parametric and nonparametric methods. Since any estimate can be described as a member of some parametric family once it has been derived, the distinction may be seen as one of spirit rather than substance.

models.[2, 3] More recently the various results of this "unscientific" approach have been "justified" theoretically; this justification usually takes the form of a principle that should be employed as a guide when the requirement of consistency with the available information leaves several alternatives. The principle is usually embodied in the form of a functional whose extreme value is to be found while the information is provided in the form of constraint equations (or inequalities) for this variational problem.

Much of the current literature is devoted to the "principle of maximum entropy" which was enunciated by Jaynes [8, 9]. If the process is zero-mean stationary and Gaussian[4] it is completely characterized by its power spectral density function, $g(\theta)$, (or "spectrum" for short) and the process entropy is expressed in terms of it by

$$Q = \int_{-\pi}^{\pi} \ln g(\theta) \, d\theta/2\pi \qquad (2.7)$$

---

[2]We shall adopt this pragmatic view later when modeling speech in an acoustically noisy environment.

[3]Sometimes a model is used in spite of its less reasonable form simply because the available parameter estimation methods yield more successful overall results. Thus AR models are employed (instead of the Pisarenko model) to estimate the frequencies of pure sinusoids in white noise from short data records.

[4]The Gaussian assumption may be avoided in the case of correlation constraints. Working directly with probability densities the Gaussian form may be derived as that which maximizes the entropy [10, p. 944].

As demonstrated by Burg [11], if the entropy is subsequently maximized subject to correlation constraints[5]

$$r_n = \int_{-\pi}^{\pi} g(\theta) \, e^{in\theta} \, d\theta/2\pi \; ; \; n=0,1,\ldots,P \qquad (2.8)$$

one may derive the AR(P) form for $g(\theta)$ as given by Equation (2.3). The AR(P) form together with the constraint Equations (2.8) are then sufficient to yield the Yule-Walker Equations (2.5) from which the model parameters may be determined. If cepstral constraints[6] are employed in place of correlation constraints the spectrum maximizing Equation (2.7) has an MA form while both correlation and cepstral constraints lead to an ARMA model. The Pisarenko model is "justified" by deriving it as the minimum energy solution under correlation constraints[7], excepting the energy ($n = 0$) constraint [12].

Another principle discussed in the recent literature is the "principle of minimum cross-entropy" [13]. Introduced by Kullback (under the name "directed divergence") as an

---

[5]The values on the left-hand side are given in terms of the data; for example, by Equation (2.6).

[6]These place constraints directly on the "cepstrum" (or log power spectrum) and are expressed by Equations (2.8) if $g(\theta)$ is replaced by its logarithm while the left-hand side values are expressed in terms of the data.

[7]It may also be related to the maximum entropy principle by noting that the AR(P) model approaches the Pisarenko model as r_0 is decreased to the point where the correlation matrix becomes singular [7, p. 355].

information measure [14], it has a number of interesting properties neatly collected in [15]. In terms of probability densities the cross-entropy is given by

$$S(q,p) = \int q(\bar{x}) \ln[q(\bar{x})/p(\bar{x})] \, d\bar{x} \qquad (2.9)$$

and measures the expected information for discrimination[8] per observation from $q(\bar{x})$ [14]. A symmetric version of this measure, $S(q,p) + S(p,q)$, was introduced earlier by Jeffreys [16] who emphasized the invariance of this measure with respect to coordinate transformations; unlike entropy, cross-entropy shares this important property.

As an inference procedure, minimum cross-entropy analysis requires a prior estimate of the density, $p(\bar{x})$, as well as new information in the form of constraints and derives a new posterior estimate of the density, $q(\bar{x})$, by minimizing $S(q,p)$ subject to the constraints [17]. In the case that the prior density is uniform the procedure is equivalent to maximum entropy; with correlation constraints the posterior density is found to be Gaussian AR(P) with parameters satisfying the Yule-Walker Equations (2.5).

---

[8]Fully, $S(q,p)$ is said to measure the expected information for discrimination in favor of the (correct) hypothesis that the density is $q(\bar{x})$ and against the (competing) hypothesis that the density is $p(\bar{x})$ per observation from $q(\bar{x})$.

Parameter Estimation

The foregoing discussion leaves the impression that the correct path to formation of a spectral estimate is clear: simply select a guiding principle (undoubtedly related to the notion of entropy), gather the available information, and solve the well defined mathematical problem that results. Seldom is the practical situation so simple.

Typically the numerical constraints are not given conveniently, say, in terms of exact knowledge of the autocorrelation function at equally spaced lags. More often, only a few irregularly spaced noise corrupted samples of the time series are available; from this data the numerical constraints must be estimated. Even when permitted the luxury of bountiful regularly spaced and noise-free data, numerous difficulties remain. Assuming a maximum entropy principle, should estimates of the autocorrelation, cepstral, or some other numerical constraints be formed? How should these estimates be formed and how many[9] of them should be formed?

The Yule AR(P) estimation procedure outlined at the beginning of this chapter provides one solution: having selected the model as AR and its order as P, form the biased autocorrelation lag estimates, Equation (2.6), and use these

---

[9]This is the problem of order determination. Various estimators of the order parameter, based upon notions of information theory, have been proposed and discussed by Akaike [18] and Parzen [19, 20] among others. Often the order parameter is selected simply upon the basis of experience with the phenomenon under study.

as if they were the true values. These autocorrelation lags then uniquely determine the AR(P) model parameters (and vice versa) via the Yule-Walker Equations (2.5). This description is explicit but fails to provide significant insight as to why it might be good. The formulation may be derived from a variety of viewpoints, each with its own merit and yielding greater understanding of the procedure.

Linear Prediction (LP) theory leads to one derivation of this formulation [21]. In this derivation the AR model is viewed as a predictor and the model parameters are determined to minimize the prediction error

$$e_n = x_n - \hat{x}_n = x_n + \sum_{m=1}^{P} a_m x_{n-m} \qquad (2.10)$$

in a mean-square sense. Depending upon the details of treatment of the ends of the data record one may derive the Yule-Walker procedure (also known as the "autocorrelation LP method") or a variant known as the "covariance LP method". Both of these methods have their proponents. The Linear Prediction theory is very similar to Yule's original considerations in which the e_n are viewed as random driving disturbances to the Pth order inhomogeneous difference Equation (2.10).

Other variants of the autocorrelation LP method are based upon a recursive lattice structure for the prediction filter [22]. In addition to the "forward" predictor $A_P(z)$, these variants consider a "backward" predictor, $B_P(z)$; both

predictors are characterized by the set of reflection coefficients $\{k_n; n=1,2,\ldots,P\}$ according to

$$A_n(z) = A_{n-1}(z) + k_n \, z^{-1} B_{n-1}(z) \; ; \; A_0(z) = 1 \qquad (2.11a)$$

$$B_n(z) = z^{-1} B_{n-1}(z) + k_n \, A_{n-1}(z) \; ; \; B_0(z) = 1 \qquad (2.11b)$$

The z-transform of the forward prediction error process after n filtering stages is simply $A_n(z) \, X(z)$; similarly the z-transform of the backward prediction error process is $B_n(z) \, X(z)$. Mean-square criteria are applied to the forward and backward error processes to obtain a variety of estimators for the reflection coefficients; one of particular importance, due to Burg [23], determines k_n to minimize the sum of the variances of the forward and backward error processes after n filtering stages. For truely ergotic processes, all these AR estimation procedures are asymptotically equivalent to the autocorrelation LP method for large values of N; as parameter estimation procedures these methods are most important for problems involving mildly nonstationary data of limited quantity.

In addition to these various "minimum mean square prediction error" formulations, another important derivation of the Yule procedure is due to Itakura and Saito [24]. Assuming an AR(P) model for the zero-mean stationary Gaussian process, they employ the maximum likelihood method and show

that the solution is obtained, asymptotically for large N, by minimizing a "spectral matching criterion"

$$I(f,g) = \int_{-\pi}^{\pi} \{[f(\theta)/g(\theta)] - \ln[f(\theta)/g(\theta)] - 1\} \, d\theta/2\pi \qquad (2.12)$$

where $f(\theta)$ is the Schuster periodogram given by Equation (2.1).

It is readily verified, by differentiating $I(f,g)$ with respect to the parameters of $g(\theta)$, that the minimum is obtained when the correlation matching property

$$\int_{-\pi}^{\pi} f(\theta) \, e^{in\theta} \, d\theta/2\pi = \int_{-\pi}^{\pi} g(\theta) \, e^{in\theta} \, d\theta/2\pi \qquad (2.13)$$

is satisfied for n=0,1,...P. By recognizing the left-hand side as the lag product autocorrelation estimates

$$r_n = \int_{-\pi}^{\pi} f(\theta) \, e^{in\theta} \, d\theta/2\pi \qquad (2.14)$$

the correlation matching property leads easily to the Yule-Walker Equations (2.5); see [25, pp. 445-6]. Recently Kay [26] has developed another variant by similarly applying the maximum likelihood method to zero-mean stationary Gaussian AR(P) processes but eliminating the large N approximation; again this variant treats the problem of limited data.

The functional (2.12), although it is usually attributed to Itakura and Saito in the current speech literature, was apparently first developed by Pinsker [27]. Assuming

only that the two processes are zero-mean and Gaussian, Pinsker showed[10]

$$\lim_{N \to \infty} \quad S(p,q)/N = I(p,q)/2 \qquad (2.15)$$

This theorem provides an information theoretic interpretation of the Itakura-Saito spectral matching criterion. Moreover, from a functional inference point of view, one might derive the Yule-Walker procedure by replacing q by an assumed AR(P) spectral model, $g(\theta)$, replacing p by a rough spectral estimate provided by $f(\theta)$, and then minimizing $I(f,g)$.

The last derivation should be contrasted with the minimum cross-entropy development discussed earlier. In that formulation the AR(P) form was derived from given correlation constraints while this formulation derives the correlation constraints from the given AR(P) form. Both developments employ (different) prior estimates and minimize a measure of information divergence between the prior and posterior estimates; however, the information divergence is not a symmetric measure and the unknown (posterior) estimate appears as the second argument in the current formulation

---

[10]The notation is somewhat abused here. On the left p and q represent the joint probability densities of N consecutive random variables; on the right p and q are power spectral density functions.

while it appears as the first argument in the minimum cross-entropy development. Nonetheless, the resultant procedures are both the same as the Yule procedure. In the next chapter a variant of this last derivation will be considered.

Noise Corruption

The problem of noise corruption to the observations pervades estimation problems. Generally all useful estimators are at least mildly tolerant of noise corruption while their performance degrades if the corruption becomes particularly severe. The most common problem considered is that of an additive independent noise process; this problem is of considerable importance in practical applications.

Upon initial reflection, the problem of estimating the parameters of both the noise and signal processes from time series observations alone may seem impossible. Indeed, the problem of determining the individual variances of two independent additive zero-mean stationary white Gaussian processes is completely confounded regardless of the quantity of data available. However, if one process is non-Gaussian, estimates of third and higher order statistics can be useful in estimating these lower order statistics. Parzen discusses the use of the "bispectrum" to estimate the spectrum of a non-Gaussian process in additive independent white Gaussian noise [28].

When both processes are Gaussian the problem is not always confounded. Since the sum of two additive

independent ARMA processes is also an ARMA process one might hope to find estimators for the parameters of the two additive processes when the number of parameters for the combined process is not exceeded by the total number of parameters of the two processes. For example, Pagano [29] discusses the problem of estimating the P + 2 parameters of additive AR(P) and white processes by first estimating the 2P + 1 parameters of a single equivalent ARMA(P,P) process and then using these 2P + 1 estimates to initialize a procedure for estimating the originally sought P + 2 parameters; it seems critical however that the order of the AR process does not degenerate (i.e. is actually nonzero).

This latter problem is fairly close in spirit to the problem considered in the following chapters. There the signal and noise processes are additive, independent, and zero-mean Gaussian; moreover, the signal process is AR(P). The problem may seem more complex because the noise process need not be white; however, a considerable simplification is achieved because the noise process spectral density (hence, all its statistics) is assumed to be known in addition to the time series observations. In practice the noise statistics are estimates provided by other observations but the large amount of data available for these estimates makes them quite reliable.

Noise Filtering

Wiener [30] considered the intimately related problem of extrapolating a time series from noise corrupted observations. When the zero-mean signal and noise processes are additive and independent with known power spectral density functions ($g(\theta)$ and $\mu(\theta)$ respectively) then the minimum variance linear extrapolating filter is the Wiener filter whose frequency response characteristic is

$$H(\theta) = g(\theta)/[g(\theta) + \mu(\theta)] \qquad (2.16)$$

This is sometimes referred to as the unrealizable Wiener filter since it is noncausal; the corresponding impulse response function extends both backward and forward in time to infinity. It is easy to show that the variance of the extrapolation can only be reduced to zero if the support of the signal spectrum has a null (or zero-measure) intersection with the support of the noise spectrum; in this case the frequency response, $H(\theta)$, will be unity on the support of $g(\theta)$ and zero elsewhere. Others, most notably Kalman [31], have since extended and refined Wiener's pioneering work.

A common procedure for dealing with additive noise is to first form a realizable estimate of the Wiener filter (or some other "optimal" filter), $\hat{H}(\theta)$, and apply it to the noise corrupted observations. The resulting data are then treated as noise-free observations of the signal process and

standard estimation procedures are employed to obtain an estimate of the signal spectrum. When the noise spectrum, $\mu(\theta)$, is known this procedure involves some mildly circular reasoning since Equation (2.16) indicates that knowledge of $H(\theta)$ is equivalent to knowledge of $g(\theta)$.[11] Nonetheless, this process has been demonstrated to be advantageous in speech analysis and other applications; a survey of these methods may be found in [32].

Much recent effort [33-39] has concentrated upon implementation structures and estimation procedures for $\hat{H}(\theta)$; typically these procedures employ side information in addition to the noise corrupted time series observations. Often the methods are nonlinear and time-varying with both theoretical and heuristic foundations. Regardless of the technique, one may always subsequently define a short-time-invariant linear equivalent frequency response characteristic in terms of the short-time input and output signal z-transforms, $X(z)$ and $Y(z)$, by

$$\hat{H}(\theta) = Y(e^{i\theta})/X(e^{i\theta}) \qquad (2.17)$$

---

[11]Hence we would have $\hat{g} = \mu\hat{H}/(1-\hat{H})$. The conceptual difficulties may be circumvented by considering the overall noise cancelling filter/spectral estimation scheme as a single estimation procedure; especially since the procedure usually does not employ (2.16) to form the final estimate of the signal spectrum.

One convenient categorization distinguishes frequency domain methods [33-36] from time domain methods [37-39]. Among the frequency domain methods, the noise cancelling filter frequency response characteristic usually appears explicitly; the simpler (and less heuristic) methods present $\hat{H}(\theta)$ as a function of the short-time signal to noise spectral density ratio estimate[12]

$$SNR(\theta;\alpha) = \{[f(\theta)]^{\alpha} \doteq [\mu(\theta)]^{\alpha}\}/[\mu(\theta)]^{\alpha} \qquad (2.18)$$

Two important classes of filter response characteristics are the subtraction class given by[13]

$$\hat{H}_1(\theta;\alpha,\beta) = \{SNR(\theta;\alpha)/[1 + SNR(\theta;\alpha)]\}^{\beta} \qquad (2.19)$$

and the soft suppression class given by

$$\hat{H}_2(\theta;\alpha,\beta) = \{[1 + \hat{H}_1(\theta;\alpha,1/2)]/2\}\{\Phi(\theta;\alpha,\beta)/[1 + \Phi(\theta;\alpha,\beta)]\}$$
$$(2.20a)$$

---

[12]Equation (2.18) employs the monus function, defined by $x \doteq y = (x-y + |x-y|)/2$, to insure a nonnegative result.

[13]Various special frequency response characteristics are worth separate mention here. The Wiener filter [30] frequency response is $\hat{H}_1(\theta;1,1)$. The power subtraction filter and the magnitude subtraction filter [35] have frequency response characteristics $\hat{H}_1(\theta;1,1/2)$ and $\hat{H}_1(\theta;1/2,1)$ respectively. Finally, the soft suppression class due to McAulay and Malpass [36] has the frequency response $\hat{H}_2(\theta;1,\beta)$.

where

$$\Phi(\theta;\alpha,\beta) = \exp[-\beta] \, I_0[2 \sqrt{\beta[1 + SNR(\theta;\alpha)]} \,] \qquad (2.20b)$$

and $I_0[\cdot]$ denotes the zeroth order modified Bessel function of the first kind. These "suppression rules" are plotted for selected values of α and β as a function of $SNR(\theta)$ in Figure 1.

Effect on Resolution

In speech applications, vocal tract resonances are not extremely sharp and are moderately well separated in frequency; consequently one is generally concerned with accurate estimation of the spectral shape and high resolution estimation is not a priority.[14] In other applications (such as sonar, radar, and medicine) accurate frequency estimation and resolution of discrete ("line") and narrowband spectra are issues of fundamental importance. Periodogram and Blackman-Tukey spectral estimates have a fundamental frequency resolution limit determined by the length of the observation interval; AR estimators have become quite popular due, in part, to their greatly improved resolving power.

---

[14]Hence, even very low resolution methods that divide the (4 kHz) voice bandwidth into fewer than two dozen "channels" can be quite effective.
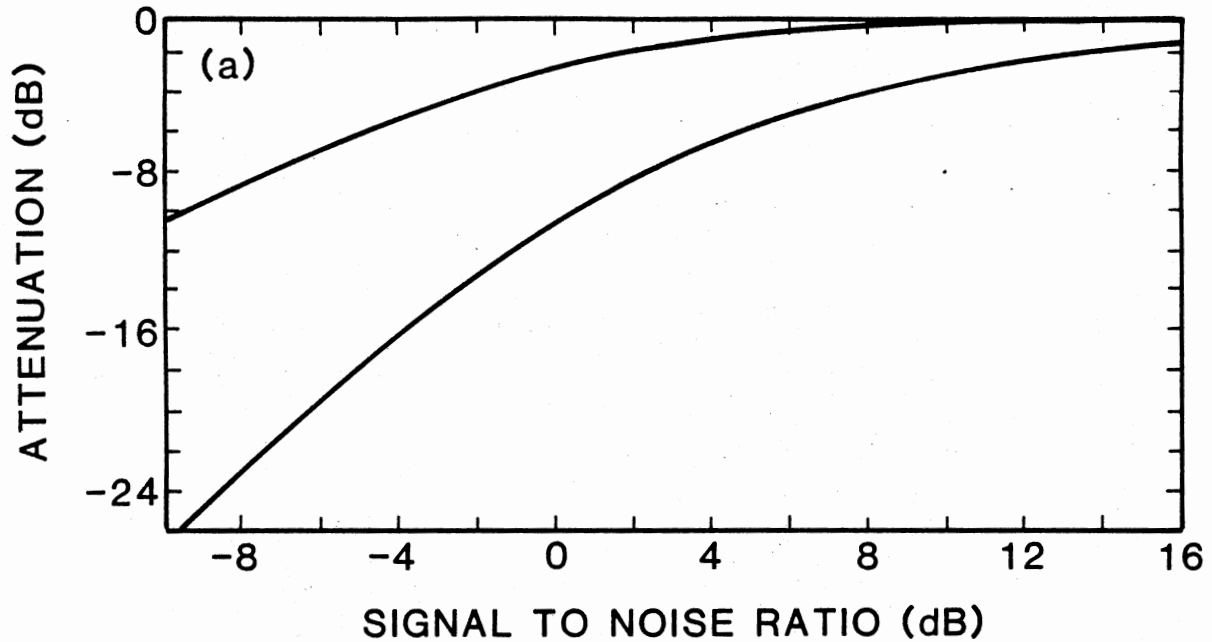
(a) Subtraction Class. Top Curve: Power Subtraction $\hat{H}_1(\theta;1,1/2)$. Bottom Curve: Magnitude Subtraction $\hat{H}_1(\theta;1/2,1)$
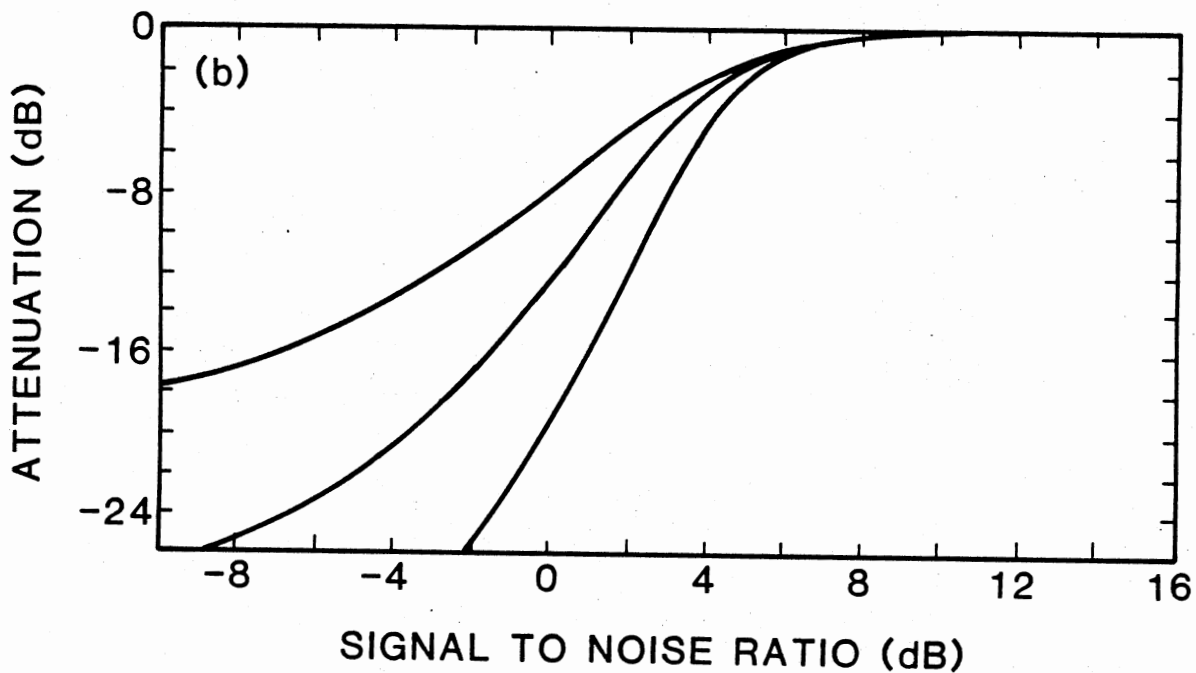
(b) Soft Suppression Class; $\hat{H}_2(\theta;1,\beta)$. Top to Bottom $\beta = 4$, 6, 8, Respectively

Figure 1. Noise Filter Characteristics

Still, the resolution (as well as other performance indicators) varies among the different AR estimators and, for each, is influenced by a variety of factors.

Noise corruption is one of the important factors limiting the resolving power of AR estimators. Several authors have considered the problem of estimating the parameters of a fixed number of sinusoids from discrete-time observations corrupted by zero-mean additive white Gaussian noise of unknown variance. For this specialized problem the Cramer-Rao performance bounds[15] may be computed [40]. As is well known, the complicated nonlinear maximum likelihood estimation procedure will achieve these bounds; Tufts and Kumaresan [41], using AR estimation procedures, have developed computationally simpler high resolution frequency estimators that nearly achieve these bounds while Cadzow, et. al. [42] claim still better performance using a singular value decomposition (SVD) approach. In many practical circumstances additional information may be available so that

---

[15]In general, the Cramer-Rao bounds indicate the minimum variance a parameter estimate can achieve [43]. An estimate achieving the minimum variance is an "efficient" estimate. In [40] the bounds upon an unbiased frequency estimate are considered (they depend upon the assumed distribution as well as the number of data points) and are presented as a function of the signal to noise ratio. In [44], the efficiency loss of any method based upon the use of correlation estimates instead of the original data is studied.

these bounds may be exceeded;[16] for example, Quirk and Liu [45] describe a simple filtering and decimation scheme (which employs knowledge of the frequency bands in which the sinusoids are located) that improves the resolution of (any) subsequent AR estimator. In a similar vein, adaptive pre-filters (that employ a reference process correlated with either the signal or noise portion of the objective process, but not both) have been devised to "enhance" narrowband signals in noise [46].

Quantization and Computation

While spectral estimation, per se, is not concerned with the problems of quantization and computation, the ultimate utility of an estimation procedure can depend strongly upon these (and other) issues. If the procedure explicitly recognizes that only one of a finite predefined set of conclusions can be reached, the situation is sometimes distinguished by referring to the "detection" (instead of the "estimation") problem.

In many digital speech recognition and communication systems the goal of spectral analysis is to solve a detection problem; in addition, the system designer must solve the problem of selecting the best finite set of models to

---

[16]More precisely, the true bounds are reduced by the availability of additional information. Consequently new estimators that account for this additional information can be devised that outperform (in terms of variance) any estimator that does not account for the additional information.

employ. Until recently, these systems would find the solution to an estimation problem and then employ a (somewhat ad hoc) quantization procedure to select a model from among the finite set. If the number of models in the finite set was sufficiently large, this procedure could be quite effective; however, one measure of goodness for the finite set of models is often how few models are in the set.

In the past decade technological advances have permitted the use of increasingly complex computational procedures while still meeting size/cost/power constraints imposed by the application. Consequently more sophisticated and effective (but previously unmanagable) techniques for estimation/detection and quantization of spectral models have been studied in earnest. The numerous variants of a class of techniques generally referred to as "vector quantization" [47-53] have recently achieved considerable success by reducing the finite number of models by about 9 orders of magnitude with only slight degradation in other measures of system performance.

Many of these vector quantization techniques are founded upon minimization of the asymptotic information divergence $I(f,g)$. Of considerable interest in the use of this measure is the triangle equality property; if $g(\theta)$ minimizes $I(f,g)$ over the set of all stable AR(P) models and $h(\theta)$ is any other model in a (possibly finite) subset then

$$I(f,h) = I(f,g) + I(g,h) \qquad (2.21)$$

As a consequence of this property one may solve the detection problem, which minimizes $I(f,h)$, by first solving the estimation problem, which minimizes $I(f,g)$, and then solving the quantization problem which minimizes $I(g,h)$.

Remarks

The general problem of spectral estimation has been discussed; this discussion has emphasized issues and methods associated with autoregressive estimation. Autoregressive spectral models are important in numerous practical applications; consequently they have received considerable attention in the literature. The AR form may be derived from either the maximum entropy or the minimum cross-entropy principle when correlation constraints are considered; alternatively the AR form may be assumed and correlation constraints derived using a linear prediction formulation. The correlation constraints, together with the AR form, are sufficient to derive the Yule-Walker equations which relate the model parameters to the prescribed correlation values.

The asymptotic maximum likelihood formulation of Itakura and Saito assumes an AR form and derives the correlation constraints; in the course of this development a "spectral matching criterion" is minimized. The earlier derivation by Pinsker of this spectral matching criterion from an asymptotic information divergence formulation makes clear that, while the AR form is necessary to derive the

Yule-Walker equations, the spectral matching criterion is applicable independant of the spectral model form.

Noise corruption pervades estimation problems and useful estimators are generally at least mildly tolerant of additive noise. Often additional data is available to help characterize or distinguish the noise and signal processes; many estimation problems are concerned with the development of effective and computationally feasible methods for incorporating this additional information. A common procedure, employed when an accurate noise spectrum estimate is known, first applies an estimated noise cancelling filter to the corrupted data and then uses the output as "noise-free" data from which to estimate the signal spectrum. Ultimately the effect of noise corruption will be to decrease the best performance possible with any spectral estimator.

In the following chapters a new spectral estimator is developed. As is common, the fundamental observations are assumed to be equally spaced samples of a zero-mean stationary Gaussian time series corrupted by additive independent zero-mean stationary Gaussian noise of known power spectral density, $\mu(\theta)$. This problem occurs in many applications involving speech analysis (as well as others) wherein the noise spectrum is estimated from data taken during speech inactivity.

The amount of data available to estimate the signal spectrum is usually limited by the nonstationary character of speech; the speech statistics are usually stationary only

over very short time intervals varying in duration. One study [54] has observed speech waveforms and subjectively judged that the duration for which a segment may be considered stationary varies from about 4 ms. to over 360 ms. with most of the distribution contained in the range of 12 ms. to 174 ms.; most speech analysis systems employ a fixed analysis interval approximately 20 to 25 ms. in duration. The use of a fixed analysis interval (with no particular attempt at optimum time alignment of end points) is simply a practical method of limiting the computational burden; while suboptimal spectral estimates are thereby achieved for long acoustic events, perhaps the most severe deleterious effect is the slurring of very short events and transitions.

In order to employ at a later time a noise estimate obtained during speech inactivity, the noise statistics are assumed to remain stationary over much longer time intervals; since one of the primary noise sources is ambient environmental noise acoustically coupled to the speech, the validity of this assumption must be checked in each situation. In many practical circumstances the noise is stationary over long intervals; for example, in aircraft, the noise statistics typically vary only with the flight condition. On the other hand, if the corrupting noise is another speech signal the assumption of long term noise stationarity is certainly invalid.

CHAPTER III

THEORETICAL FORMULATION

In this chapter several related procedures for esti-
mating AR(P) process parameters from noise corrupted time
series observations are developed. In the first section the
problem is motivated as one arising in speech applica-
tions. In the next section an ideal formulation is discus-
sed; unfortunately the resulting nonlinear system of
equations is sufficiently complicated to make analytical
solution intractable.[1] In the third section a first ap-
proximation to the ideal formulation is developed and shown
to be essentially equivalent to the noise filtering pro-
cedures discussed in Chapter II. In the fourth section a
second, improved, approximate formulation employing a
weighted information measure is developed;[2] some important

---

[1]Numerical solution may be feasible in some cases but
this is not investigated in the present work.

[2]This weighted information formulation assumes a cen-
tral role in this work. In fact, this was the original
foundation and was developed heuristically following the
work of Chu and Messerschmitt [55, 56]. The theoretical
foundation (as an approximation to the "ideal" formulation)
was subsequently developed because the heuristic development
could only specify the weight function qualitatively and a
more quantitative characterization was required.

properties of the weighted information measure are derived in the fifth section. Finally, the last section reflects upon these formulations, their relationship to other estimation procedures, and problems of spectral estimation and speech analysis to which they may be applied.

Application to Speech Analysis

Acoustic events in speech are often modeled as a white zero-mean Gaussian stationary excitation of a linear system. The linear system response is usually identified with the vocal cavity response which depends upon the position of speech articulators (tongue, lips, teeth, etc.); the excitation is usually assumed to be physically localized although its position may vary with different speech events.

The linear system model may be criticized in various ways; still it has had considerable success in practical situations. The particular case of an AR (or all-pole linear) system model can be justified on the basis of a lossless acoustic tube of varying cross-sectional area. The analogy of an acoustic tube with the oral or nasal cavity alone is clear; however, some speech sounds reflect the combined response characteristics of the oral and nasal cavities indicating that a full ARMA model would be more appropriate. A more complete discussion of acoustic tube modeling of the vocal tract may be found in [21].

Based upon the considerable success of AR models in speech applications, as well as the physical analogies that

may be drawn between AR models and the vocal tract via acoustic tube modeling, the AR speech model is adopted here. In most applications the deleterious effects of the pressure transducer, analog amplifier, anti-aliasing prefilter, and the digitizer have been carefully minimized and may be ignored. Some applications permit the system designer to ensure that the pressure transducer response reflect only the speech of the intended speaker; more often, conflicting goals deny the designer this flexibility so that the microphone transduces other ambient environmental acoustic events that appear as unwanted "noise" in the observed signal. Consequently, while the AR model is adopted for the speech spectrum, it is inadequate as a model for the observed signal spectrum.

Some ambient noise is a direct environmental response to the speech itself (e.g. echoes) or is short, transient, and generally unpredictable by nature (e.g. a gunshot, dropped book, engine backfire, cough, etc). Other ambient noise is repetitive (e.g. machine-gun fire) or steady by nature (e.g. drone of engines, rushing air, running water, whine of a turbine). This last (steady) type of noise is the primary focus of many speech analysis systems; typically these systems exploit the steady nature of the noise to determine noise statistics during speech activity from signal observations made during speech inactivity. With multiple transducers (or other clever system design techniques) the statistics of a much broader class of noises may be

known during speech activity. In the following it is only assumed that, during each analysis interval, the noise in the primary (objective) observation signal be zero-mean Gaussian stationary additive and independent of the speech; the noise is, therefore, completely characterized by a spectral density function, $\mu(\theta)$, which is assumed to be known.

The goals of speech analysis are many and varied. In communications the goal is often to achieve a minimal data rate subject to a quality or communicability constraint. In artificial intelligence the goal is usually to "understand" the speech with phonetic or written transcription often arising as an intermediate step. Some other goals include the identification of the speaker, the identification of the language, translation of the voice of one speaker to that of another in the same or a different language, and the screening/diagnosis of disease (e.g. laryngeal cancer). Spectral estimation is at the foundation of speech analysis for all these goals and accurate AR model estimation in noise is fundamental to the estimation of speech spectra in practical environments.

Ideal Formulation

Let

$$h(\theta) = g(\theta) + \mu(\theta) \tag{3.1}$$

be the observed process power spectral density model where $\mu(\theta)$ is the known additive noise process spectrum and $g(\theta)$ is the unknown AR(P) power spectral density model characterizing the signal process; see Equation (2.3). Let $f(\theta)$ be the Schuster periodogram defined for the N time series observations by Equation (2.1). If the signal and noise processes are independent zero-mean real stationary Gaussian processes then the maximum likelihood method is asymptotically equivalent, for large N, to minimizing $I(f,h)$ with respect to the AR(P) process parameters. Any parameter set minimizing $I(f,h)$ and corresponding to a stable AR(P) process shall be considered here to be an ideal solution to the estimation problem.

This formulation of the estimation problem as a minimization problem may also be derived from an information theoretic viewpoint. Let $\widetilde{f}(\theta)$ be the true observed process power spectral density so that $I(\widetilde{f},h)$ represents the asymptotic information divergence between the true spectrum and an arbitrary model spectrum. Clearly it is desirable to find the model $h(\theta)$ minimizing $I(\widetilde{f},h)$; if the minimum value is zero then $h(\theta) = \widetilde{f}(\theta)$ almost everywhere. Since $\widetilde{f}(\theta)$ is unavailable, replace it by a rough estimate, $f(\theta)$, and find $h(\theta)$ to minimize $I(f,h)$.

Minimization of $I(f,h)$ is subject to several interesting interpretations; the maximum likelihood and minimum information divergence interpretations have been given

above, a third noise filtering interpretation is now provided. Notice that $I(f,h) = I(Hf,g)$ where $H(\theta)$ is the frequency response of the Wiener filter given in Equations (2.16). The quantity $H(\theta)f(\theta)$ may be interpreted as a rough estimate of the spectrum of a process obtained by passing the observed process through a filter whose power[3] spectral response characteristic is $H(\theta)$; minimization of $I(f,h) = I(Hf,g)$ may then be understood as a standard LP (or maximum entropy, etc.) fit to the noise filtered process. Of course, $H(\theta)$ is not known but is a function of the unknown parameters of $g(\theta)$; one must simply imagine finding a parameter set defining $H(\theta)$ that also corresponds to the best LP fit, $g(\theta)$, to the output process.

The functional $I(f,h)$ is minimized by computing its derivative with respect to each parameter of $g(\theta)$ and setting the result to zero. For an arbitrary parameter, ξ, this is

$$\int_{-\pi}^{\pi} \{[H(\theta)\ g(\theta) - H^2(\theta)\ f(\theta)]/g^2(\theta)\}\ (\partial g(\theta)/\partial \xi)\ d\theta/2\pi = 0$$

$$(3.2)$$

---

[3]This is not to say that the observed process is passed through a Wiener filter whose frequency response is $H(\theta)$. Recall that the Wiener filter is designed to minimize the mean-square prediction error; the output process doing this does not have the signal process spectrum, $g(\theta)$, but instead the spectrum $H(\theta)g(\theta)$. Alternatively, $H(\theta)f(\theta)$ may be interpreted as a rough estimate of the cross-spectrum between the input and output processes of the Wiener filter.

Using Equations (2.3) and (2.4), the partial derivatives of $g(\theta)$ are

$$\partial g(\theta)/\partial \sigma^2 = g(\theta)/\sigma^2 \tag{3.3a}$$

and, for $\ell = 1, 2, \ldots, P$

$$\partial g(\theta)/\partial a_\ell = -g^2(\theta) \sum_{m=0}^{P} 2a_m \cos[(\ell-m)\theta]/\sigma^2 \tag{3.3b}$$

Defining[4]

$$V_n = \int_{-\pi}^{\pi} \{H^2(\theta)\, f(\theta) - H(\theta)\, g(\theta)\}\, e^{in\theta}\, d\theta/2\pi \tag{3.4}$$

and substituting Equation (3.3) in Equation (3.2) yields

$$\sum_{m=0}^{P} (a_m/\sigma^2) \sum_{\ell=0}^{P} (a_\ell/\sigma^2)\, V_{\ell-m} = 0 \tag{3.5a}$$

and, for $\ell = 1, 2, \ldots, P$

$$\sum_{m=0}^{P} (a_m/\sigma^2)\, V_{\ell-m} = 0 \tag{3.5b}$$

while a little further manipulation of Equations (3.5)

---

[4]It is worth noting that the quantities, V_n, defined by Equations (3.4) are the components of the gradient vector of $I(Hf,g)$ where differentiation is defined with respect to the inverse correlation parametrization of $g(\theta)$; see Equation (3.22).

yields, for $\ell=0,1,\ldots,P$

$$\sum_{m=0}^{P} a_m V_{\ell-m} = 0 \qquad (3.6)$$

The symmetry of the functions $f(\theta)$, $g(\theta)$, and $H(\theta)$ may be used to demonstrate that $V_{-n} = V_n$ while it is easy to see that Equations (3.6) are satisfied if

$$V_n = 0 \; ; \; n=0,1,\ldots,P \qquad (3.7)$$

To show that Equations (3.7) must be satisfied if a stable filter is to be obtained, rewrite the system of Equations (3.6) in matrix form as

$$\left\{ \begin{bmatrix} 1 & a_1 & \cdots & a_{P-1} & a_P \\ a_1 & a_2 & \cdots & a_P & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{P-1} & a_P & \cdots & 0 & 0 \\ a_P & 0 & \cdots & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & a_{P-2} & \cdots & 1 & 0 \\ 0 & a_{P-1} & \cdots & a_1 & 1 \end{bmatrix} \right\} \begin{bmatrix} V_0 \\ V_1 \\ \vdots \\ V_{P-1} \\ V_P \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

$$\qquad (3.8)$$

The coefficients of a stable $P-1^{\text{st}}$ order predictor $\{\hat{a}_n; \; n = 1,2,\ldots,P-1\}$ are given recursively in terms of a stable P^{th} order predictor according to

$$[I + k_P \, J]^{-1} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_{P-1} \\ a_P \end{bmatrix} = \begin{bmatrix} 1 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_{P-1} \\ 0 \end{bmatrix} \qquad (3.9)$$

where I is the identity matrix, J is the reversal matrix

$$J = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix} \qquad (3.10)$$

$k_P = a_P$ is a reflection coefficient[5] and

$$[I + k_P J]^{-1} = [I - k_P J]/(1 - k_P^2) \qquad (3.11)$$

Applying the nonsingular transformation[6] $[I + k_P J]^{-1}$ to Equation (3.8) does not change the solution and yields

$$\left\{ \begin{bmatrix} 1 & \hat{a}_1 & \cdots & \hat{a}_{P-1} & 0 \\ \hat{a}_1 & \hat{a}_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \hat{a}_{P-1} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & \hat{a}_{P-2} & \cdots & 1 & 0 \\ 0 & \hat{a}_{P-1} & \cdots & \hat{a}_1 & 1 \end{bmatrix} \right\} \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{P-1} \\ v_P \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

$$(3.12)$$

---

[5]These are the same reflection coefficients used in the forward-backward recursion; see Equation (2.11).

[6]Bounded input, bounded output (BIBO) stability requires and is guaranteed by the condition $|k_n| < 1$ for $n = 1,2,\ldots,P$ which also guarantees that the indicated transformation is nonsingular.

The last equation shows V_P to be a linear combination of $V_O, V_1, \ldots, V_{P-1}$ and the reduced system

$$\left\{\begin{bmatrix} 1 & \hat{a}_1 & \cdots & \hat{a}_{P-1} \\ \hat{a}_1 & \hat{a}_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \hat{a}_{P-1} & 0 & \cdots & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & \hat{a}_{P-1} & \cdots & 1 \end{bmatrix}\right\} \begin{bmatrix} V_O \\ V_1 \\ \vdots \\ V_{P-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (3.13)$$

is of the same form as Equation (3.8). Consequently, stability requires that each V_n be a linear combination of the previous V_ℓ, $\ell = 0, 1, \ldots, n-1$, while the final reduced system is simply $V_O = 0$. Hence, if only stable minima of $I(f,h)$ are sought these minima must satisfy Equations (3.7) which may be rewritten, for $n = 0, 1, \ldots, P$, as

$$\int_{-\pi}^{\pi} H(\theta) \, H(\theta) \, f(\theta) \, e^{in\theta} \, d\theta/2\pi = \int_{-\pi}^{\pi} H(\theta) \, g(\theta) \, e^{in\theta} \, d\theta/2\pi \quad (3.14)$$

This is a highly complicated nonlinear system of equations that appears to be very difficult to solve analytically. Note that, in the absence of noise, $\mu(\theta) = 0$ and $H(\theta) = 1$ so that the system reduces to Equations (2.13) as expected; in this case it is well known that the system always possesses a unique stable solution.

In general no admissable solution exists; the following example will serve to illustrate. Consider an AR(0) process corrupted by white noise of known variance μ. The system of

equations reduces to

$$r_0 = \int_{-\pi}^{\pi} f(\theta)\ d\theta/2\pi = \sigma^2 + \mu \qquad (3.15)$$

If $r_0 \geq \mu$ the system is solved by $\sigma^2 = r_0 - \mu$ which yields the minimum value $I(f,h) = I(f,r_0) = 0$. If $r_0 < \mu$ the system does not possess a real solution; however, $I(f,h)$ is always minimized by selecting $\sigma^2 = r_0 \doteq \mu$.

Noise Filtering Formulation

Since Equations (3.14) appear so difficult to solve, it is natural to consider alternate formulations. From the observation that $I(f,h) = I(Hf,g)$ and the interpretation of $H(\theta)$ as the power spectral response of a noise filter a simple and reasonable procedure is to replace $H(\theta)$, which depends upon unknowns, by an estimate $\hat{H}(\theta)$. Several classes of estimates have been presented in Equations (2.19) and (2.20).

Once the data has been processed by the filter with power response $\hat{H}(\theta)$ a "noise-free" rough estimate is available

$$\hat{f}(\theta) = \hat{H}(\theta)\ f(\theta) \qquad (3.16)$$

Then, minimization of $I(\hat{f},g) = I(\hat{H}f,g)$ is achieved by the

solution to the equations

$$\int_{-\pi}^{\pi} \hat{H}(\theta) \ f(\theta) \ e^{in\theta} \ d\theta/2\pi = \int_{-\pi}^{\pi} g(\theta) \ e^{in\theta} \ d\theta/2\pi \qquad \text{for } n=0,1,\ldots,P \qquad (3.17)$$

This, of course, leads easily to the Yule-Walker equations with the difference that the estimated correlation values are now given by the left-hand side of Equation (3.17); the reader is urged to compare this equation with Equations (3.14) and (2.13).

Weighted Information Formulation

The previous approximate formulation encompasses a wide variety of estimation procedures that have been studied in recent years. If $\hat{f}(\theta)$, given by Equation (3.16), is a good rough estimate of the noise-free power spectral density the resultant model parameters can be expected to be accurate. Consequently, considerable effort has been expended trying to find the best form of $\hat{H}(\theta)$ and, ultimately, the best means of computing the correlation values on the left hand side of Equation (3.17).

Generally speaking, any estimate can be expected to be more accurate if there is less corrupting noise; in particular, $\hat{f}(\theta)$ can be expected to be more accurate in those spectral regions where the signal to noise density ratio is large. Since the reliability of the rough estimate $\hat{f}(\theta)$ varies with frequency, the criteria for fitting a model to

$\hat{f}(\theta)$ should reflect this variation in reliability. The frequency weighted spectral distance measure introduced by Chu and Messerschmitt [55, 56] provides precisely the required flexibility for such a criteria. The criteria is derived from the asymptotic information divergence, $I(f,g)$, by noting that the integrand in Equation (2.17) is a nonnegative error measure; the frequency weighted variant is obtained by introducing a multiplicative nonnegative weight function to the integrand of $I(f,g)$ to yield

$$I_W(f,g) = \int_{-\pi}^{\pi} W(\theta)\{[f(\theta)/g(\theta)] - \ln[f(\theta)/g(\theta)] - 1\} \, d\theta/2\pi$$

$$(3.18)$$

If $W(\theta)$ is constant, minimization of $I_W(\hat{f},g) = I_W(\hat{H}f,g)$ is equivalent to minimization of $I(\hat{f},g) = I(\hat{H}f,g)$. To reflect the greater reliability of $\hat{f}(\theta)$ in some spectral regions, $W(\theta)$ should be selected to be large where the signal to noise density ratio is large. To remain consistent with AR estimation procedures that work well in the absence of noise, $\hat{H}(\theta)$ should approach unity and $W(\theta)$ should approach a constant as $\mu(\theta)$ approaches zero. Specific procedures for selecting $\hat{H}(\theta)$ have been studied in the past [32-39] and important examples are given in Equations (2.19) and (2.20); the above considerations provide a qualitative understanding of an appropriate selection for $W(\theta)$ but a more specific, quantitative understanding is required.

To minimize $I_W(\hat{H}f,g)$ Equation (3.18) is differentiated with respect to the parameters of $g(\theta)$ and the results are set to zero. The procedure is the same, mutatis mutandis, as that followed for minimizing $I(Hf,g)$ and yields the system of equations

$$\int_{-\pi}^{\pi} W(\theta) \; \hat{H}(\theta) \; f(\theta) \; e^{in\theta} \; d\theta/2\pi = \int_{-\pi}^{\pi} W(\theta) \; g(\theta) \; e^{in\theta} \; d\theta/2\pi \quad (3.19)$$

Comparison of Equations (3.19) to Equations (3.14), which result from the ideal formulation, immediately suggests the required quantitative criteria for selecting $W(\theta)$. Specifically, $W(\theta)$ should be selected so that, at least approximately,

$$W(\theta) = \hat{H}(\theta) \quad\quad\quad (3.20)$$

and $\hat{H}(\theta)$ should estimate $H(\theta)$. This selection is supported by the previous heuristic considerations which indicated that $W(\theta)$ should be large where the signal to noise density ratio is large.

Properties of the Weighted Information

In this section three important results concerning the weighted information measure, $I_W(f,g)$, are developed. These results also apply to the asymptotic information divergence, $I(f,g)$, as a special case where $W(\theta) = 1$. The first result generalizes the triangle equality property for $I(f,g)$, see

Equation (2.21); that this property generalizes appropriately is of interest to the use of the weighted information measure in place of the (unweighted) asymptotic information divergence for vector quantization.

The Kullback information number and the asymptotic information divergence are well known to be convex with respect to general classes of probability and spectral densities. With the appropriate definition for convex superposition of AR(P) spectra, the second important result is that the class of stable AR(P) spectra is convex and the weighted information measure is strictly convex with respect to this class.[7] As a consequence, $I_W(\hat{H}f, g)$ can have at most one local minimum with respect to this class; moreover, if such a minimum exists it is also a global minimum.

Finally, the third result shows that the second mixed partial derivative of $I_W(\hat{H}f, g)$ defines a positive definite quadratic form. This shows that any stable solution to Equation (3.19) is a local minimum of $I_W(\hat{H}f, g)$; this could also have been demonstrated using the strict convexity. Combined with the previous result this shows that Equation

---

[7]A set, $\mathscr{S}$, is convex if it always contains the convex superposition of two elements in the set. A convex superposition is a map $x_3 = CS(x_1, x_2; \gamma)$ defined for $0 \leq \gamma \leq 1$ and all $x_1, x_2 \in \mathscr{S}$ such that $x_3 = x_1$ if $\gamma = 1$ and $x_3 = x_2$ if $\gamma = 0$; if $x_1 = x_2$ then $x_3 = x_2 = x_1$ for all γ. A function $f(x)$ defined on a convex set $\mathscr{S}$ is said to be convex if $\gamma f(x_1) + (1-\gamma) f(x_2) \geq f(x_3)$ and strictly convex if equality implies $x_1 = x_2$.

(3.19) can have at most one stable solution (although unstable solutions can, and often do, exist); moreover, if such a solution exists, it is the global minimum among stable AR(P) spectral models.

The question of existence is not addressed in this set of results. The existence of a stable solution to Equations (3.19) is assumed but remains an open question in general; existence can be demonstrated in special cases, e.g. $W(\theta) = 1$, while experimental results are discussed in Chapter V. Because the proofs are nonconstructive, they do not assist with the question of existence nor do they provide algorithms for computation of a solution; computational procedures are discussed in Chapter IV. It is worth noting that if no solution to Equations (3.19) exists then, since $I_W(\hat{H}f, g)$ must possess a minimum in the closure of the set of stable AR(P) spectra, the minimum occurs as a limit point of the set.

To simplify the following discussion the set of stable AR(P) spectra shall be denoted $\mathscr{R}_P$. Each element of the set may be characterized by a P+1-tuple of real parameter values satisfying appropriate (stability) criteria. Four characterizations of $\mathscr{R}_P$ are presented below:

Predictor Coefficients. Let $A_P(z)$ be given by Equation (2.4) with all roots of $A_P(z)$ inside the unit circle. Then $(\sigma, a_1, a_2, \ldots, a_P)$ denotes an arbitrary element of $\mathscr{R}_P$ if $\sigma > 0$.

Reflection Coefficients. Let $A_P(z)$ be given by Equation (2.11) with $|k_n| < 1$ for $n=1,2,\ldots,P$. Then $(\sigma,k_1,k_2,\ldots,k_P)$ denotes an arbitrary element of $\mathcal{R}_P$ if $\sigma > 0$.

Autocorrelation Coefficients. Let the real symmetric Töeplitz quadratic form given by

$$T(\bar{x}) = \sum_{m,n=0}^{P} r_{|m-n|}\, x_m\, x_n \qquad (3.21)$$

be positive definite. Then $(r_0,r_1,\ldots,r_P)$ denotes an arbitrary element of $\mathcal{R}_P$.

Inverse Correlation Coefficients. Let

$$1/g(\theta) = \sum_{n=-P}^{P} u_{|n|}\, e^{in\theta} \qquad (3.22)$$

be a positive function of θ in $[-\pi,\pi)$. Then $(u_0,u_1,\ldots,u_P)$ denotes an arbitrary element of $\mathcal{R}_P$.

These represent only a few of the infinitely many ways of characterizing $\mathcal{R}_P$. The first three parametrizations are well known with the corresponding terminology well established in the literature. Each set of predictor coefficients is related to a unique set of reflection coefficients by a continuous bijection defined by the Levinson-Durbin recursion. Each set of autocorrelation coefficients defines a unique set of predictor coefficients according to the Yule-Walker equations while the autocorrelation coefficients

may be retrieved from the predictor coefficients using Equations (2.3) and (2.8).

The last parametrization is less common than the other three; these parameters have been denoted "inverse correlation coefficients" since they are the autocorrelation coefficients of a moving-average process whose spectral density function is inverse to that of the defined AR(P) spectrum. Each set of predictor coefficients uniquely defines the inverse correlation coefficients according to

$$u_n = \sum_{m=0}^{P-n} a_m \, a_{m+n}/\sigma^2 \quad ; \quad a_o = 1 \quad ; \quad n=0,1,\ldots,P \qquad (3.23)$$

That the predictor coefficients may be retrieved in a unique fashion from the inverse correlation coefficients is more difficult to establish. Positivity of Equation (3.22) generally establishes only the possibility of several appropriate predictor coefficient sequences; closer inspection reveals that only one of these sequences satisfies the stability requirements. The question is taken up in somewhat greater detail by Blackman and Tukey [5, pp. 126-7].

The first result follows easily using the inverse correlation coefficient parametrization of the AR(P) spectral density, Equation (3.22), together with Equations (3.19) and (3.18).

Theorem 3-1. (Triangle Equality). Let $g_1(\theta)$ be an AR(P) spectral density satisfying Equation (3.19) and let $g_2(\theta)$ be any other AR(P) spectral density. Then

$$I_W(\hat{H}f, g_2) = I_W(\hat{H}f, g_1) + I_W(g_1, g_2) \tag{3.24}$$

The inverse correlation coefficient parametrization of AR(P) models in $\mathscr{R}_P$ is used here to define the convex superposition of two models according to

$$\overline{u}_3 = CS\ (\overline{u}_1, \overline{u}_2; \gamma) = \gamma\overline{u}_1 + (1-\gamma)\ \overline{u}_2 \tag{3.25}$$

for $0 \leq \gamma \leq 1$. Since (3.22) remains a strictly positive function for $\overline{u}_3$ when $\overline{u}_1$ and $\overline{u}_2$ define strictly positive functions, this shows $\mathscr{R}_P$ to be a convex set and leads to the second result.

Lemma 3-1. (Strict Convexity). Let $g_3(\theta)$ be a stable AR(P) spectrum defined by the convex superposition of the two stable AR(P) spectra $g_1(\theta)$ and $g_2(\theta)$. Then

$$I_W(f, g_3) \leq \gamma I_W(f, g_1) + (1-\gamma)\ I_W(f, g_2) \tag{3.26}$$

for $0 \leq \gamma \leq 1$ with equality only if $g_1(\theta) = g_2(\theta)$.

Proof. Using the inverse correlation coefficient parametrization and the definition of convex superposition for AR(P) spectra it is easy to show that

$$g_3(\theta) = 1/\{[\gamma/g_1(\theta)] + [(1-\gamma)/g_2(\theta)]\} \qquad (3.27)$$

Together with Equations (3.18) this yields

$$\gamma I_W(f,g_1) + (1-\gamma) I_W(f,g_2) - I_W(f,g_3)$$
$$= \int_{-\pi}^{\pi} W(\theta) \ln\{[g_1(\theta)]^\gamma [g_2(\theta)]^{1-\gamma}/g_3(\theta)\} \, d\theta/2\pi \qquad (3.28)$$

From the theorem on geometric and harmonic means the argument of the logarithm in Equation (3.28) is not less than one and equals one only if $g_1(\theta) = g_2(\theta)$. The lemma follows easily.

Theorem 3-2.   (Uniqueness). $I_W(f,g)$ can have at most one local minimum in $\mathscr{R}_p$; if such a minimum exists it is also a global minimum.

Proof.   Let $g_1(\theta)$ and $g_2(\theta)$ be two distinct local minima and form their convex superposition $g_3(\theta)$. Without loss of generality assume $I_W(f,g_1) \geq I_W(f,g_2)$. With $\gamma \neq 1$ the previous lemma gives

$$I_W(f,g_3) < \gamma I_W(f,g_1) + (1-\gamma) I_W(f,g_2) < I_W(f,g_1) \qquad (3.29)$$

But $g_3(\theta)$ is arbitrarily close[8] to $g_1(\theta)$ for γ arbitrarily

---

[8]The Euclidean metric applied to the inverse correlation coefficients shall suffice to define closeness here.

close to one, so that this inequality contradicts the assumption that $g_1(\theta)$ is a local minimum. The second part of the theorem follows by assuming $g_1(\theta)$ is a local minimum while $g_2(\theta)$ is any distinct element of $\mathscr{R}_P$ such that $I_W(f,g_2) \leq I_W(f,g_1)$ and then repeating the above argument.

In order to establish the final theorem of this section the second mixed partial derivative of $I_W(\hat{H}f,g)$ is shown to define a positive definite quadratic form. The variables

$$v_n = \begin{cases} u_o & \text{for} \quad n=0 \\ \\ 2u_n & \text{for} \quad n \neq 0 \end{cases} \tag{3.30}$$

are defined for $n=0,1,\ldots,P$ so that the first partial derivatives are

$$\begin{aligned} \hat{V}_n &= \partial I_W(\hat{H}f,g)/\partial v_n \\ &= \int_{-\pi}^{\pi} W(\theta)\{\hat{H}(\theta)\ f(\theta) - g(\theta)\}\ \cos(n\theta)\ d\theta/2\pi \end{aligned} \tag{3.31}$$

and the second mixed partial derivatives are

$$\begin{aligned} L_{nm} &= \partial \hat{V}_n/\partial v_m \\ &= \int_{-\pi}^{\pi} W(\theta)[g(\theta)]^2\ \cos(n\theta)\ \cos(m\theta)\ d\theta/2\pi \end{aligned} \tag{3.32}$$

Clearly, the quadratic form

$$\sum_{m,n=0}^{P} x_n \, x_m \, L_{nm} = \int_{-\pi}^{\pi} W(\theta)[g(\theta)]^2 \left[\sum_{n=0}^{P} x_n \, \cos(n\theta)\right]^2 d\theta/2\pi \tag{3.33}$$

is positive definite. This proves the following

Theorem 3-3. (Absence of False Solutions). Any stable AR(P) solution to Equations (3.19) is a local minimum.

Note that this does not eliminate the possibility of unstable solutions to Equations (3.19), nor does it establish the existence of a stable solution. Since the previous theorem has established the uniqueness of a minimum this theorem establishes the

Corrollary 3-1. Equations (3.19) can have at most one stable AR(P) solution. If such a solution exists it is the unique absolute minimum of $I_W(\hat{H}f,g)$ over $\mathscr{R}_P$.

Remarks

Three general formulations for estimating the parameters of an AR(P) process in noise have been discussed. The first "ideal" formulation has theoretical foundations resting upon principles of information theory as well as the maximum likelihood method. The second two formulations are developed as approximations to the first.

The need for approximate formulations arises due to the difficulty posed by the nonlinear equations resulting from the ideal formulation. The first approximate formulation

leads to the Yule-Walker equations but with modified corre-
lation values; algorithms for solving the Yule-Walker equa-
tions are computationally simple and well understood while
methods for evaluating the modified correlation values have
been carefully studied in recent years.

While this first, noise filtering, approach has led to
demonstrable performance improvements in noise environments
over the standard noise free formulation (and reduces to the
noise free formulation in noise free environments), still
better performance is desired. Rather than attempt direct
solution of the ideal formulation the second approximate
formulation is developed. Evidence that this weighted in-
formation formulation leads to improved performance over the
noise filtering formulation is presented in Chapter V;
neither approximate formulation is expected to perform as
well as the "ideal" formulation.

The weighted information formulation is related to
other techniques that have appeared in the literature.
Consider the situation wherein the desired signal spectrum
is essentially zero outside the region $\theta \in [-\pi/\mathfrak{D}, \pi/\mathfrak{D})$ while
the noise spectrum is essentially zero inside this region.
The foregoing theory indicates that an appropriate selection
for the weight function is

$$
W(\theta) = \hat{H}(\theta) = \begin{cases} 1 & \theta \in [-\pi/\mathfrak{D}, \pi/\mathfrak{D}) \\ 0 & \text{otherwise} \end{cases} \qquad (3.34)
$$

so that the weighted information is

$$I_W(\hat{H}f,g) = \int_{-\pi/\mathfrak{D}}^{\pi/\mathfrak{D}} \{[f(\theta)/g(\theta)] - \ln[f(\theta)/g(\theta)] - 1\} \, d\theta/2\pi \quad (3.35)$$

With the change of variable $\tilde{\theta}/\mathfrak{D} = \theta$ this may be rewritten

$$I_W(\hat{H}f,g) = (1/\mathfrak{D}) \int_{-\pi}^{\pi} \{[f(\tilde{\theta}/\mathfrak{D})/g(\tilde{\theta}/\mathfrak{D})] - \ln[f(\tilde{\theta}/\mathfrak{D})/g(\tilde{\theta}/\mathfrak{D})] - 1\} \, d\tilde{\theta}/2\pi \quad (3.36)$$

Clearly the indication here is to low pass filter and decimate the observed signal before fitting the AR(P) model to the resulting data. This is precisely the technique employed by Quirk and Liu [45] to improve the resolution of AR(P) estimation in noise; they considered the use of AR(P) estimators to determine the frequencies of sinusoids in noise and demonstrated that the filtering/decimation scheme is clearly advantageous when the sinusoids are a priori known to lie in some fixed frequency range.

The problem which motivates the present work concerns signal and noise spectra that are both generally nonzero throughout the entire frequency range, $[-\pi, \pi)$; hence the luxury of simple filtering/decimation schemes is not permitted. On the other hand, the difficulties associated with very limited quantities of data are not the primary focus of this work so that the asymptotic formulation is considered adequate.

Computational issues for the weighted information formulation are discussed in Chapter IV. Equations (3.19) are cast in algebraic form and their (exact) analytical solution is discussed. Approximate (numerical) solution methods might be developed based upon the resulting analytical system of equations or directly upon minimization of $I_W(\hat{H}f,g)$; the latter approach is adopted to develop a simple iterative procedure based upon the notion of a contraction mapping. In addition, computational procedures appropriate to the use of the weighted information for vector quantization are discussed. Since in many applications the "vector quantization codebook" may be designed "off-line" using noise free speech data, questions associated with the codebook design problem are not discussed; instead, computational procedures for the "on-line" minimization of $I_W(\hat{H}f,g)$ over the finite codebook are developed.

CHAPTER IV

COMPUTATIONAL FORMULATION

In this chapter computational procedures for the solution of Equations (3.19) are discussed. In the first section the system is reduced to an algebraic form by assuming the weight function to take the form of an AR(M) power spectral density; once cast as a nonlinear algebraic system of equations, analytic procedures for solving the system are discussed. In the second section, techniques for evaluating the coefficients of the system are discussed.

Analytic solution of the nonlinear algebraic system becomes increasingly difficult as the order of the weight function, M, is increased. While numerical polynomial root solving procedures could be systematically applied, the third section develops instead an iterative procedure based upon the idea of a contraction mapping. Together with sampled frequency domain processing techniques, these iterative procedures do not restrict the weight function to an all-pole form. The fourth section develops computational formulae required for the use of the weighted information in vector quantization; an extension of Jensen's theorem is developed to permit closed form evaluation of the appropriate integrals when the weight function assumes an

AR(M) form. Finally, the last section concludes this chapter with some final remarks concerning these computational methods.

Reduction to Algebraic Form

Let

$$\rho_n = \int_{-\pi}^{\pi} W(\theta) \; \hat{H}(\theta) \; f(\theta) \; e^{in\theta} \; d\theta/2\pi \tag{4.1}$$

$$n = 0,1,\ldots,P$$

denote the coefficients appearing on the left hand side of Equations (3.19). Let

$$\hat{\rho}_n = \int_{-\pi}^{\pi} W(\theta) \; g(\theta) \; e^{in\theta} \; d\theta/2\pi \tag{4.2}$$

$$n = 0,1,\ldots,P+M$$

denote the quantities appearing on the right hand side of Equations (3.19). Observe that the index of $\hat{\rho}_n$ is permitted to range beyond P to P+M. If $W(\theta)$ is an AR(M) spectrum given by

$$W(\theta) = \sigma_W^2/[B_M(e^{i\theta}) \; B_M(e^{-i\theta})] \tag{4.3a}$$

where

$$B_M(z) = \sum_{m=0}^{M} b_m \, z^{-m}; \quad b_0 = 1 \tag{4.3b}$$

and if $g(\theta)$ is an AR(P) spectrum given by Equations (2.3) and (2.4) then their product is an AR(P+M) spectrum given by

$$W(\theta) \, g(\theta) = \sigma_W^2 \, \sigma^2 / [C_{P+M}(e^{i\theta}) \, C_{P+M}(e^{-i\theta})] \tag{4.4a}$$

where

$$C_{P+M}(z) = A_P(z) \, B_M(z) = \sum_{m=0}^{P+M} c_m z^{-m}; \quad c_0 = 1 \tag{4.4b}$$

The quantities defined by Equation (4.2) are related to the polynomial coefficients in Equation (4.4b) by the Yule-Walker equations

$$\begin{bmatrix} \hat{\rho}_0 & \hat{\rho}_1 & \cdots & \hat{\rho}_{P+M} \\ \hat{\rho}_1 & \hat{\rho}_0 & \cdots & \hat{\rho}_{P+M-1} \\ & & & \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \hat{\rho}_{P+M} & \hat{\rho}_{P+M-1} & \cdots & \hat{\rho}_0 \end{bmatrix} \begin{bmatrix} 1 \\ c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_{P+M} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \sigma_W^2 \sigma^2 \tag{4.5}$$

Equations (3.19) assign numerical values to some of the entries in the coefficient matrix according to

$$\hat{\rho}_n = \rho_n \; ; \quad n = 0, 1, \ldots P \tag{4.6}$$

while the remaining entries are to be considered as unknowns. The elements of the column vector are defined as a linear combination of the coefficients of the unknown polynomial, $A_P(z)$, by Equation (4.4b) which may be rewritten in matrix form as

$$\begin{bmatrix} 1 \\ c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_{P+M} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ b_1 & 1 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & b_M \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_P \end{bmatrix} \qquad (4.7)$$

Equations (4.5), (4.6), and (4.7) define a nonlinear system of P+M+1 multivariate polynomials in the P+M+1 unknowns σ, a_1, a_2, $\ldots$, a_P, $\hat{\rho}_{P+1}$, $\hat{\rho}_{P+2}$, $\ldots$, $\hat{\rho}_{P+M}$. Each polynomial is a first order function of each unknown while each term in these polynomials may involve up to two distinct unknowns. The properties of the weighted information developed in Chapter III indicate that this system of equations can have at most one stable solution; if a stable solution exists it is the solution sought.

Assuming the AR(M) weight function to be stable the product polynomial, $C_{P+M}(z)$, also has all its roots inside the unit circle and may be expressed recursively in terms of a set of reflection coefficients according to

$$C_n(z) = C_{n-1}(z) + k_n \, z^{-n} \, C_{n-1}(z^{-1}); \quad C_0(z) = 1 \qquad (4.8)$$

for n = 1,2,...,P+M. If the coefficient matrix in Equation (4.5) were entirely known then the Levinson-Durbin recursion[1] could be applied to yield $C_{P+M}(z)$. Since some of the entries in the coefficient matrix are unknown, the Levinson-Durbin recursion cannot proceed beyond the determination of $C_P(z)$; the remaining reflection coefficients $\{k_{P+1}, k_{P+2}, \ldots, k_{P+M}\}$ are unspecified (beyond the stability requirement that $|k_n| < 1$) by Equations (4.5) and may be considered as new unknowns replacing $\{\hat{\rho}_{P+1}, \hat{\rho}_{P+2}, \ldots, \hat{\rho}_{P+M}\}$.

These remaining reflection coefficients should be selected so that $C_{P+M}(z) = 0$ modulo $B_M(z)$. Once these have been determined the solution may be obtained by simple polynomial division from

$$A_P(z) = C_{P+M}(z)/B_M(z) \tag{4.9}$$

together with

$$\sigma^2 = (\rho_0/\sigma_W^2) \prod_{n=1}^{P+M} (1-k_n^2) \tag{4.10}$$

To determine the remaining reflection coefficients it is generally simpler to consider the polynomials

---

[1]This well-known algorithm may be found in many fairly recent publications; for example, see [21, p. 55ff]. An exposition by the authors is contained in [57] and [58].

$$\widetilde{C}_{P+M}(z) = z^{-(P+M)} C_{P+M}(z^{-1}) \qquad (4.11)$$

$$\widetilde{B}_{M}(z) = z^{-M} B_{M}(z^{-1}) \qquad (4.12)$$

so that the condition to be satisfied is

$$\widetilde{C}_{P+M}(z) = 0 \text{ modulo } \widetilde{B}_{M}(z) \qquad (4.13)$$

Modulo reduction is then accomplished more simply by re-peated use of the substitution

$$z^{-M} = - \sum_{\ell=0}^{M-1} b_{M-\ell} \; z^{-\ell} \qquad (4.14)$$

in $\widetilde{C}_{P+M}(z)$ until all powers of z^{-1} larger than M-1 have been eliminated. The reduction process is facilitated by using the recursion (4.8) to express $\widetilde{C}_{P+M}(z)$ as

$$\widetilde{C}_{P+M}(z) = \widetilde{C}_{P}(z) \; \widetilde{E}_{M}(z) + z^{-P} \; \widetilde{C}_{P}(z^{-1}) \; \widetilde{F}_{M}(z) \qquad (4.15)$$

where

$$\widetilde{E}_{n}(z) = z^{-1} \; \widetilde{E}_{n-1}(z) + k_{P+n} \; z^{-(n-1)} \; \widetilde{F}_{n-1}(z^{-1}); \; \widetilde{E}_{0}(z) = 1 \qquad (4.16a)$$

$$\widetilde{F}_{n}(z) = z^{-1} \; \widetilde{F}_{n-1}(z) + k_{P+n} \; z^{-(n-1)} \; \widetilde{E}_{n-1}(z^{-1}); \; \widetilde{F}_{0}(z) = 0 \qquad (4.16b)$$

and

$$\widetilde{C}_P(z) = z^{-P} C_P(z^{-1}) \qquad\qquad (4.17)$$

With these formulae the reduction is accomplished in part by determining

$$\widetilde{D}_{M-1}(z) = \widetilde{C}_P(z) \bmod \widetilde{B}_M(z) \qquad\qquad (4.18a)$$

and

$$D_{M-1}(z) = C_P(z) \bmod \widetilde{B}_M(z) = z^{-P} \widetilde{C}_P(z^{-1}) \bmod \widetilde{B}_M(z) \qquad (4.18b)$$

The condition to be satisfied is then

$$\widetilde{D}_{M-1}(z) \widetilde{E}_M(z) + D_{M-1}(z) \widetilde{F}_M(z) = 0 \bmod \widetilde{B}_M(z) \qquad\qquad (4.19)$$

Modulo reduction of the left-hand side of Equation (4.19) leads to an M-1st order polynomial whose M coefficients must be equated to zero; this yields a system of M nonlinear polynomial equations in the M unknowns $\{k_{P+1}, k_{P+2}, \ldots, k_{P+M}\}$. While these equations are nonlinear some reflection will reveal that each polynomial equation is linear (i.e., of first degree) in each of the unknowns; the nonlinearity enters by way of terms involving products of different unknowns.

Because of this structure, systematic algebraic elimination[2] will yield an Mth order polynomial in a single unknown; each acceptable root of this polynomial will yield an $M-1^{st}$ order polynomial in a second unknown. Continuing in this fashion one successively solves M^{th}, $M-1^{st}$, ... order polynomial equations possibly generating M factorial potential solutions of which at most one satisfies the stability criteria. This method is feasible for small values of M (e.g. $M \leq 4$) but for larger values of M one must generally resort to numerical polynomial root solving procedures.[3]

For the case $M=2$, let

$$\widetilde{D}_1(z) = \widetilde{d}_o + \widetilde{d}_1 \, z^{-1} \tag{4.20a}$$

$$D_1(z) = d_o + d_1 \, z^{-1} \tag{4.20b}$$

---

[2] Several methods (such as those due to Euler, Bezout, or Sylvester) are available; one should take care not to introduce extraneous roots. For a general discussion see [59, Vol. II, p. 70ff] or [60, p. 277ff].

[3] The recommendation that M not exceed four is made based upon the fact that general polynomial equations of degree five and higher cannot be solved algebraically [59, Vol. II, p. 286]. Of course this does not eliminate the possibility of transcendental solutions [59, Vol. I, p. 274] or the possibility that some special structure, unrecognized by the present author, may be discovered (or imposed) to aid in the solution.

and let

$$G_3(z) = \widetilde{D}_1(z)\,\widetilde{E}_2(z) + D_1(z)\,\widetilde{F}_2(z) = \sum_{m=0}^{3} g_m\, z^{-m} \qquad (4.21)$$

denote the left hand side of Equation (4.19). Using Equations (4.16) these coefficients are

$$g_0 = d_0\, k_{P+2} \qquad\qquad (4.22a)$$

$$g_1 = \widetilde{d}_0\, k_{P+1}\, k_{P+2} + d_0\, k_{P+1} + d_1\, k_{P+2} \qquad\qquad (4.22b)$$

$$g_2 = \widetilde{d}_0 + \widetilde{d}_1\, k_{P+1}\, k_{P+2} + d_1\, k_{P+1} \qquad\qquad (4.22c)$$

$$g_3 = \widetilde{d}_1 \qquad\qquad (4.22d)$$

while modulo reduction yields

$$g_0 - b_2\, g_2 + b_1\, b_2\, g_3 = 0 \qquad\qquad (4.23a)$$

$$g_1 - b_1\, g_2 + (b_1^2 - b_2)\, g_3 = 0 \qquad\qquad (4.23b)$$

Expanding Equations (4.23) yields

$$p_0\, k_{P+2} + p_1 = k_{P+1}(q_0\, k_{P+2} + q_1) \qquad\qquad (4.24a)$$

$$\hat{p}_0\, k_{P+2} + \hat{p}_1 = k_{P+1}(\hat{q}_0\, k_{P+2} + \hat{q}_1) \qquad\qquad (4.24b)$$

where

$$p_o = d_o \tag{4.25a}$$

$$p_1 = \tilde{d}_1 \, b_1 \, b_2 - \tilde{d}_o \, b_2 \tag{4.25b}$$

$$\hat{p}_o = d_1 \tag{4.25c}$$

$$\hat{p}_1 = \tilde{d}_1 (b_1^2 - b_2) - \tilde{d}_o \, b_1 \tag{4.25d}$$

$$q_o = \tilde{d}_1 \, b_2 \tag{4.25e}$$

$$q_1 = d_1 \, b_2 \tag{4.25f}$$

$$\hat{q}_o = \tilde{d}_1 \, b_1 - \tilde{d}_o \tag{4.25g}$$

$$\hat{q}_1 = d_1 \, b_1 - d_o \tag{4.25h}$$

So that the solutions are given, upon elimination, by

$$k_{P+1} = (p_o \, k_{P+2} + p_1)/(q_o \, k_{P+2} + q_1) \tag{4.26}$$

and

$$k_{P+2} = \left[-s_1 \pm \sqrt{s_1^2 - 4s_o \, s_2} \right] / 2s_2 \tag{4.27}$$

where

$$p_o = d_o \tag{4.25a}$$

$$p_1 = \tilde{d}_1 \, b_1 \, b_2 - \tilde{d}_o \, b_2 \tag{4.25b}$$

$$\hat{p}_o = d_1 \tag{4.25c}$$

$$\hat{p}_1 = \tilde{d}_1 (b_1^2 - b_2) - \tilde{d}_o \, b_1 \tag{4.25d}$$

$$q_o = \tilde{d}_1 \, b_2 \tag{4.25e}$$

$$q_1 = d_1 \, b_2 \tag{4.25f}$$

$$\hat{q}_o = \tilde{d}_1 \, b_1 - \tilde{d}_o \tag{4.25g}$$

$$\hat{q}_1 = d_1 \, b_1 - d_o \tag{4.25h}$$

So that the solutions are given, upon elimination, by

$$k_{P+1} = (p_o \, k_{P+2} + p_1)/(q_o \, k_{P+2} + q_1) \tag{4.26}$$

and

$$k_{P+2} = \left[-s_1 \pm \sqrt{s_1^2 - 4s_o \, s_2} \right] /2s_2 \tag{4.27}$$

By simply cascading two of these filters a new filter is created whose power spectral response is $\hat{H}(\theta)\,\hat{H}(\theta)$. The coefficients in Equation (4.1) may then be computed in the usual manner (lag products of the windowed data) from the output of the cascaded filter structure. This scheme, depicted in Figure 2, assumes the relationship expressed by Equation (3.20) although this relationship may generally be avoided by replacing one of the filters in the cascade by a filter with $W(\theta)$ as its power spectral response. For each data window, a "snapshot" of the impulse response of the FIR filter could be used to estimate the parameters of $W(\theta)$. Since the response of the FIR filter may differ slightly from the response of the weight function a somewhat more consistent procedure would use the weight function parameters to implement an infinite impulse response (IIR) filter as the second filter in the cascade.

Frequency domain noise filtering methods generally provide greater flexibility in response function selection than is available with time domain methods. These methods involve an explicit transformation to the frequency domain, often by using the discrete Fourier transform (DFT), and determine the multiplicative response function, $\hat{H}(\theta)$, in sampled form using a formula such as Equation (2.19) or (2.20). The sampled form of $\hat{H}(\theta)$ may be used to estimate the parameters of $W(\theta)$. If the noise filtered signal is not required, frequency samples of the weight function may be used multiplicatively before evaluating the coefficients;
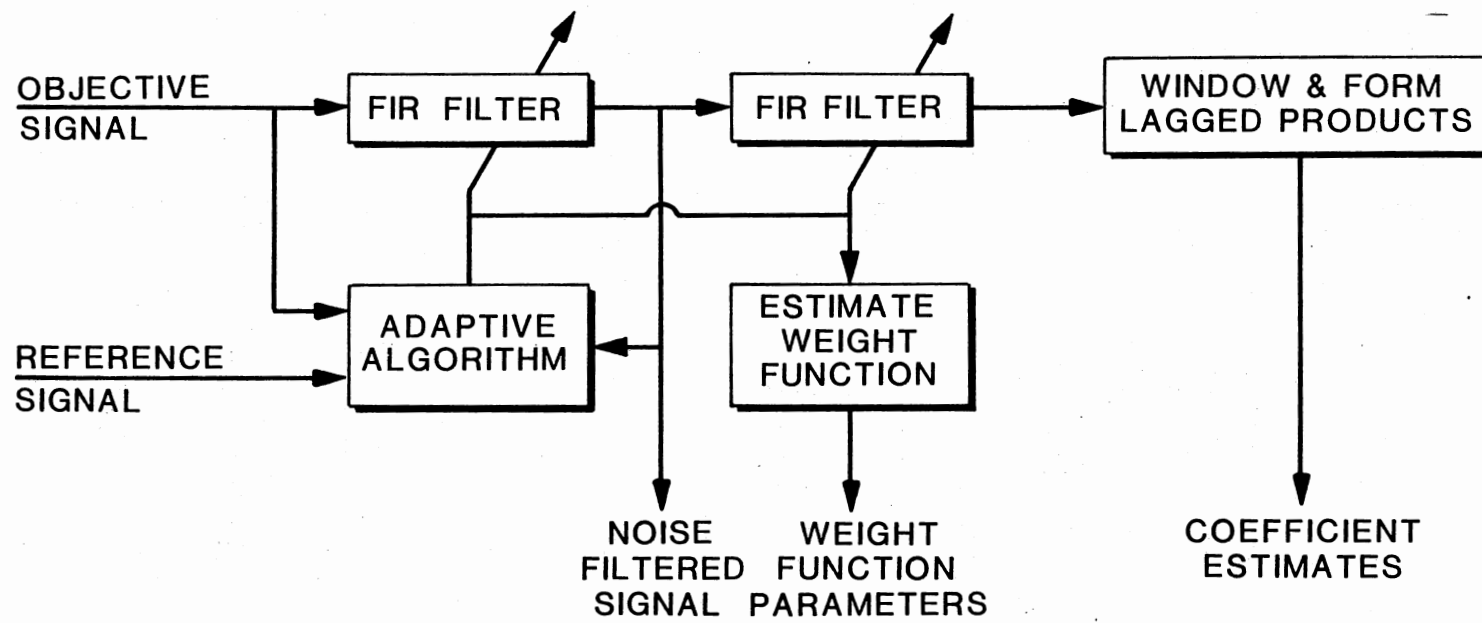
Figure 2. Time Domain Coefficient Evaluation

alternatively, one may avoid re-evaluating the weight func-
tion and simply apply $\hat{H}(\theta)$ twice. This latter alternative
is depicted in Figure 3.

A mixed time-frequency domain method is employed to
obtain some of the results presented in Chapter V. In this
method a Hamming window is applied to the observed data
which is then zero-extended before computing the DFT. A
sampled noise spectrum estimate is used together with these
transform values to compute a noise filter spectral re-
sponse, $\hat{H}(\theta)$, according to Equations (2.19) or (2.20).[4]
This frequency sampled noise filter response is applied
multiplicatively to the transform values and an inverse DFT
of these modified transform values (with their original
phase values) is computed. A random phase characteristic is
computed and introduced to the frequency sampled noise
filter spectral response which is inverse transformed to
obtain an impulse response characteristic. Standard (auto-
correlation method) LP analysis is applied to this impulse
response characteristic to determine the parameters of the
weight function. These parameters are used to implement a

---

[4]It is generally found to be useful to modify the fre-
quency response characteristic slightly by smoothing the
response obtained from (2.19) or (2.20) across frequency.
The smoother should eliminate features narrower than those
expected in the final signal spectrum while retaining
broader features; a recursive median filter with a total
length of about 2.5% of the single-sided bandwidth is a
current favorite of this author. End conditions (near the
DC and Nyquist frequencies) can be properly handled using
the known periodic nature of the frequency response.
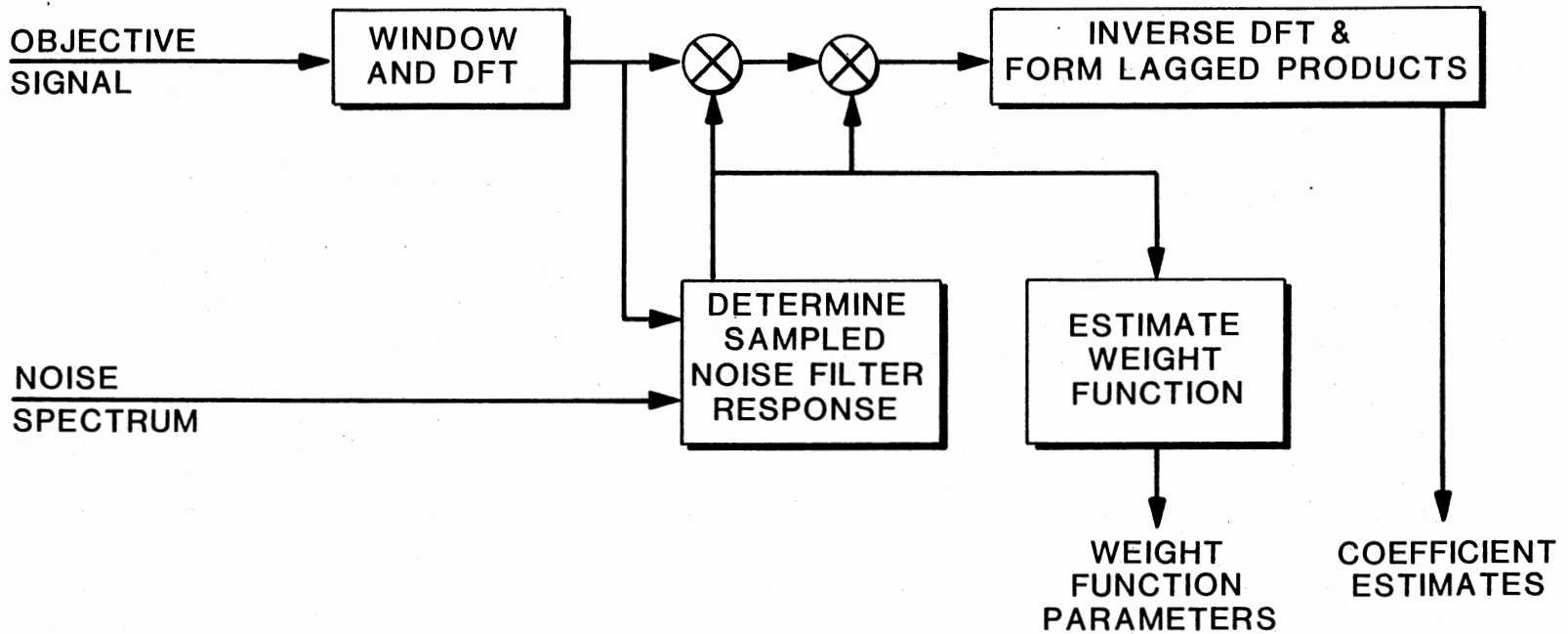
Figure 3.   Frequency Domain Coefficient Evaluation

(lattice structure) filter; beginning in the all-zero state the noise filtered (inverse transformed) data values are passed through this filter which is then permitted to "ring" awhile.[5] Lag products computed from this output then provide the required coefficient estimates; the overall procedure is depicted in Figure 4.

Finally, it is worth mentioning that each of these methods has recommended computing the final coefficient estimates as lagged products. The reason for this is that various quantization effects may occur up to the point of obtaining the modified data samples; however, if full precision is maintained in the final lag product computations, the resulting coefficient estimates will define a positive definite symmetric Töeplitz quadratic form in all but a very few highly exceptional cases (such as all modified data samples being identically zero).

Iterative Techniques

Equations (3.19) may be solved when the weight function has an AR(M) form by using the algebraic procedures described in the first section of this chapter; this method is appropriate if $M \leq 4$. Unfortunately, it is expected that accurate estimation of speech spectra will require weight functions with greater variation than is possible with an

---

[5]That is to say that a zero input is applied to the filter after all the noise filtered data values have been applied as input.
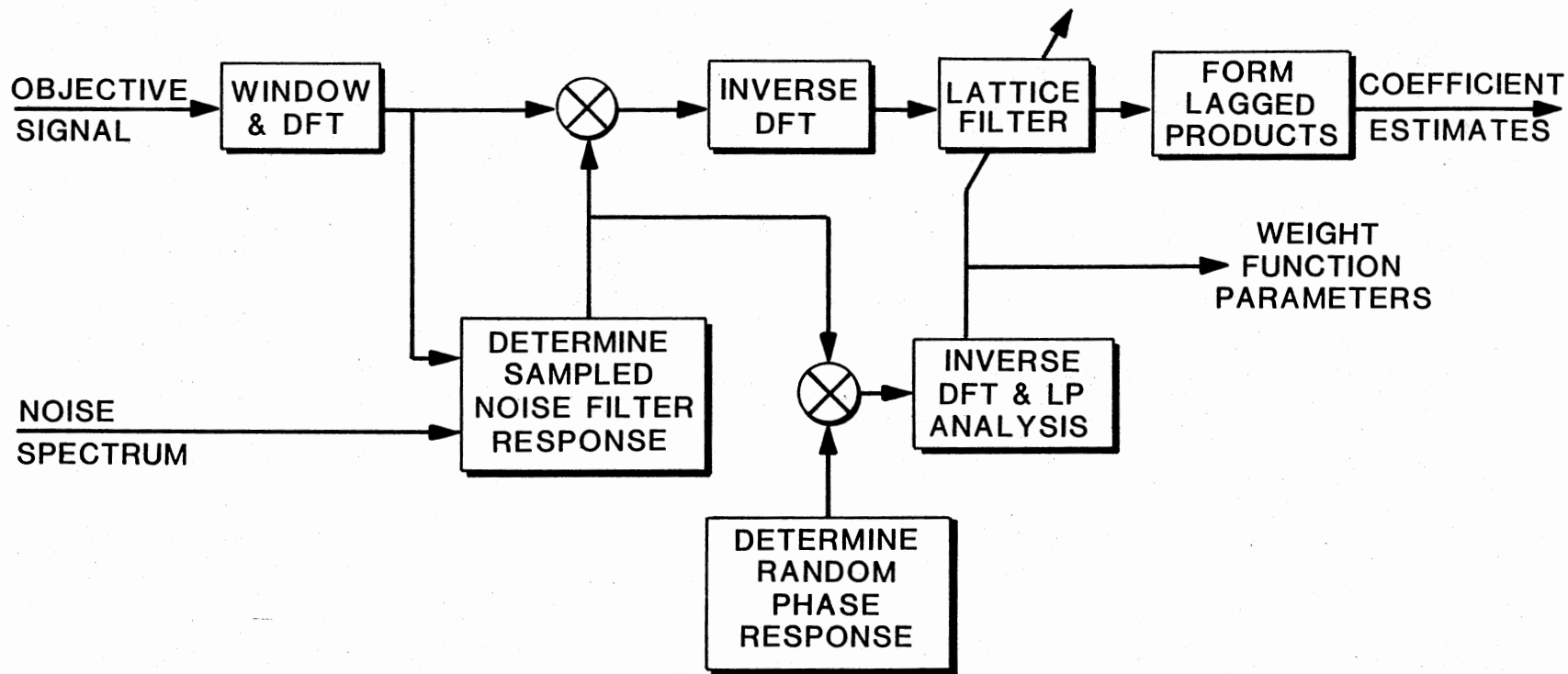
Figure 4.   Mixed Time-Frequency Domain Coefficient Evaluation.

AR(4) form.   The procedures of the first section might be extended by applying numerical polynomial root solving procedures when M becomes large but at present such an approach appears somewhat cumbersome.[6]   In this section alternate numerical formulations are discussed that do not make specific (parametric) assumptions as to the form of the weight function; these techniques are iterative and based upon the notion of a contraction mapping.   A good general reference for this section is Collatz [61].

Most (single-step) iterative procedures can be expressed in the form[7]

$$\bar{v}^{(n+1)} = \bar{\varphi}(\bar{v}^{(n)})$$
(4.29)

---

[6]For the reader wishing to pursue this approach it is worth noting that one stumbling block is that the previous uniqueness theorem has not eliminated the possibility of an unstable (or imaginary) solution to Equations (4.5), (4.6), and (4.8) for which some (but not all) of the reflection coefficients are real and in the interval (-1, 1).   If one could devise a method which guarantees that only the solution sought has real parameters isolated in (-1, 1), or some other known interval, the development of a numerical algorithm would be greatly facilitated.   The reader is referred to [60, p. 99ff] or any similar discussion of numerical methods for determining real roots of polynomials.

[7]Parenthesized superscripts shall denote instances of the parameter vector while subscripts shall denote components of the parameter vector.

where $\bar{v}^{(n)}$ is the nth iterate of the parameter vector $\bar{v}$. The solution sought is a fixed point of the map $\bar{\varphi}$. If $\bar{\varphi}$ satisfies a Lipschitz condition[8]

$$\| \bar{\varphi}(\bar{v}^{(1)}) - \bar{\varphi}(\bar{v}^{(2)}) \| \leq \mathscr{L} \| \bar{v}^{(1)} - \bar{v}^{(2)} \| \qquad (4.30)$$

for some $0 < \mathscr{L} < 1$ then $\bar{\varphi}$ is said to be a contraction map. Contraction maps are often used to prove existence theorems because the sequence of iterates generated by (4.29) is Cauchy.

The problem of designing an iterative procedure for solving a system of equations can be viewed as the problem of finding a contraction map whose fixed points coincide with the solutions sought. One usually begins with a map having the appropriate fixed points and then tries to show it satisfies a Lipschitz condition; often one employs the mean value theorem which states that if φ_n is a continuously differentiable function of the parameter vector $\bar{v}$ then[9]

---

[8]The map $\bar{\varphi}$ is assumed to have its domain in a Banach space with norm $\| \cdot \|$ and its range contained by the domain.

[9]Two notational conventions are introduced here. First $\varphi_{n/\ell}$ denotes $\partial \varphi_n / \partial v_\ell$ and second the Einstein summation convention (with respect to repeated subscripts) is employed. The summation range is $0, 1, \ldots, P$ so that the Einstein convention implies summation with respect to the subscript ℓ (only) over this range on the right hand side of (4.31). These conventions are used in this section only.

$$\varphi_n(\bar{v}^{(1)}) - \varphi_n(\bar{v}^{(2)}) = \varphi_{n/\ell} (\gamma \bar{v}^{(1)} + [1-\gamma] \bar{v}^{(2)}) \{v_\ell^{(1)} - v_\ell^{(2)}\}$$

$$(4.31)$$

for some $0 \leq \gamma \leq 1$. If one can determine a constant $\mathscr{L} < 1$ majorizing the norm of the matrix with components $\varphi_{n/\ell}$ then $\bar{\varphi}$ has been demonstrated to satisfy a Lipschitz condition.

Using Equations (3.22), (3.30), (3.31), (3.32) and (4.1) the system of Equations (3.19) may be expressed as

$$\hat{V}_n = 0 \; ; \quad n = 0,1,\ldots,P \qquad\qquad (4.32)$$

where

$$\hat{V}_n = \rho_n - L_{nm} v_m \qquad\qquad (4.33)$$

Defining

$$L_{nm\ell} = \int_{-\pi}^{\pi} W(\theta)[g(\theta)]^3 \cos(n\theta) \cos(m\theta) \cos(\ell\theta) \, d\theta/2\pi \quad (4.34)$$

and

$$\delta_{nm} = \begin{cases} 0 & n \neq m \\ \\ 1 & n = m \end{cases} \qquad\qquad (4.35)$$

the following relations may be easily verified

$$L_{nm/\ell} = -2\ L_{nm\ell} \tag{4.36}$$

$$\hat{V}_{n/\ell} = -\ L_{nm/\ell}\ v_m - L_{nm}\ v_{m/\ell}$$

$$= 2\ L_{nm\ell}\ v_m - L_{nm}\ \delta_{m\ell}$$

$$= 2\ L_{n\ell} - L_{n\ell} = L_{n\ell} \tag{4.37}$$

Consider the map $\bar{\varphi}$ with components[10]

$$\varphi_n = v_n - \lambda\ [L_o^{-1}]_{nm}\ \hat{V}_m \tag{4.38}$$

where λ is a nonzero scalar constant. Use of this map for an iterative procedure is essentially a modified Newton method. First observe that $\bar{\varphi}$ has a fixed point if and only if the second term on the right hand side of (4.38) vanishes. This term vanishes if and only if Equations (4.32) are satisfied since, as shown in Chapter III, L (and so also L^{-1} and L_o^{-1}) is positive definite.

---

[10]If L denotes the matrix with entries L_{nm} and L^{-1} the inverse of this matrix then L_o^{-1} shall denote L^{-1} evaluated at the initial iterate $\bar{v}^{(0)}$ and $[L_o^{-1}]_{nm}$ its entries.

Next, using (4.37), consider

$$\varphi_{n/\ell} = v_{n/\ell} - \lambda [L_0^{-1}]_{nm} \hat{V}_{m/\ell}$$

$$= \delta_{n\ell} - \lambda [L_0^{-1}]_{nm} L_{m\ell} \qquad (4.39)$$

which, if evaluated at $\bar{v} = \bar{v}^{(0)}$, is

$$\varphi_{n/\ell}^{(0)} = (1-\lambda) \delta_{n\ell} \qquad (4.40)$$

Clearly, (4.40) is majorized by $\mathscr{L} = |1-\lambda|$ so that λ should be selected in the range $0 < \lambda < 2$ if the Lipschitz condition is to be satisfied. More generally, since the last term in (4.39) is positive definite, λ should be selected in the range $0 < \lambda < 2/\lambda_{max}$ where

$$\lambda_{max} \geq \sup_{\|q\|=1} q_n [L_0^{-1}]_{nm} L_{m\ell} q_\ell \qquad (4.41)$$

bounds the matrix norm. With this selection

$$\inf_{\|q\|=1} q_n \varphi_{n/\ell} q_\ell = 1 - \lambda \sup_{\|q\|=1} q_n [L_0^{-1}]_{nm} L_{m\ell} q_\ell$$

$$\geq 1 - \lambda \lambda_{max} > -1 \qquad (4.42)$$

and the matrix norm of $\varphi_{n/\ell}$ is bounded by one.

Apparently the choice $\lambda = 1/\lambda_{max}$ would lead to the most rapid convergence while smaller values would lead to slower

convergence and guarantee that $\varphi_{n/\ell}$ is positive definite. Unfortunately, the right hand side of inequality (4.41) is a function of the parameter vector $\bar{v}$ and cannot be bounded by a constant, λ_{max}, for all $\bar{v}$ in $\mathscr{R}_P$; consequently the Lipschitz condition cannot be satisfied everywhere in $\mathscr{R}_P$.

If a solution, $g_*(\theta)$, exists in $\mathscr{R}_P$ it is possible to find a constant G_{max} sufficiently large such that

$$g_*(\theta) \leq G_{max} \tag{4.43}$$

for all $\theta \epsilon [-\pi, \pi)$. For such a constant the solution will be contained in that portion of $\mathscr{R}_P$ for which

$$g(\theta) \leq G_{max} \tag{4.44}$$

for all $\theta \epsilon [-\pi, \pi)$. Then from

$$\sup_{\|q\|=1} q_n [L_o^{-1}]_{nm} L_{m\ell} q_\ell$$

$$\leq \sup_{\|q\|=1} q_n L_{nm} q_m / \varsigma$$

$$\leq W_{max} G_{max}^2 / \varsigma \tag{4.45}$$

where

$$W(\theta) \leq W_{max} \tag{4.46}$$

for all $\theta \epsilon [-\pi, \pi)$ and

$$\mathfrak{S} = \inf_{\|q\|=1} q_n L_{nm}^{(0)} q_m > 0 \qquad (4.47)$$

it is clear that any choice

$$\lambda_{max} \geq W_{max} \, G_{max}^2 / \mathfrak{S} \qquad (4.48)$$

will suffice to satisfy the Lipschitz condition for that portion of $\mathscr{R}_p$.

To recapitulate, the map $\bar{\varphi}$, defined by (4.38), has fixed points coinciding with the solutions to (4.32). Moreover, if there exists a solution in $\mathscr{R}_p$ and the domain of $\bar{\varphi}$ is suitably restricted to a subset of $\mathscr{R}_p$ containing this solution then there exists $\lambda > 0$ sufficiently small such that $\bar{\varphi}$ satisfies a Lipschitz condition on this subset and (4.39) is positive definite. This implies that application of the map $\bar{\varphi}$ to any element of the subset will generate a new parameter vector closer (in norm) to the solution. Hopefully, repeated application of $\bar{\varphi}$ will generate a sequence of parameter vectors approaching the solution; this will be the case if each new parameter vector is also in the restricted domain of $\bar{\varphi}$.

Providing a guarantee that each new parameter vector will be within the restricted domain of $\bar{\varphi}$ is not a simple task. Without such a guarantee it is possible to devise a computational test to check for this condition; then, if the

test is violated, some method must be devised to restart the iterations. In practice the situation is not expected to be quite so pathological; if λ is selected to be conservatively small (smaller if the solution is expected to be a sharply peaked spectrum) and a reasonably good initial estimate is provided, one does not expect to encounter convergence difficulties. This more optimistic approach shall be taken in the following.

To implement the iterative procedure assume $W(\theta)$ is available in sampled form. The components of the nth iterate parameter vector may be used to evaluate

$$g_n(\theta) = 1/\{ \sum_{\ell=0}^{P} v_{\ell}^{(n)} \cos(\ell\theta) \} \tag{4.49}$$

in sampled form. If the sample mesh is equally spaced at

$$\theta_k = \pi k/N \; ; \quad k = -N,\ldots,0,1,\ldots,N-1 \tag{4.50}$$

then the components $\hat{V}_m^{(n)}$ may be computed from

$$\hat{V}_m^{(n)} = \rho_m - \sum_{k=-N}^{N-1} W(\theta_k) \, g_n(\theta_k) \, \cos(m\theta_k)/2N \tag{4.51}$$

and the components of the next iterate are provided by

$$v_{\ell}^{(n+1)} = v_{\ell}^{(n)} - \lambda [L_0^{-1}]_{\ell m} \, \hat{V}_m^{(n)} \tag{4.52}$$

A crude test that the nth iterate is in $\mathscr{R}_p$ is provided in the course of these computations by verifying that the denominator of (4.49) is positive on the sample mesh.

The procedure can be initialized by the solution to the Yule-Walker equations where the elements of the coefficient matrix are given by ρ_m. Equations (3.23) and (3.30) may then be used to evaluate $v_\ell^{(0)}$ while the elements $[L_o^{-1}]_{nm}$ may be obtained by inverting the real symmetric matrix with entries

$$[L_o]_{nm} = \sum_{k=-N}^{N-1} W(\theta_k)\,[g_o(\theta_k)]^2\,\cos(n\theta_k)\,\cos(m\theta_k)/2N \quad (4.53)$$

The coefficients ρ_m may be evaluated from

$$\rho_m = \sum_{k=-N}^{N-1} W(\theta_k)\,\hat{H}(\theta_k)\,f(\theta_k)\,\cos(m\theta_k)/2N \quad (4.54)$$

Alternatively, the computational methods described in the previous section may be employed to evaluate the ρ_m as lagged products of modified data values.

A simple test for iteration completion is to simply check that

$$\psi_n = \sum_{m=0}^{P} [\hat{v}_m^{(n)}]^2 \quad (4.55)$$

is less than some small preselected value. Finally, to obtain filter coefficients as are required by many

applications, it is perhaps simplest to first compute correlation values from

$$r_m = \sum_{k=-N}^{N-1} g_n(\theta_k) \cos(m\theta_k) \qquad (4.56)$$

and then solve the Yule-Walker equations.

If at some step prior to iteration completion an iterate falls outside $\mathcal{R}_P$, one may attempt to reinitialize the procedure using one of the last few iterates inside $\mathcal{R}_P$.

Formulae for Vector Quantization

In this section formulae relevant to the problem of minimizing $I_W(\hat{H}f,g)$ over a specified finite collection of AR(P) model spectra are developed. Consider first that according to Equation (3.24) this problem is equivalent to minimizing $I_W(g_1,g)$ where $g_1(\theta)$ is an AR(P) model spectrum satisfying Equation (3.19). Next, observe that minimizing $I_W(g_1,g)$ is equivalent to minimizing

$$J_W(g_1,g) = \int_{-\pi}^{\pi} [W(\theta) \; g_1(\theta)/g(\theta) + W(\theta) \ln g(\theta)] \; d\theta/2\pi \quad (4.57)$$

Since $g(\theta)$ is an AR(P) model given by Equation (3.22) the first term in Equation (4.57) may be rewritten as

$$\int_{-\pi}^{\pi} W(\theta) \; g_1(\theta)/g(\theta) \; d\theta/2\pi = \sum_{n=-P}^{P} u_{|n|} \; \rho_{|n|} \qquad (4.58a)$$

where the fact that $g_1(\theta)$ satisfies Equation (3.19) has been used together with Equation (4.1). Similarly, the second term in Equation (4.57) may be rewritten as

$$\int_{-\pi}^{\pi} W(\theta) \ln g(\theta) \, d\theta/2\pi = \ln(\sigma^2) \int_{-\pi}^{\pi} W(\theta) \, d\theta/2\pi$$

$$- \int_{-\pi}^{\pi} W(\theta) \ln[A_P(e^{i\theta}) A_P(e^{-i\theta})] \, d\theta/2\pi \qquad (4.58b)$$

In general $J_W(g_1, g)$ will be minimized over the finite collection of AR(P) spectra by evaluating this quantity for each model spectrum in the collection. For any given model spectrum the first term may be easily evaluated using (4.58a); the coefficients ρ_n may be determined from the data using one of the methods outlined in the second section of this chapter. The second term presents somewhat greater difficulty; when $W(\theta) = 1$ the last term in (4.58b) may be shown to vanish as a consequence of Jensen's theorem but, in general, this term will not vanish.

When $W(\theta)$ has an AR(M) form an extension of Jensen's theorem, which shall be developed presently, permits the evaluation of this term from a simple formula. In order to establish the general theorem it shall be necessary to first establish the following lemma.

Lemma 4-1. Let

$$A_P(z^{-1}) = \prod_{m=1}^{P} (1 - \eta_m z) \qquad (4.59)$$

have no roots inside the unit circle, Γ. If τ_k and τ_ℓ are also within the unit circle then

$$T_{\ell k} = \oint_\Gamma \{\ln A_P(z^{-1})\}/\{(z-\tau_k)(1-\tau_\ell z)\} \, dz/2\pi i$$

$$= \{\ln A_P(\tau_k^{-1})\}/(1-\tau_k \tau_\ell) \tag{4.60}$$

Proof. The method of proof is essentially the same as that used for Jensen's theorem by Hille [62, pp. 256-7]. Assume without loss of generality that a narrow strip from τ_k to $\nu_k = \tau_k/|\tau_k|$ is free of the η_m and consider the integral

$$\Xi_{\ell k} = \oint_\mathscr{C} \{\ln[(z-\tau_k)/(1-\tau_\ell z)]\} \, d[\ln A_P(z^{-1})]/2\pi i \tag{4.61}$$

around the contour, $\mathscr{C}$, depicted in Figure 5. The logarithm, determined so that $\ln(-1) = \pi i$, is analytic within $\mathscr{C}$ and $A_P(z^{-1})$ has neither poles nor zeros within $\mathscr{C}$ so $\Xi_{\ell k} = 0$. As the radius of the circular portion of the contour, $\mathscr{C}$, surrounding the singularity τ_k tends to zero it offers no contribution to this integral. As the distance between the two straight sections of the contour tends to zero they provide the contribution

$$\aleph_k = \int_{z=\nu_k}^{z=\tau_k} [A_P(z^{-1})]^{-1} \, d[A_P(z^{-1})]$$

$$= \ln A_P(\tau_k^{-1}) - \ln A_P(\nu_k^{-1}) \tag{4.62}$$
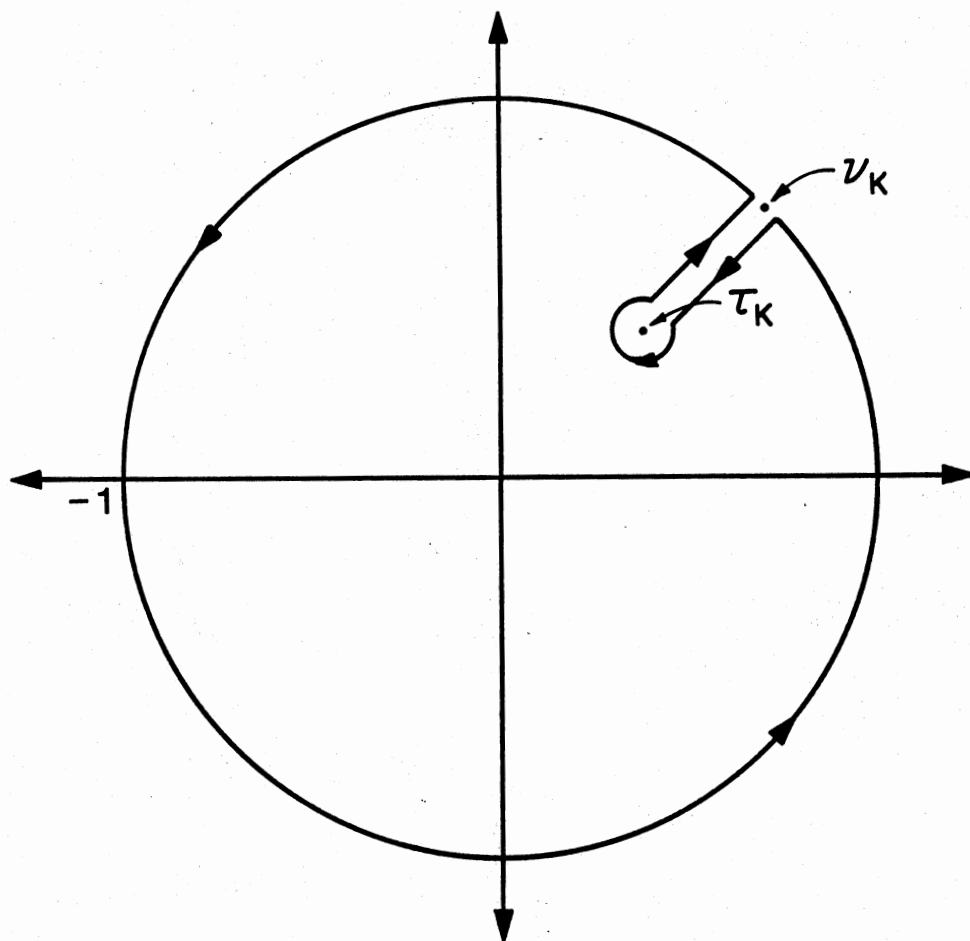
Figure 5.   The Contour $\mathscr{C}$ in the Complex Z-Plane

For the remaining portion of the contour, integration by parts yields

$$(1 - \tau_k \tau_\ell) \, T_{\ell k} = \aleph_k - \Xi_{\ell k} + \Delta_\Gamma / 2\pi i \qquad (4.63)$$

where the integrated part is

$$\Delta_\Gamma = \{\ln A_P(z^{-1}) \, \ln[(z-\tau_k)/(1-\tau_\ell z)]\} \, \Big|_{z \, = \, \nu_k}^{z \, = \, \nu_k^{(+)}}$$

$$= \ln A_P(\nu_k^{-1})\{2\pi i + \ln[\nu_k - \tau_k)/(1 - \tau_\ell \nu_k)]\}$$

$$-\ln A_P(\nu_k^{-1}) \, \{\ln[(\nu_k - \tau_k)/(1 - \tau_\ell \nu_k)]\}$$

$$= 2\pi i \, \ln A_P(\nu_k^{-1}) \qquad (4.64)$$

Substitution of (4.62) and (4.64) along with $\Xi_{\ell k} = 0$ into Equation (4.63) completes the proof.

A simple variable substitution may be used to obtain the related formula

$$T_{k\ell} = \oint_\Gamma \{\ln A_P(z)\}/\{(z-\tau_k)(1-\tau_\ell z)\} \, dz/2\pi i$$

$$= \{\ln A_P(\tau_\ell^{-1})\}/(1 - \tau_k \tau_\ell) \qquad (4.65)$$

which together with (4.60) establishes the

Corollary 4-1.

$$T_{\ell k} + T_{k\ell} = \oint_{\Gamma} \{\ln A_P(z) \, A_P(z^{-1})\}/\{(z-\tau_k)(1- \tau_\ell z)\} \, dz/2\pi i$$

$$= \{\ln A_P(\tau_k^{-1}) \, A_P(\tau_\ell^{-1})\}/(1- \tau_k \tau_\ell) \tag{4.66}$$

Finally, sufficient background has now been presented to establish the

Theorem 4-1. Let $W(\theta)$ have an AR(M) form given by

$$W(\theta) = \left| \Omega(e^{i\theta}) \right|^2 \tag{4.67}$$

where $\Omega(z)$ has the partial fraction expansion

$$\Omega(z) = \sum_{\ell=1}^{M} \omega_\ell/(1- \tau_\ell z^{-1}) = \sigma_W/B_M(z) \tag{4.68}$$

with $|\tau_\ell| < 1$. Then with $g(\theta)$ given by equation (2.3) the second term in (4.57) is

$$\int_{-\pi}^{\pi} W(\theta) \ln g(\theta) \, d\theta/2\pi = \ln \sigma^2 \int_{-\pi}^{\pi} W(\theta) \, d\theta/2\pi - T \tag{4.69}$$

where

$$T = 2 \sum_{k=1}^{M} \omega_k \, \Omega(\tau_k^{-1}) \ln A_P(\tau_k^{-1}) \tag{4.70}$$

Proof. Using (2.3), (4.67), and (4.68)

$$T = \int_{-\pi}^{\pi} \left| \Omega(e^{i\theta}) \right|^2 \ln \left| A_P(e^{i\theta}) \right|^2 d\theta/2\pi$$

$$= \sum_{k,\ell=1}^{M} \omega_k \omega_\ell \oint_{\Gamma} \{ \ln A_P(z) A_P(z^{-1}) \}/\{(z-\tau_k)(1-\tau_\ell z)\} \; dz/2\pi i \qquad (4.71)$$

Together with the above corollary this yields

$$T = \sum_{k,\ell=1}^{M} \{\omega_k \omega_\ell /(1-\tau_k \tau_\ell)\} \ln A_P(\tau_k^{-1}) A_P(\tau_\ell^{-1}) \qquad (4.72)$$

and (upon splitting the logarithm and collecting terms) Equation (4.70).

With $W(\theta) = 1$ this theorem yields

$$\int_{-\pi}^{\pi} \ln g(\theta) \; d\theta/2\pi = \ln \sigma^2 \qquad (4.73)$$

which is a special case of Jensen's theorem [62, Theorem 9.2.5]. The first term in Equation (4.69) is easy to compute while the second term, T, given by Equation (4.70) may offer the reader some difficulty. First observe that (4.70) requires knowledge of the parameters of the partial fraction expansion (4.68). These are fairly easy to determine once

the roots τ_k of $B_M(z)$ are known by recognizing that ω_k equals[11]

$$(z-\tau_k) \; z^{-P} \; \Omega(z) \; = \; \sigma_W(z-\tau_k)/\{z^P \; B_M(z)\} \qquad (4.74)$$

evaluated at $z = \tau_k$. Hence, the basic difficulty is that of determing the roots, τ_k.

Since extracting the roots of $B_M(z)$ can be a difficult problem for large values of M it is advantageous if $B_M(z)$ is already known as a product of low order factors. To accomplish this, recall that $B_M(z)$ is determined so that $W(\theta)$ approximates $\hat{H}(\theta)$. If $W(\theta)$ is a product of known AR(2) models

$$W(\theta) = W_1(\theta) \; W_2(\theta) \; \dots \; W_{M/2}(\theta) \qquad (4.75)$$

then $B_M(z)$ is easily known as a product of second order factors. In order to determine $W(\theta)$ in this manner one may first determine $W_1(\theta)$ to approximate $\hat{H}(\theta)$, then $W_2(\theta)$ to approximate $\hat{H}(\theta)/W_1(\theta)$, then $W_3(\theta)$ to approximate $\hat{H}(\theta)/[W_1(\theta) \; W_2(\theta)]$ and so on. To obtain the best overall approximation it is probably advantageous to develop some simple ad hoc method to force the approximation at each

---

[11]This assumes the roots, τ_k, are distinct. The formulae become mildly more complicated when this is not the case.

stage to fit no more than one strong resonance in the function being approximated.

Remarks

This chapter has explored computational procedures related to the weighted information estimation formulation developed in Chapter III; it is worth noting that the author does not consider any of these methods entirely satisfactory for all applications.

The first section employed an assumed AR(M) form for the weight function which enabled the problem to be cast in the form of a nonlinear system of polynomial equations. Solution of the system was found to be a relatively simple task for small values of M but one that becomes rapidly more complex as M is increased beyond four. As a general approach, the assumption of a parametric form for the weight function has considerable promise for the development of efficient computational methods; the basic difficulty is that of finding a clever parametrization which provides sufficient flexibility in the form of the weight function (for the given application) while leading to a simple and efficient computational algorithm.

The second section discussed the computation of various coefficients that arise within the computational formulae. Choice of a specific procedure will ultimately be influenced by the demands of the specific application; interdependant

factors to be considered include the quantity of data available, rounding/truncation effects, fixed/floating point representation format, algorithm structure, memory requirements, and computational speed. The coefficient evaluation procedures discussed are variants of methods proposed (and sometimes implemented) for real time speech analysis applications.

The third section discussed single-step iterative methods within the general framework provided by the notion of a contraction mapping. Multi-step methods were not discussed; in general, convergence characteristics are more difficult to prove for multi-step methods in spite of the fact that they tend to converge faster in practice.[12] These iterative methods offer significantly more flexibility in the form of the weight function[13] at the expense of a greater computational cost. The notion of a contraction map, sometimes employed for nonconstructive existence proofs, provides a useful general framework within which a

---

[12]Faster convergence, in terms of a reduced number of iterations, should not be confused with reduced computational cost. Each iteration of a multi-step method generally is more expensive computationally than a comparable single-step method so that a detailed analysis is usually required to compare costs.

[13]That is, compared to the parametric approach to weight function selection discussed in the first section. In this sense one might describe these methods with a seemingly contradictory phrase such as "nonparametric autoregressive estimation".

variety of iterative methods may be discussed; the specific method presented is a modified Newtonian iteration chosen as a tradeoff between simplicity and effectiveness. A possibly more effective iterative procedure would be a steepest descent method; generally such a procedure attempts to minimize a scalar function $U = U(\bar{v})$ by using a map with components

$$\varphi_n = v_n - \lambda \, \partial U / \partial v_n \qquad (4.76)$$

where the scalar function $\lambda = \lambda(\bar{v})$ is chosen to minimize $U(\bar{\varphi})$ at each iteration.

The fourth section considers the problem of minimizing $I_W(\hat{H}f, g)$ over a given finite collection of AR(P) models. The procedure involves the computation of a cost function for each model in the collection. The cost function involves two terms; the first term is evaluated quite simply (regardless of the form of the weight function) using formula (4.58a) which is identical to one arising in "standard" (unweighted) vector quantization. The second term is usually quite simple in "standard" vector quantization, see Equation (4.73), but becomes far more complex when the weighted information formulation is employed.

An extension of Jensen's theorem provides a formula which may be employed to evaluate this term when $W(\theta)$ has an AR(M) form; however, the reader is admonished to bear in

mind that it is probably far simpler to discretize this integral and evaluate it numerically as a sum of products from

$$\int_{-\pi}^{\pi} W(\theta) \ln g(\theta) \, d\theta/2\pi \;=\; \sum_{k=-N}^{N-1} W(\theta_k) \; \mathscr{G}_k \tag{4.77}$$

where

$$\mathscr{G}_k \;=\; [\ln g(\theta_k)]/(2N) \tag{4.78}$$

This has the additional advantage of not imposing an AR(M) form upon the weight function. More generally, $W(\theta)$ might be expressed as a sum of perhaps only a dozen nonnegative "shape functions" by

$$W(\theta) \;=\; \sum_k \xi_k \, \mathscr{W}_k(\theta) \tag{4.79}$$

so that, if the quantities

$$\mathscr{F}_k \;=\; \int_{-\pi}^{\pi} \mathscr{W}_k(\theta) \ln g(\theta) \, d\theta/2\pi \tag{4.80}$$

are precomputed for each AR model in the finite collection, the second term may be easily evaluated from

$$\int_{-\pi}^{\pi} W(\theta) \ln g(\theta) \, d\theta/2\pi \;=\; \sum_k \xi_k \, \mathscr{F}_k \tag{4.81}$$

CHAPTER V

RESULTS

In this chapter the weighted information estimation formulation is demonstrated to provide improved performance relative to the noise filtering formulation. It is worth noting that, although existence has not been proven in previous chapters, several thousand data frames have been analyzed using the weighted information formulation and not one counterexample has been encountered.

Gaussian Signals

In order to study the performance of the weighted information formulation pseudorandom sequences were generated. A zero-mean white Gaussian process was simulated using a congruential multiplicative random number generator; the resulting sequence of independent uniformly distributed samples was transformed to Gaussian form using the Box-Müller transformation followed by Central-Limit averaging.[1] Zero-mean AR(P) Gaussian processes were simulated

---

[1] In theory, the Box-Müller transformation is adequate. However, if the input deviates from a uniform distribution the output will, correspondingly, deviate from a Gaussian distribution; Central-Limit averaging will tend to reduce any such deviations from a Gaussian form.

by applying the simulated white Gaussian process to an all-pole (lattice structure) digital filter; the first few thousand output values from the filter were consistently ignored in order to avoid the transient response of the filter.

By adding two independent zero-mean Gaussian AR processes at a specified signal to noise ratio appropriate test data was produced. For many of the examples the "signal" process had an AR(2) spectrum defined by the reflection coefficient values

$$k_1 = -.8 \quad \text{and} \quad k_2 = -.9 \qquad (5.1)$$

This signal process spectrum, evaluated from these parameter values, is displayed in Figure 6a. While some examples employ a white Gaussian corrupting noise process, others employ an AR(2) process defined by the reflection coefficient values

$$k_1 = +.8 \quad \text{and} \quad k_2 = -.9 \qquad (5.2)$$

This "colored" noise process spectrum is displayed in Figure 6b.

As a basis for comparison, the standard autocorrelation analysis method was applied to 100 different 400 sample Hamming windowed frames of data from the uncorrupted signal process. Each resulting estimate is characterized by a pair

LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY)

*** *** AR(2) SIGNAL SPECTRUM; {K1,K2} = {-.8,-.9} *** ***
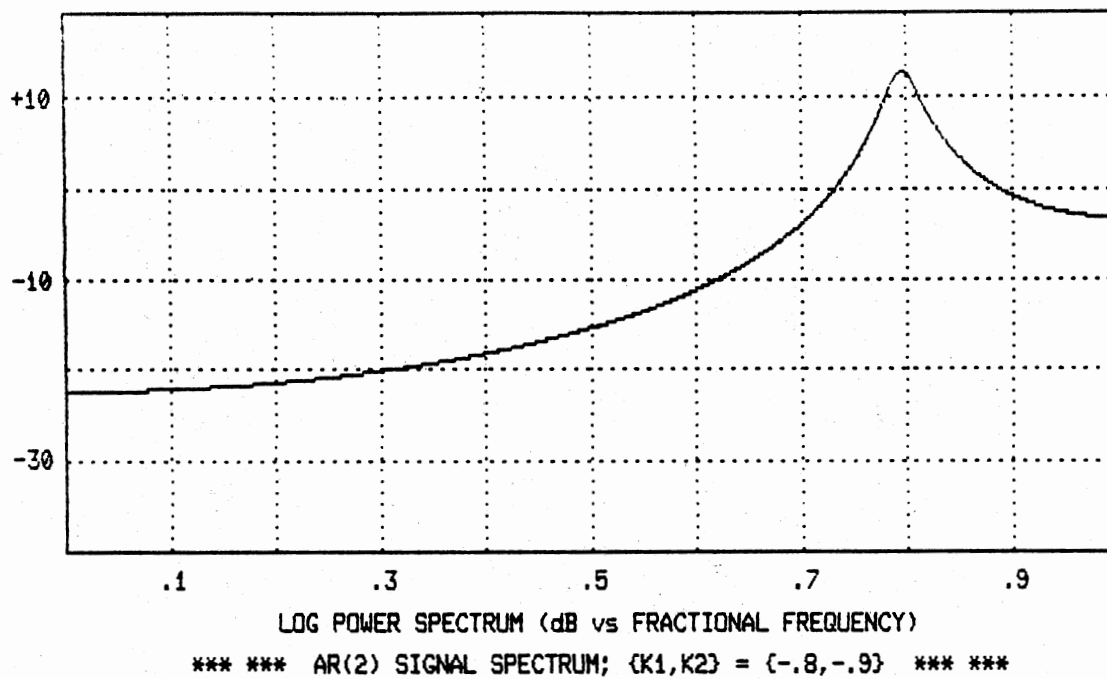
Figure 6a. True Spectrum; Test Signal Process

LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY)

*** *** AR(2) NOISE SPECTRUM; {K1,K2} = {+.8,-.9} *** ***
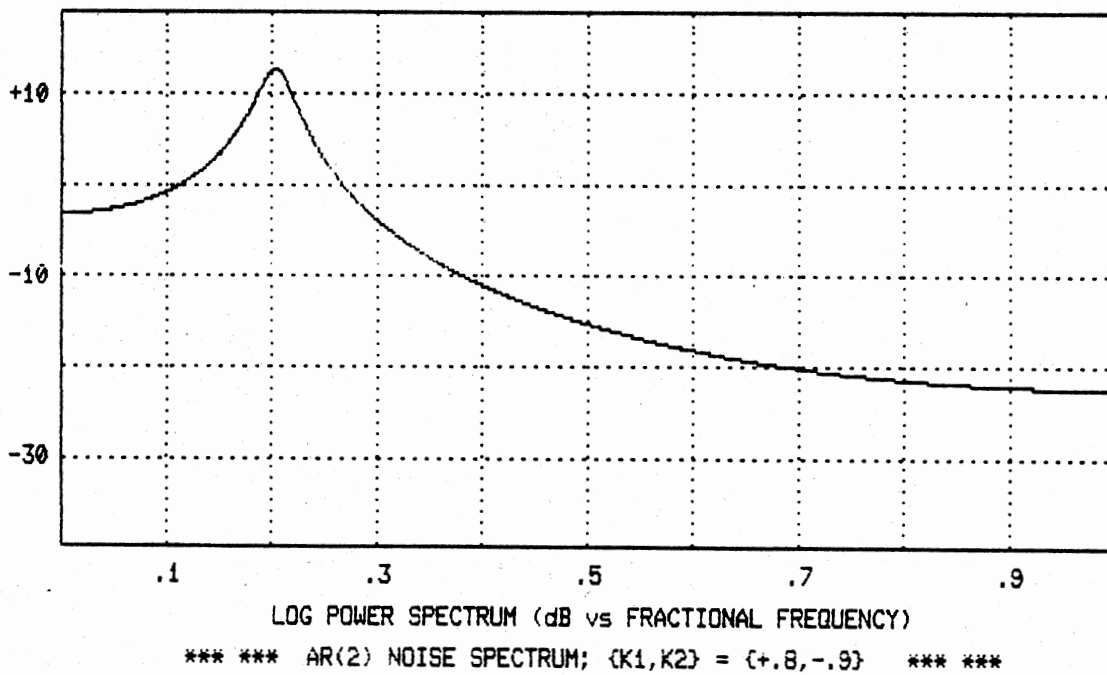
Figure 6b. True Spectrum; Test Noise Process

Figure 6. Test Signal Spectra

of reflection coefficients which define a single dot in Figure 7. For this "scatter plot" (and all subsequent scatter plots) the ordinate and the abcissa correspond to the first and second reflection coefficients, respectively; for convenience, cross-hairs indicate the location of the true parameter values.

Figures 8 and 9 each present various estimates of a single 200 sample Hamming windowed frame of data. In both cases the frame of data consists of the signal and colored noise processes summed at a 10 dB signal to noise ratio. The periodogram estimates in Figures 8a and 9a clearly display the signal resonance (near the fractional frequency value of .8) and the noise resonance (near the fractional frequency value of .2).

Figures 8b and 9b display power spectrum estimates obtained using the noise filtering formulation. The estimate presented in Figure 8b is a result of using the noise filter response displayed in Figure 8c which was determined by using the power subtraction rule;[2] similarly, Figure 9b results from the use of the noise filter response displayed in Figure 9c which was determined by using the magnitude subtraction rule.

---

[2]As indicated in the caption, the noise filter response was smoothed across frequencies before being applied. Although many smoothing algorithms are possible, only a recursive median smoother (with a length about 2.5% of the displayed bandwidth) was ever employed to obtain results presented in this chapter.

AUTOCORRELATION (YULE–WALKER) ANALYSIS
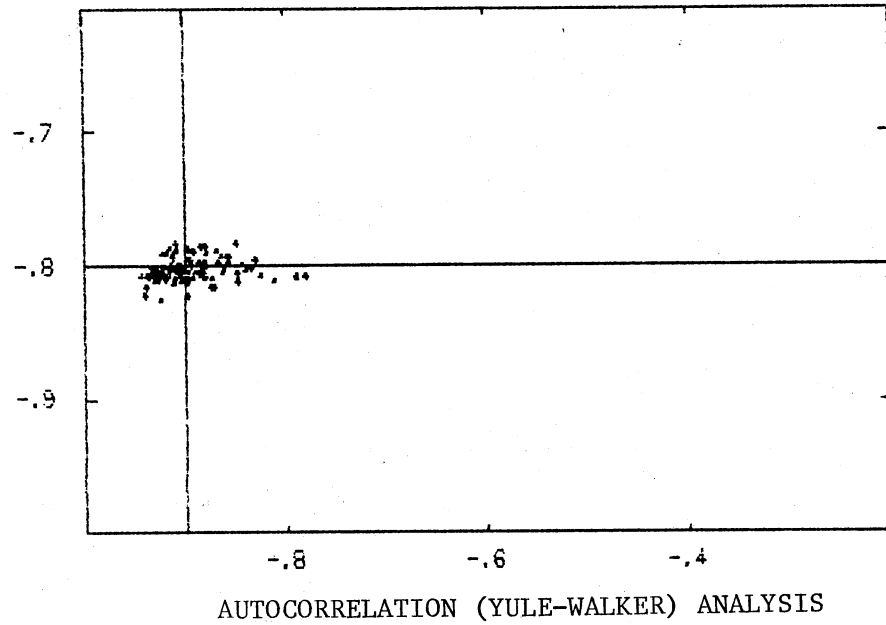
Figure 7. Scatter Plot; AR(2) Estimates
of Noiseless Signal Process
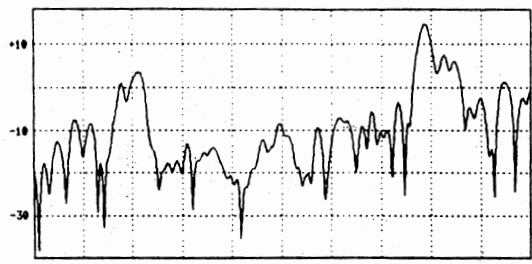
(a) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); PERIODO-
GRAM ESTIMATE OF SIGNAL + COLORED NOISE: SNR = 10 dB
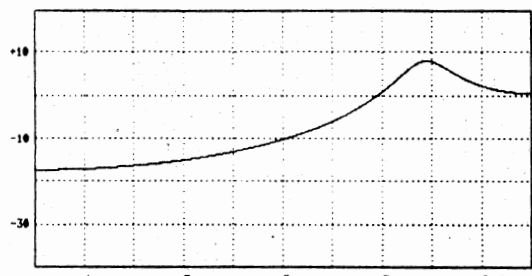
(b) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); AR(2)
ESTIMATE USING SMOOTHED POWER SUBTRACTION FILTER
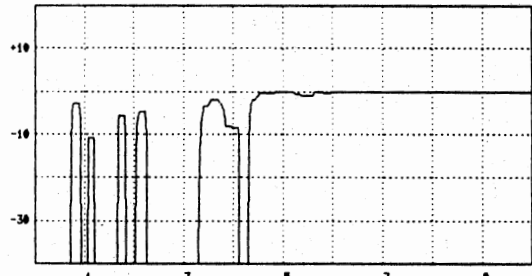
(c) SUPPRESSION FUNCTION (ATTENUATION [dB]) vs FRACTIONAL FREQUENCY;
NOISE FILTER RESPONSE--SMOOTHED POWER SUBTRACTION RULE
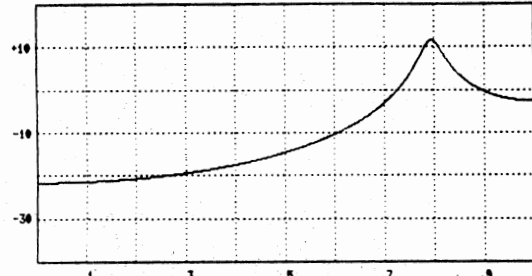
(d) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); AR(2) ESTI-
MATE USING WEIGHTED INFO. & SMOOTHED POWER SUBTRACTION

(e) WEIGHT FUNCTION (dB vs FRACTIONAL FREQUENCY); AR(2)
FIT TO SMOOTHED POWER SUBTRACTION FILTER RESPONSE

Figure 8. Gaussian Signal &
Colored Noise (10 dB)
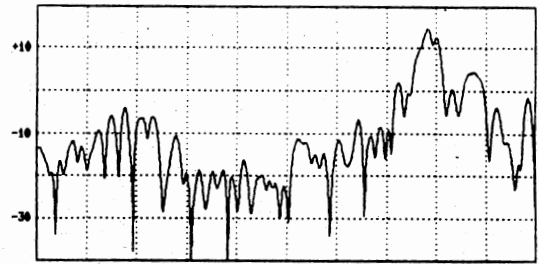
(a) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); PERIODO-
GRAM ESTIMATE OF SIGNAL + COLORED NOISE: SNR = 10 dB

(b) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); AR(2)
ESTIMATE USING SMOOTHED MAGNITUDE SUBTRACTION RULE

(c) SUPPRESSION FUNCTION (ATTENUATION [dB]) vs FRACTIONAL FREQUENCY;
NOISE FILTER RESPONSE--SMOOTHED MAGNITUDE SUBTRACTION RULE
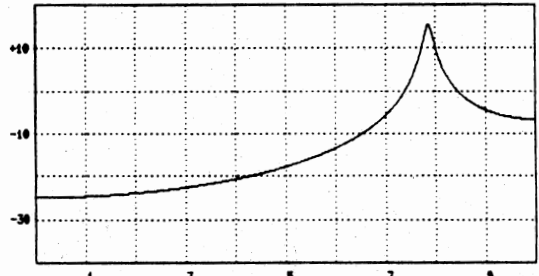
(d) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); AR(2) ESTI-
MATE USING WEIGHTED INFO. & SMOOTHED MAGNITUDE SUBTRACTION

(e) WEIGHT FUNCTION (dB vs FRACTIONAL FREQUENCY); AR(2) FIT
TO SMOOTHED MAGNITUDE SUBTRACTION FILTER RESPONSE

Figure 9. Gaussian Signal &
Colored Noise (10 dB)

Figures 8d and 9d display power spectrum estimates obtained using the weighted information formulation. The algebraic solution method, which requires an AR(M) form for the weight function, was used in both cases; coefficient evaluation was performed using the mixed time-frequency domain method presented in Figure 4. The same noise filter response functions were employed and the weight functions, displayed in Figures 8e and 9e, were determined as an AR(2) fit to their respective noise filter response functions.

By comparing Figures 8 and 9 to the true signal spectrum shown in Figure 6a the deficiencies of these typical estimates becomes apparent. In Figure 8b the noise filtering formulation leads to an estimate which is overly flat; the weighted information formulation (Figure 8d) has improved the estimate by raising the peak and lowering the valleys. In Figure 9b the noise filtering formulation leads to an estimate which is overly sharp; the weighted information formulation (Figure 9d) has improved the estimate by lowering the peak and raising the valleys. Since the weight functions are similar in both figures it is apparent that frequency weighting cannot be simply interpreted as increasing or decreasing the sharpness of a spectral estimate; rather, the weight function reduces distortions in the estimate by requiring a more accurate fit to the data in those spectral regions where the weight function is large.

Figures 10 and 11 present the result of analyzing 100 different 400 sample Hamming windowed frames of data using
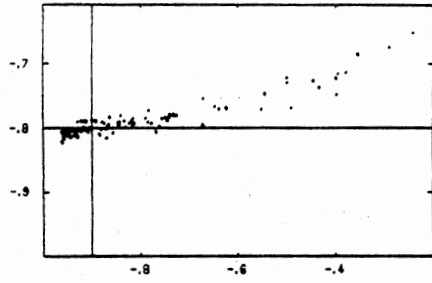
(a) POWER SUBTRACTION - SMOOTHED
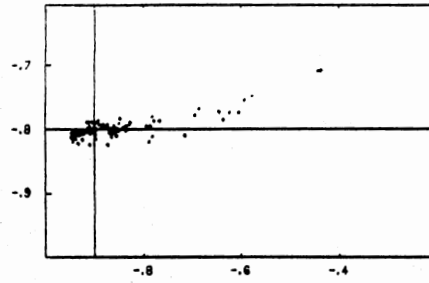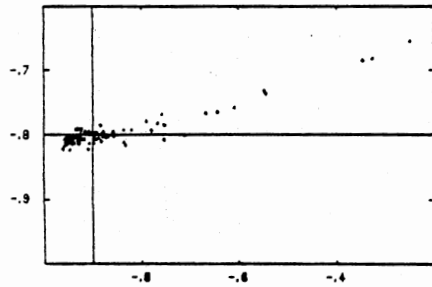
(b) SOFT(4) SUPPRESSION - SMOOTHED

(c) SOFT(6) SUPPRESSION - SMOOTHED
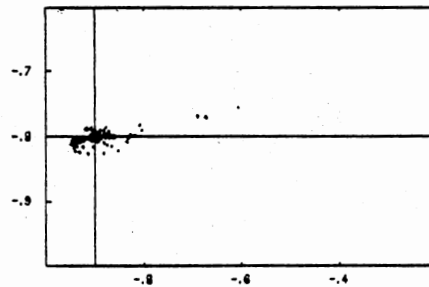
(d) SOFT(8) SUPPRESSION - SMOOTHED
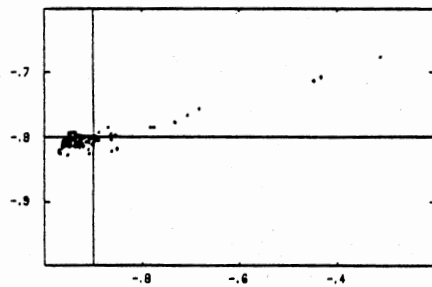
(e) MAGNITUDE SUBTRACTION - SMOOTHED
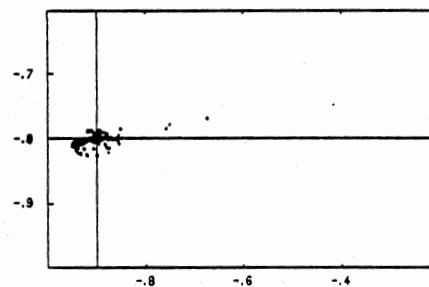
(a) POWER/SMOOTHED & AR(2) WEIGHTED INFO.
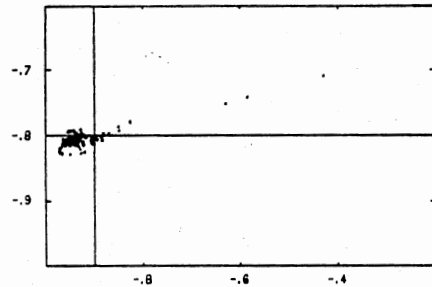
(b) SOFT(4)/SMOOTHED & AR(2) WEIGHTED INFO.
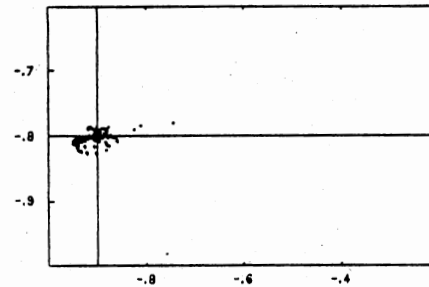
(c) SOFT(6)/SMOOTHED & AR(2) WEIGHTED INFO.
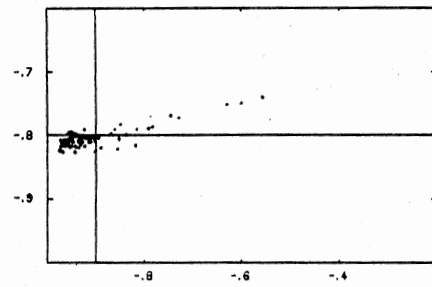
(d) SOFT(8)/SMOOTHED & AR(2) WEIGHTED INFO.
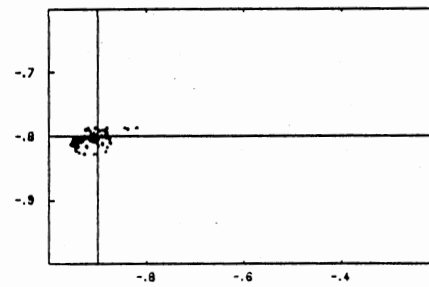
(e) MAGNITUDE/SMOOTHED & AR(2) WEIGHTED INFO.

Figure 10. Noise Filtering:
Gaussian Signal & Colored
Noise (10 dB)

Figure 11. Weighted Informa-
tion: Gaussian Signal &
Colored Noise (10 dB)

various different methods. Figure 10 presents the results obtained using the noise filtering formulation; the smoothed noise filter response was determined using different rules ranging (roughly) from the least severe rule in Figure 10a to the most severe in Figure 10e. The results presented in Figure 11 represent an analysis of the same 100 data frames and the same noise filter response functions but the analysis uses the weighted information formulation with an AR(2) weight function fit to the noise filter response function.

It is clear that in each case (a through e) the estimation error is reduced by the weighted information formulation. The best results in both figures are obtained by the most severe rules. Figure 10, while exhibiting less variance, shows an increased deviation (bias) of the main cluster from the true values for the more severe rules; apparently, variance error of the noise filtering formulation may be reduced at the expense of increased bias error by using the more severe rules. Comparing, for example, Figures 10e and 11e it is apparent that the weighted information formulation achieves still greater variance reduction while correcting the bias error. Comparison of Figures 11e and 7 indicate that one has little, if any, hope of achieving significantly better performance than that provided by the weighted information formulation in this case.

Figures 12 and 13 show similar results for the same 100 frames of data; the analysis methods used to produce these

(a) POWER SUBTRACTION - UNSMOOTHED

(b) SOFT(4) SUPPRESSION - UNSMOOTHED

(c) SOFT(6) SUPPRESSION - UNSMOOTHED

(d) SOFT(8) SUPPRESSION - UNSMOOTHED

(e) MAGNITUDE SUBTRACTION - UNSMOOTHED
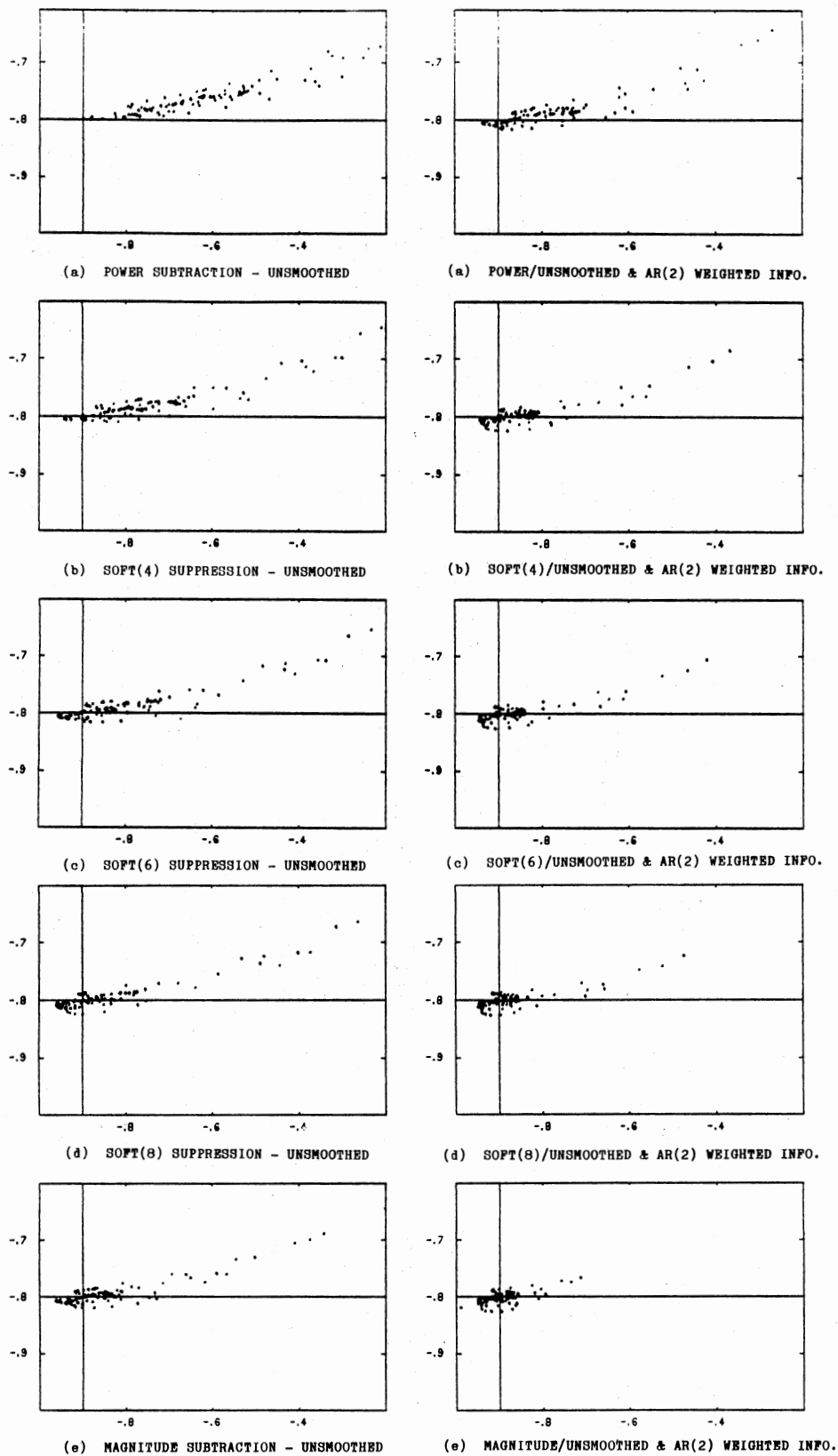
Figure 12. Noise Filtering: Gaussian Signal & Colored Noise (10 dB)

(a) POWER/UNSMOOTHED & AR(2) WEIGHTED INFO.

(b) SOFT(4)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(c) SOFT(6)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(d) SOFT(8)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(e) MAGNITUDE/UNSMOOTHED & AR(2) WEIGHTED INFO.

Figure 13. Weighted Information: Gaussian Signal & Colored Noise (10 dB)

figures differ from the method used to produce Figures 10 and 11 only in that no smoothing algorithm was applied to the noise filter response. All the same trends are apparent in figures 12 and 13 as were apparent in Figures 10 and 11; somewhat greater variance is exhibited in these figures indicating that smoothing produces a generally beneficial effect in this case.

Figures 14, 15, 16, and 17 display similar results for the case of white corrupting noise at a 10 dB signal to noise ratio. Again, each plot represents analysis of the same 100 different 400 sample Hamming windowed frames of data. For each method of determining noise filter response, the weighted information formulation leads to less variance and bias error than the comparable (unweighted) noise filtering formulation. As may be expected,[3] all these estimators yield poorer performance in this white noise case than in the previous colored noise case.

Figures 18, 19, 20, and 21 again present similar results; while the corrupting noise is still white the signal to noise ratio is now zero dB. One small difference is worth noting: in Figures 10 through 17 the parts b, c, and d employed a soft suppression rule with suppression factors

---

[3]The reader will recall that if the signal and noise processes are completely separated in frequency (i.e., do not have overlapping spectra), the Wiener filter can completely eliminate the noise. Since the colored noise case exhibits greater spectral separation from this signal process than the white noise case, an estimate can be expected to provide superior performance.

(a) POWER SUBTRACTION - SMOOTHED

(b) SOFT(4) SUPPRESSION - SMOOTHED

(c) SOFT(6) SUPPRESSION - SMOOTHED

(d) SOFT(8) SUPPRESSION - SMOOTHED

(e) MAGNITUDE SUBTRACTION - SMOOTHED

(a) POWER/SMOOTHED & AR(2) WEIGHTED INFO.

(b) SOFT(4)/SMOOTHED & AR(2) WEIGHTED INFO.

(c) SOFT(6)/SMOOTHED & AR(2) WEIGHTED INFO.

(d) SOFT(8)/SMOOTHED & AR(2) WEIGHTED INFO.

(e) MAGNITUDE/SMOOTHED & AR(2) WEIGHTED INFO.

Figure 14. Noise Filtering:
Gaussian Signal & White
Noise (10 dB)

Figure 15. Weighted Informa-
tion: Gaussian Signal &
White Noise (10 dB)
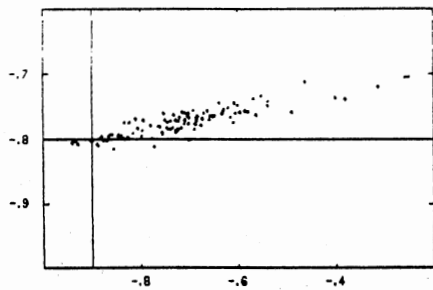
(a) POWER SUBTRACTION – UNSMOOTHED
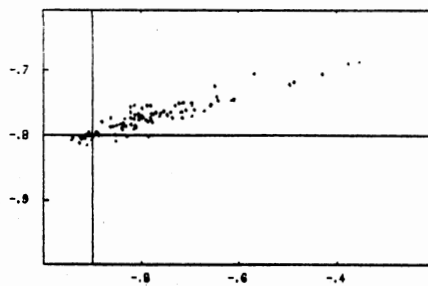
(a) POWER/UNSMOOTHED & AR(2) WEIGHTED INFO.
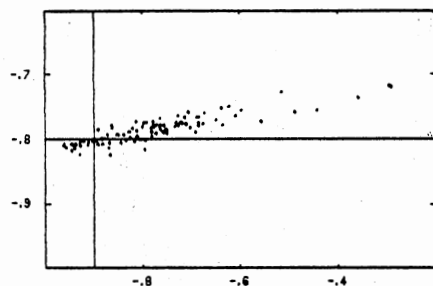
(b) SOFT(4) SUPPRESSION – UNSMOOTHED
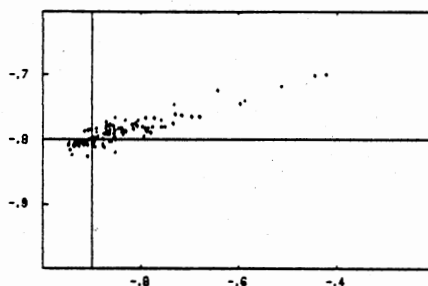
(b) SOFT(4)/UNSMOOTHED & AR(2) WEIGHTED INFO.
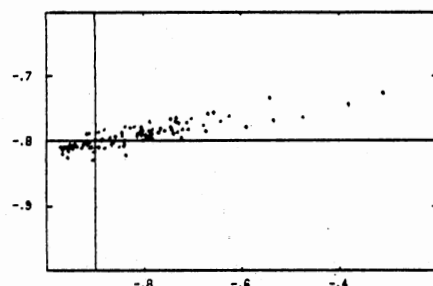
(c) SOFT(6) SUPPRESSION – UNSMOOTHED
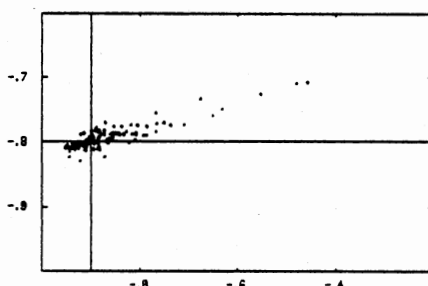
(c) SOFT(6)/UNSMOOTHED & AR(2) WEIGHTED INFO.
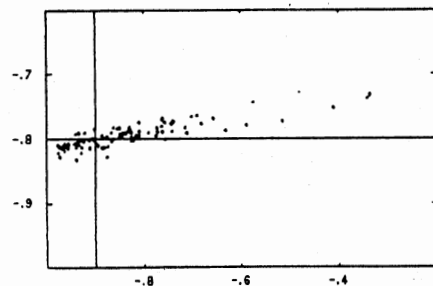
(d) SOFT(8) SUPPRESSION – UNSMOOTHED
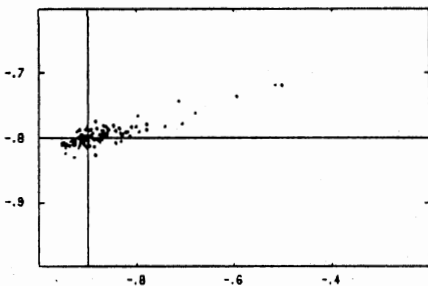
(d) SOFT(8)/UNSMOOTHED & AR(2) WEIGHTED INFO.
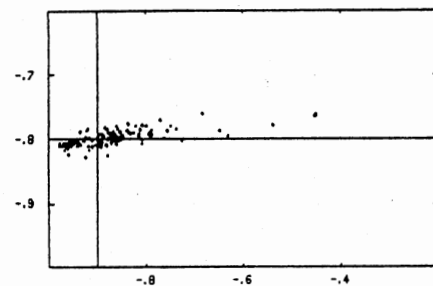
(e) MAGNITUDE SUBTRACTION – UNSMOOTHED
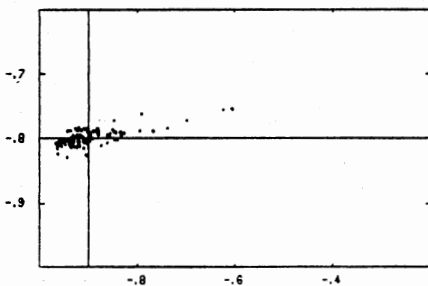
(e) MAGNITUDE/UNSMOOTHED & AR(2) WEIGHTED INFO.

Figure 16. Noise Filtering: Gaussian Signal & White Noise (10 dB)

Figure 17. Weighted Information: Gaussian Signal & White Noise (10 dB)
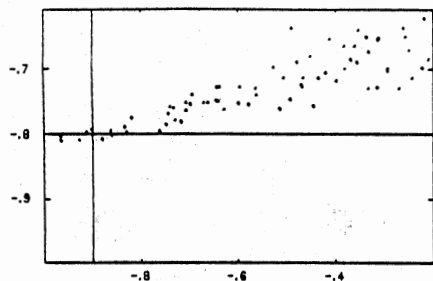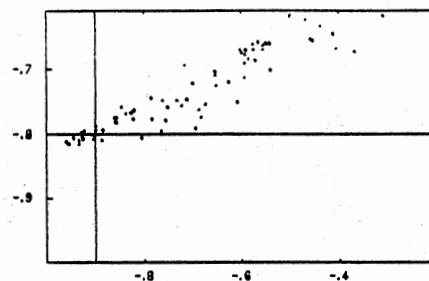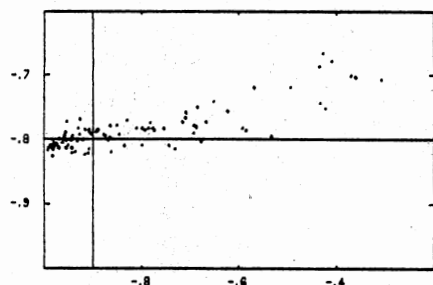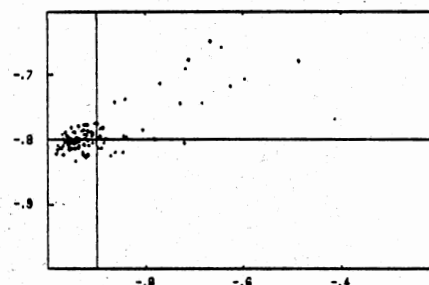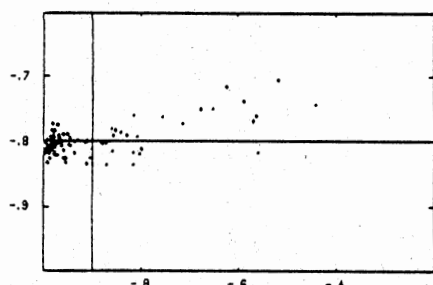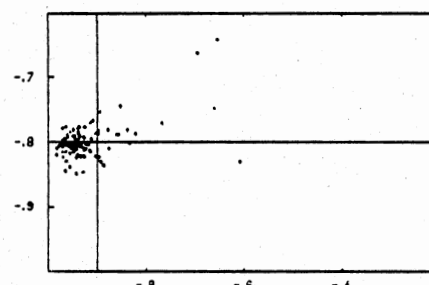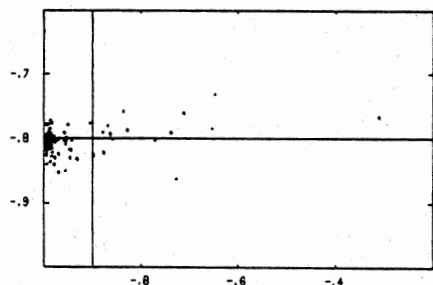
(a)   POWER SUBTRACTION - SMOOTHED

(b)   SOFT(6) SUPPRESSION - SMOOTHED

(c)   SOFT(8) SUPPRESSION - SMOOTHED

(d)   SOFT(10) SUPPRESSION - SMOOTHED
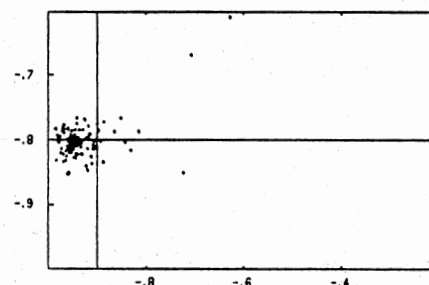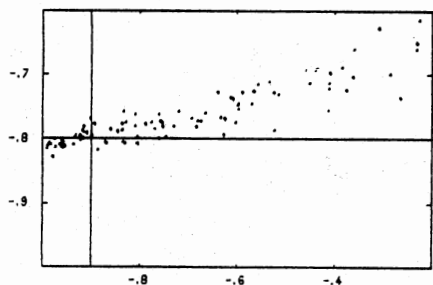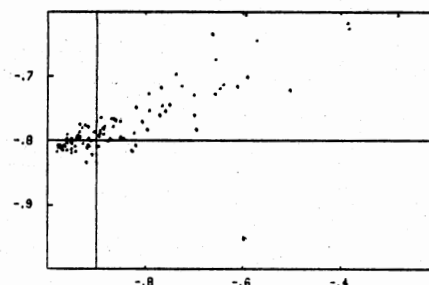
(e)   MAGNITUDE SUBTRACTION - SMOOTHED

Figure 18. Noise Filtering:
Gaussian Signal & White
Noise (0 dB)

(a)   POWER/SMOOTHED & AR(2) WEIGHTED INFO.

(b)   SOFT(6)/SMOOTHED & AR(2) WEIGHTED INFO.

(c)   SOFT(8)/SMOOTHED & AR(2) WEIGHTED INFO.

(d)   SOFT(10)/SMOOTHED & AR(2) WEIGHTED INFO.

(e)   MAGNITUDE/SMOOTHED & AR(2) WEIGHTED INFO.

Figure 19. Weighted Informa-
tion: Gaussian Signal &
White Noise (0 dB)

(a)   POWER SUBTRACTION - UNSMOOTHED

(b)   SOFT(6) SUPPRESSION - UNSMOOTHED

(c)   SOFT(8) SUPPRESSION - UNSMOOTHED

(d)   SOFT(10) SUPPRESSION - UNSMOOTHED

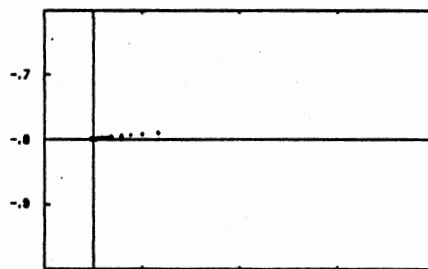(e)   MAGNITUDE SUBTRACTION - UNSMOOTHED

(a)   POWER/UNSMOOTHED & AR(2) WEIGHTED INFO.

(b)   SOFT(6)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(c)   SOFT(8)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(d)   SOFT(10)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(e)   MAGNITUDE/UNSMOOTHED & AR(2) WEIGHTED INFO.

Figure 20. Noise Filtering: Gaussian Signal & White Noise (0 dB)

Figure 21. Weighted Information: Gaussian Signal & White Noise (0 dB)

of 4, 6, and 8 respectively; in Figures 18 through 21 the parts b, c, and d again employ a soft suppression rule but with increased suppression factors of 6, 8, and 10 respectively.

Speech and Speech-Like Signals

Many speech waveforms exhibit a nonrandom periodic character; their spectra display a fine harmonic structure (with peaks separated by integral multiples of the pitch frequency) with a roughly AR modulation. The harmonic structure is generally attributed to the periodic glottal pulses while the AR modulation is generally attributed to the response characteristics of the vocal tract.

To simulate such waveforms the all pole filter with frequency response displayed in figure 6a was excited with a periodic stream of impulses (with a period of 79 samples). No figure comparable to Figure 7 is included here since, in the absence of noise, the analysis of 100 different 400 sample Hamming windowed frames of data (with a random distribution of phase displacement) presents no apparent estimation error.[4] Consequently, while part of the apparent estimation error in the scatter plots of Figures 10 through 21 must be attributed to the random character of the signal

---

[4]That is, on the scale used for these scatter plots. On a greatly enlarged scale, a small amount of bias and variance error may be observed.

itself, all of the apparent estimation error in the following scatter plots (Figures 24 through 35) may be attributed to the presence of noise.

Figures 22 and 23 each present various estimates of a single 200 sample Hamming windowed frame of data. In both cases the frame of data consists of the aforementioned periodic signal process and a colored Gaussian noise process summed at a 10 dB signal to noise ratio. The periodogram estimates in Figures 22a and 23a clearly display the fine harmonic structure of the signal spectrum near the filter resonance (fractional frequency of .8) while this structure breaks down near the noise resonance (fractional frequency of .2).

Figures 22b and 23b display estimates obtained using the noise filtering formulation; Figures 22c and 23c display the noise filter response characteristics that produced these estimates. Clearly the estimate appearing in Figure 22b is overly flat while the estimate appearing in Figure 23b is overly sharp. Figures 22d and 23d display the estimates obtained using the weighted information formulation; comparison with Figure 6a reveals that both these estimates are improved relative to their counterparts in Figures 22b and 23b. Finally the AR(2) weight functions approximating the noise filter response functions are presented in Figures 22e and 23e.

Figures 24, 25, 26, and 27 display a variety of scatter plots; each scatter plot presents the result of analyzing
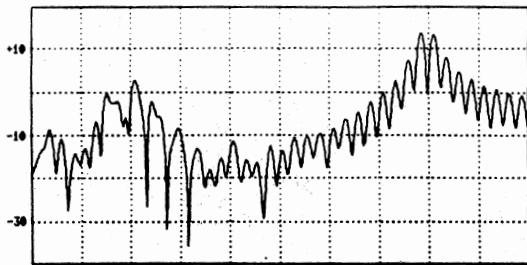
(a) LOW POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); PERIODO-
GRAM ESTIMATE OF SIGNAL + COLORED NOISE: SNR = 10 dB
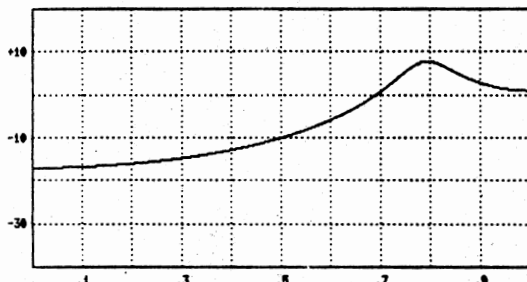
(b) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); AR(2)
ESTIMATE USING SMOOTHED POWER SUBTRACTION FILTER
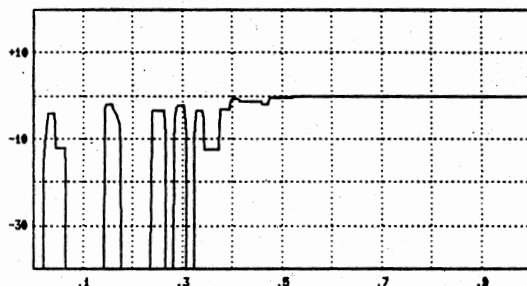
(c) SUPPRESSION FUNCTION (ATTENUATION [dB]) vs FRACTIONAL FRE-
QUENCY; NOISE FILTER RESPONSE--SMOOTHED POWER SUB. RULE
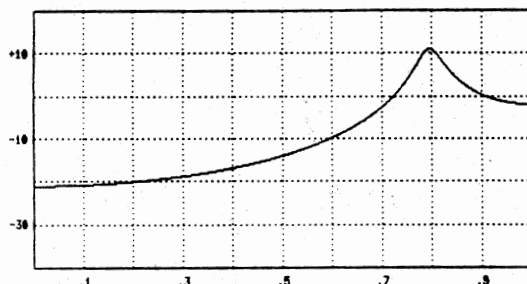
(d) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); AR(2)
ESTIMATE USING WEIGHTED INFO. & SMOOTHED POWER SUB.
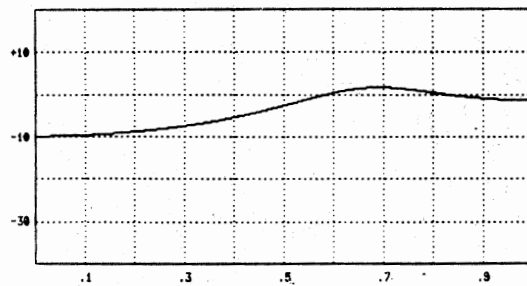
(e) WEIGHT FUNCTION (dB vs FRACTIONAL FREQUENCY); AR(2) FIT
TO SMOOTHED POWER SUBTRACTION FILTER RESPONSE

Figure 22. Periodic Signal &
Colored Noise (10 dB)
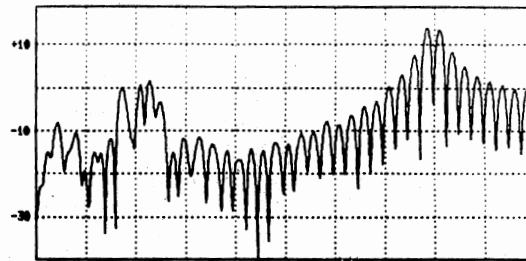
(a) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); PERIODO-
GRAM ESTIMATE OF SIGNAL + COLORED NOISE: SNR = 10 dB

(b) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); AR(2)
ESTIMATE USING SMOOTHED MAGNITUDE SUBTRACTION RULE

(c) SUPPRESSION FUNCTION (ATTENUATION [dB]) vs FRACTIONAL FREQUENCY;
NOISE FILTER RESPONSE--SMOOTHED MAGNITUDE SUBTRACTION RULE

(d) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); AR(2) ESTI-
MATE USING WEIGHTED INFO. & SMOOTHED MAGNITUDE SUBTRACTION

(e) WEIGHT FUNCTION (dB vs FRACTIONAL FREQUENCY); AR(2) FIT
TO SMOOTHED MAGNITUDE SUBTRACTION FILTER RESPONSE

Figure 23. Periodic Signal &
Colored Noise (10 dB)
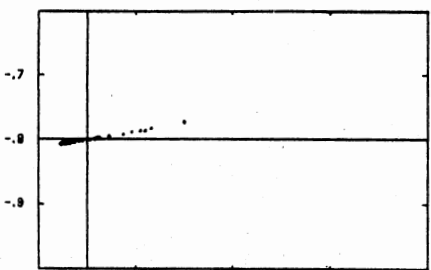
(a) POWER SUBTRACTION - SMOOTHED

(b) SOFT(4) SUPPRESSION - SMOOTHED

(c) SOFT(6) SUPPRESSION - SMOOTHED

(d) SOFT(8) SUPPRESSION - SMOOTHED

(e) MAGNITUDE SUBTRACTION - SMOOTHED

Figure 24. Noise Filtering:
Periodic Signal & Colored
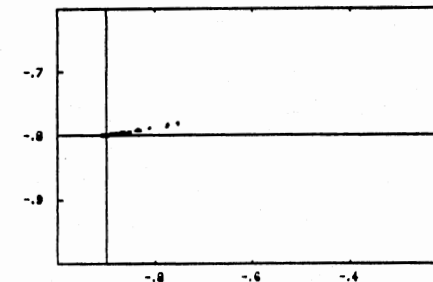Noise (10 dB)

(a) POWER/SMOOTHED & AR(2) WEIGHTED INFO.

(b) SOFT(4)/SMOOTHED & AR(2) WEIGHTED INFO.
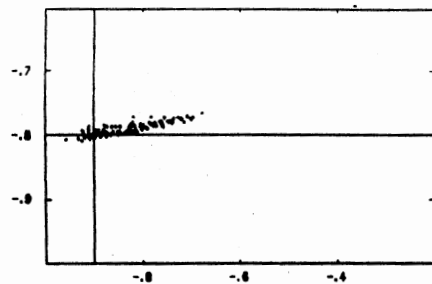
(c) SOFT(6)/SMOOTHED & AR(2) WEIGHTED INFO.
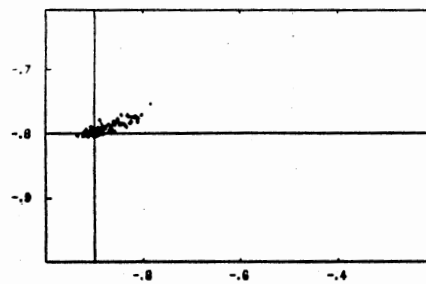
(d) SOFT(8)/SMOOTHED & AR(2) WEIGHTED INFO.

(e) MAGNITUDE/SMOOTHED & AR(2) WEIGHTED INFO.

Figure 25. Weighted Informa-
tion: Periodic Signal &
Colored Noise (10 dB)
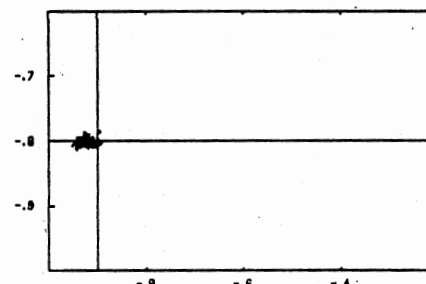
(a) POWER SUBTRACTION - UNSMOOTHED

(b) SOFT(4) SUPPRESSION - UNSMOOTHED

(c) SOFT(6) SUPPRESSION - UNSMOOTHED

(d) SOFT(8) SUPPRESSION - UNSMOOTHED

(e) MAGNITUDE SUBTRACTION - UNSMOOTHED

Figure 26. Noise Filtering:
Periodic Signal & Colored
Noise (10 dB)

(a) POWER/UNSMOOTHED & AR(2) WEIGHTED INFO.

(b) SOFT(4)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(c) SOFT(6)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(d) SOFT(8)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(e) MAGNITUDE/UNSMOOTHED & AR(2) WEIGHTED INFO.

Figure 27. Weighted Informa-
tion: Periodic Signal &
Colored Noise (10 dB)
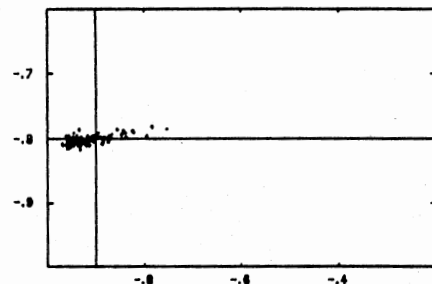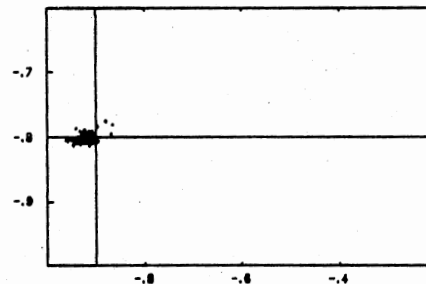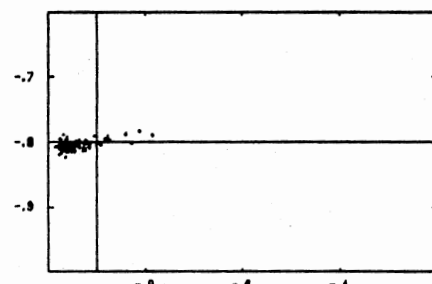
(a) POWER SUBTRACTION - SMOOTHED

(b) SOFT(4) SUPPRESSION - SMOOTHED

(c) SOFT(6) SUPPRESSION - SMOOTHED

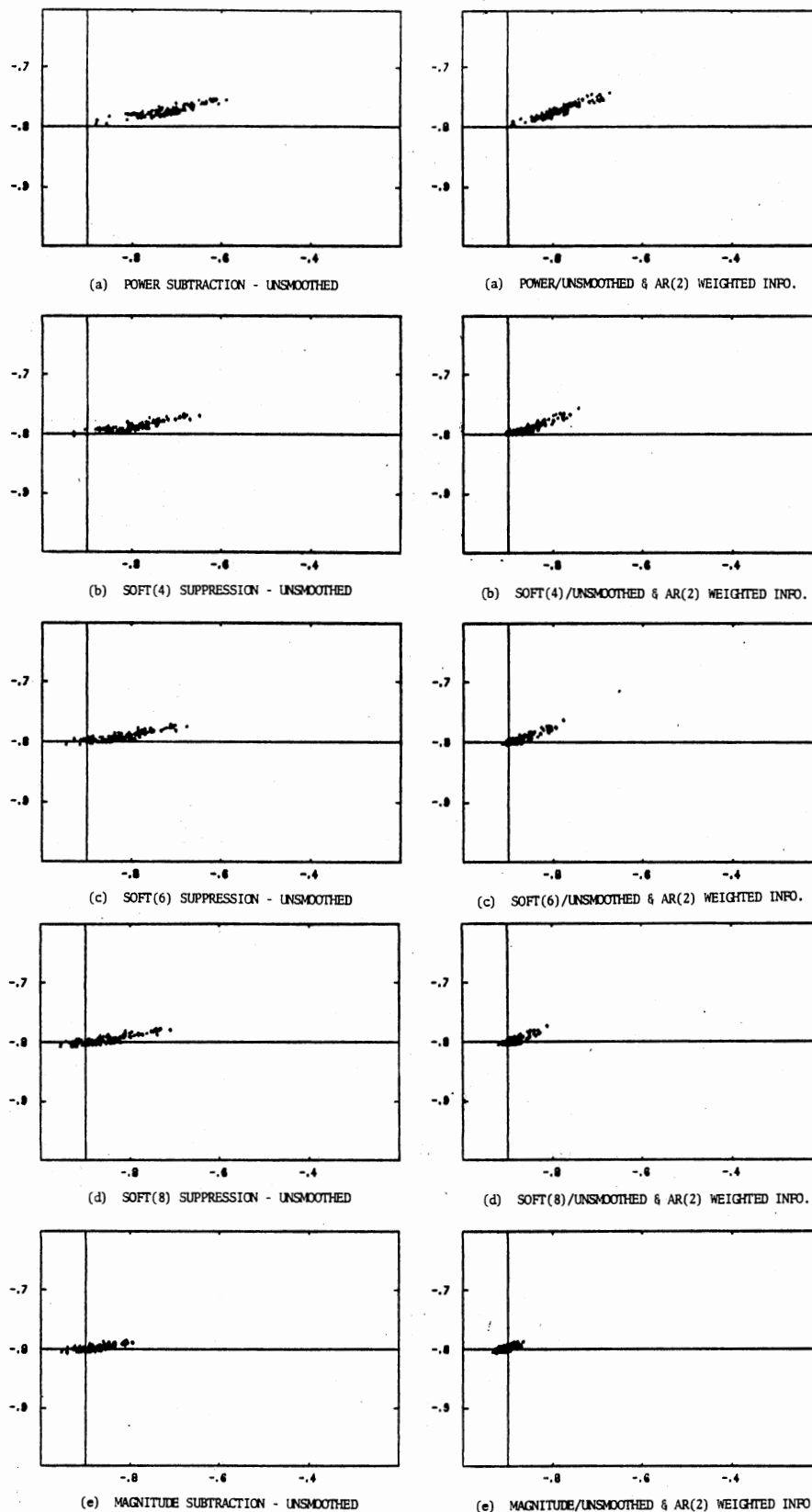(d) SOFT(8) SUPPRESSION - SMOOTHED

(e) MAGNITUDE SUBTRACTION - SMOOTHED

Figure 28. Noise Filtering:
Periodic Signal & White
Noise (10 dB)

(a) POWER/SMOOTHED & AR(2) WEIGHTED INFO.

(b) SOFT(4)/SMOOTHED & AR(2) WEIGHTED INFO.

(c) SOFT(6)/SMOOTHED & AR(2) WEIGHTED INFO.

(d) SOFT(8)/SMOOTHED & AR(2) WEIGHTED INFO.

(e) MAGNITUDE/SMOOTHED & AR(2) WEIGHTED INFO.

Figure 29. Weighted Informa-
tion: Periodic Signal &
White Noise (10 dB)

(a) POWER SUBTRACTION - UNSMOOTHED

(b) SOFT(4) SUPPRESSION - UNSMOOTHED

(c) SOFT(6) SUPPRESSION - UNSMOOTHED

(d) SOFT(8) SUPPRESSION - UNSMOOTHED

(e) MAGNITUDE SUBTRACTION - UNSMOOTHED

(a) POWER/UNSMOOTHED & AR(2) WEIGHTED INFO.

(b) SOFT(4)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(c) SOFT(6)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(d) SOFT(8)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(e) MAGNITUDE/UNSMOOTHED & AR(2) WEIGHTED INFO.
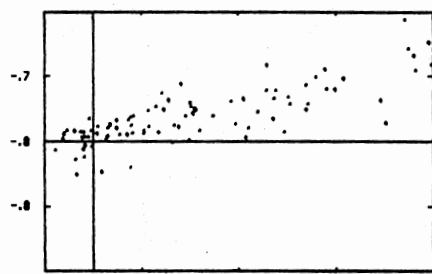
Figure 30. Noise Filtering: Periodic Signal & White Noise (10 dB)

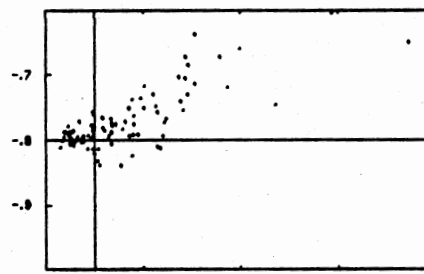Figure 31. Weighted Information: Periodic Signal & White Noise (10 dB)
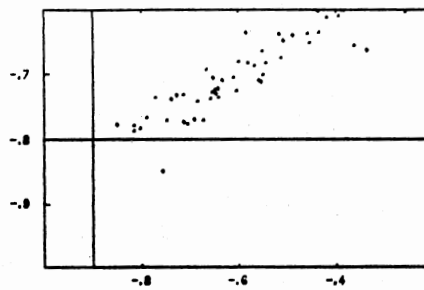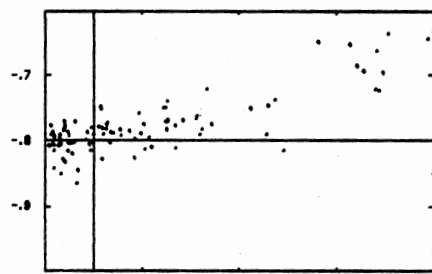
(a) POWER SUBTRACTION - SMOOTHED

(b) SOFT(6) SUPPRESSION - SMOOTHED

(c) SOFT(8) SUPPRESSION - SMOOTHED

(d) SOFT(10) SUPPRESSION - SMOOTHED

(e) MAGNITUDE SUBTRACTION - SMOOTHED

(a) POWER/SMOOTHED & AR(2) WEIGHTED INFO.
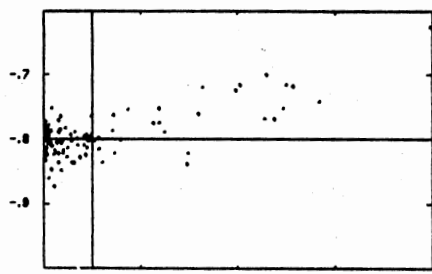
(b) SOFT(6)/SMOOTHED & AR(2) WEIGHTED INFO.
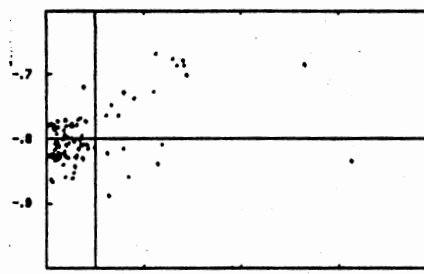
(c) SOFT(8)/SMOOTHED & AR(2) WEIGHTED INFO.
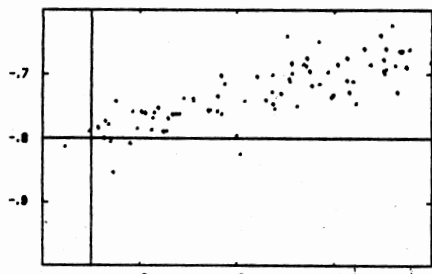
(d) SOFT(10)/SMOOTHED & AR(2) WEIGHTED INFO.
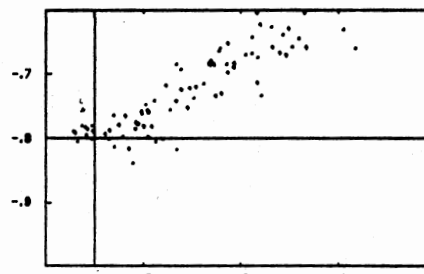
(e) MAGNITUDE/SMOOTHED & AR(2) WEIGHTED INFO.

Figure 32. Noise Filtering: Periodic Signal & White Noise (0 dB)

Figure 33. Weighted Information: Periodic Signal & White Noise (0 dB)
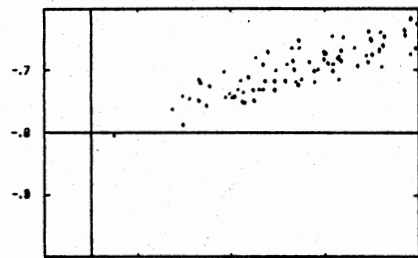
(a) POWER SUBTRACTION - UNSMOOTHED

(b) SOFT(6) SUPPRESSION - UNSMOOTHED
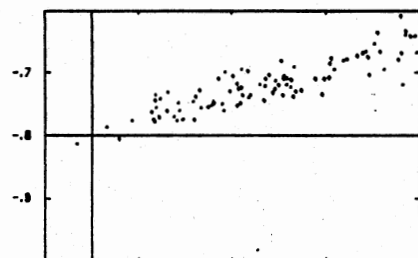
(c) SOFT(8) SUPPRESSION - UNSMOOTHED
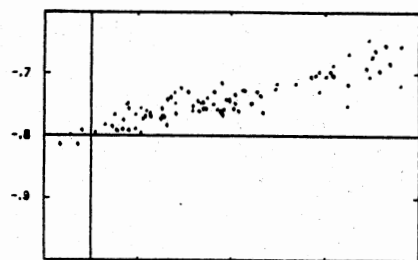
(d) SOFT(10) SUPPRESSION - UNSMOOTHED
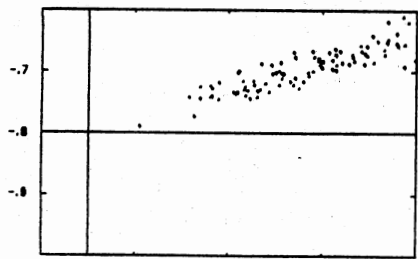
(e) MAGNITUDE SUBTRACTION - UNSMOOTHED

Figure 34. Noise Filtering: Periodic Signal & White Noise (0 dB)
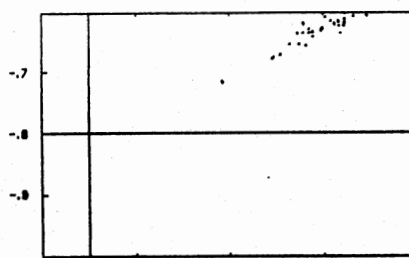
(a) POWER/UNSMOOTHED & AR(2) WEIGHTED INFO.
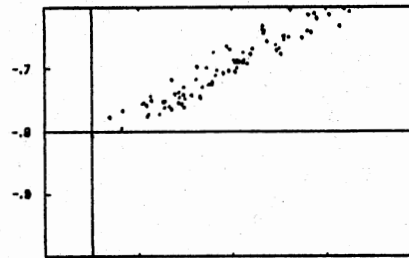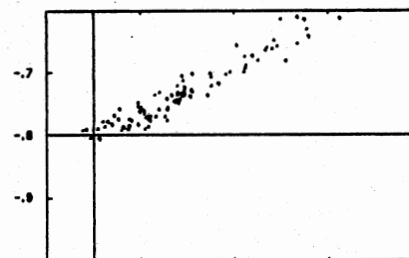
(b) SOFT(6)/UNSMOOTHED & AR(2) WEIGHTED INFO.

(c) SOFT(8)/UNSMOOTHED & AR(2) WEIGHTED INFO.
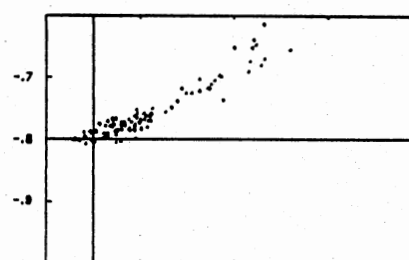
(d) SOFT(10)/UNSMOOTHED & AR(2) WEIGHTED INFO.
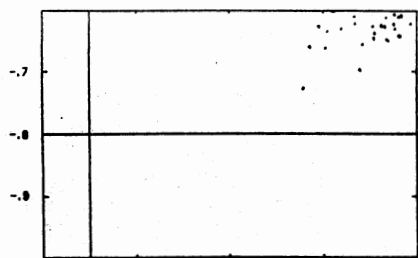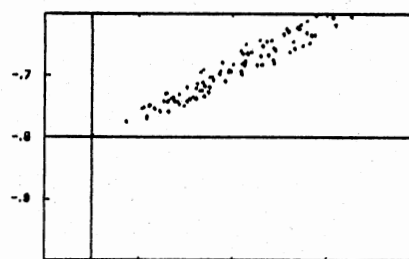
(e) MAGNITUDE/UNSMOOTHED & AR(2) WEIGHTED INFO.

Figure 35. Weighted Information: Periodic Signal & White Noise (0 dB)

100 different 400 sample Hamming windowed frames of data; the same 100 data frames were employed for each plot. As mentioned above, because the signal process is periodic and not random in character all of the apparent estimation error can be attributed to the added colored Gaussian noise (SNR = 10 dB).

Figures 24 and 25 employ smoothed noise filter response characteristics while Figures 26 and 27 employ the unsmoothed characteristics.[5] Figures 24 and 26 display the results obtained with the noise filtering formulation while Figures 25 and 27 display the results obtained with the AR(2) weighted information formulation. Once again, the weighted information formulation leads to less estimation error than the comparable noise filtering formulation; in Figures 25d and 25e the estimation error is so small as to be almost imperceptible on the scale employed for these plots. Smoothing still appears to display a generally beneficial effect.

Figures 28, 29, 30, and 31 present similar results for the case of white Gaussian noise corruption to the periodic signal processes (SNR = 10 dB). As with the Gaussian signal

----

[5]Some caution is advised regarding the use of smoothers here. The dimensions of the lobes within the fine harmonic structure are controlled by the length and shape of the data window so that a smoother that works well with one frame length may not work well with longer frames or a differently shaped window.

process, all the estimates present degraded performance in this white noise case as compared to the colored noise case.

To complete these simulations, Figures 32, 33, 34, and 35 present analysis results for the case of white Gaussian noise corruption to the periodic signal process at a zero dB signal to noise ratiio. As with the Gaussian signal process, parts b, c, and d of these figures employ soft suppression rules with increased suppression factors of 6, 8,and 10 respectively.

The following summarizes the description of these scatter plots. Figures 10-13 and 24-27 correspond to colored noise corruption at a 10 dB signal to noise ratio; Figures 14-17 and 28-31 correspond to white noise corruption at a 10 dB signal to noise ratio; Figures 18-21 and 32-35 correspond to white noise corruption at a 0 dB signal to noise ratio. Figures 10-21 correspond to a Gaussian random signal; Figures 24-35 correspond to a periodic (period = 79 samples) signal. Figures 10, 11, 14, 15, 18, 19, 24, 25, 28, 29, 32, and 33 employ a smoothed noise filter response while the remainder employ an unsmoothed response; parts a and e of each of these figures determine the noise filter response using the power and magnitude subtraction rules respectively while parts b, c, and d employ the soft suppression rules. In Figures 10-17 and 24-31 the suppression factors for parts b, c, and d are 4, 6, and 8 respectively; in Figures 18-21 and 32-35 the suppression factors and 6, 8, and 10 respectively. Finally, Figures 10, 12, 14, 16, 18, 20, 24,

26, 28, 30, 32 and 34 display the results of the (un-weighted) noise filtering analysis while Figures 11, 13, 15, 17, 19, 21, 25, 27, 29, 31, 33, and 35 display the results of the AR(2) weighted information analysis.

Before concluding this chapter, several examples resulting from the analysis of a real speech segment are provided. Figure 36a shows a periodogram estimate obtained from a Hamming windowed 400 sample segment taken from the vowel portion of the word "wrap";[6] from the fine harmonic structure it is apparent that the pitch of this segment is about 135 Hz (about 59 samples). Figure 36b shows a tenth order AR estimate of the spectrum obtained as the result of an autocorrelation method analysis of the same Hamming windowed segment; four vocal tract resonances are clearly visible.[7]

Figures 37a and 37b show periodogram and tenth order AR estimates obtained from this same vowel segment after adding white noise at a 10 dB signal to noise ratio. Clearly, the fine harmonic structure of the periodogram estimate has been partially obscured and, while four resonances are still visible, the AR estimate is severely distorted.

---

[6]The word, spoken in context by an adult male in a quiet environment, was taken from the sentence "Don't gift wrap the tall glass." and was appropriately filtered before sampling at 8 kHz.

[7]Lower and higher order analyses were applied to this segment and it was judged from plots such as these that a tenth order model is appropriate.
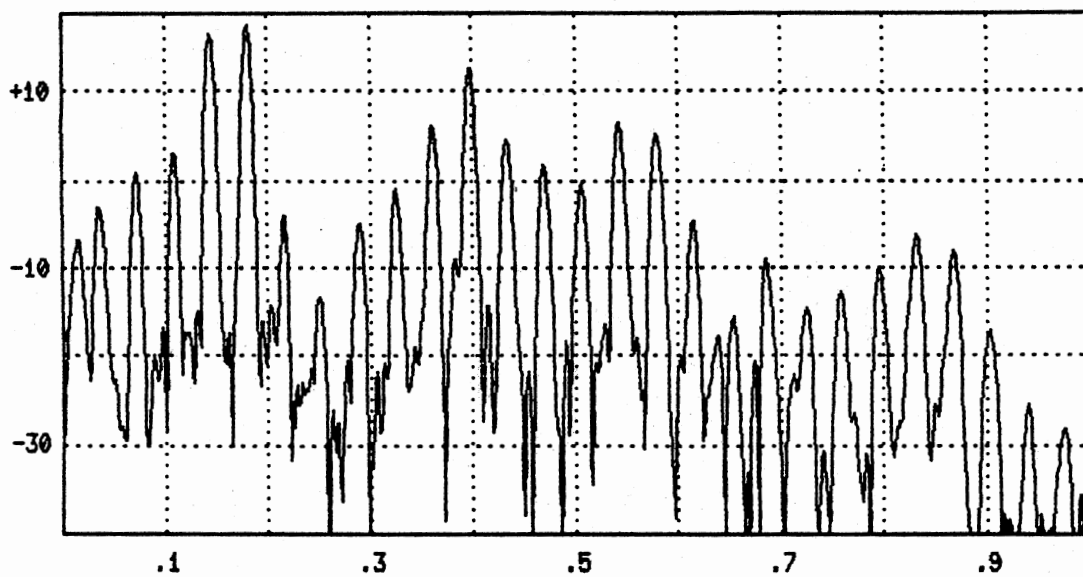
(a) Log Power Spectrum (dB vs Fractional Frequency); Periodogram Estimate of Vowel Spectrum; Noise Free
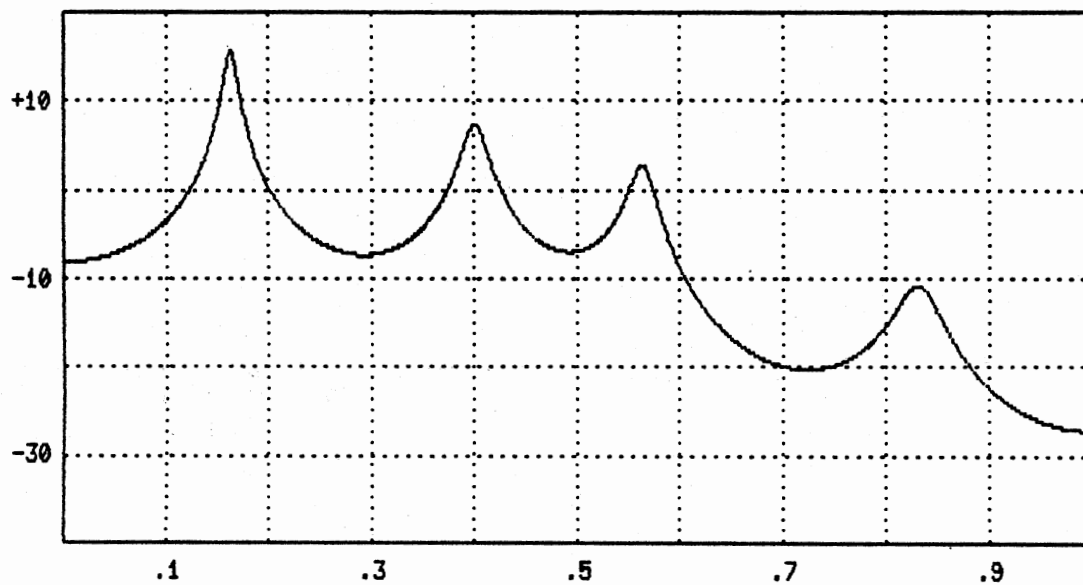
(b) Log Power Spectrum (dB vs Fractional Frequency); AR(10) Estimate Using Autocorrelation Method; Noise Free

Figure 36. Vowel Spectrum in Quiet Environment
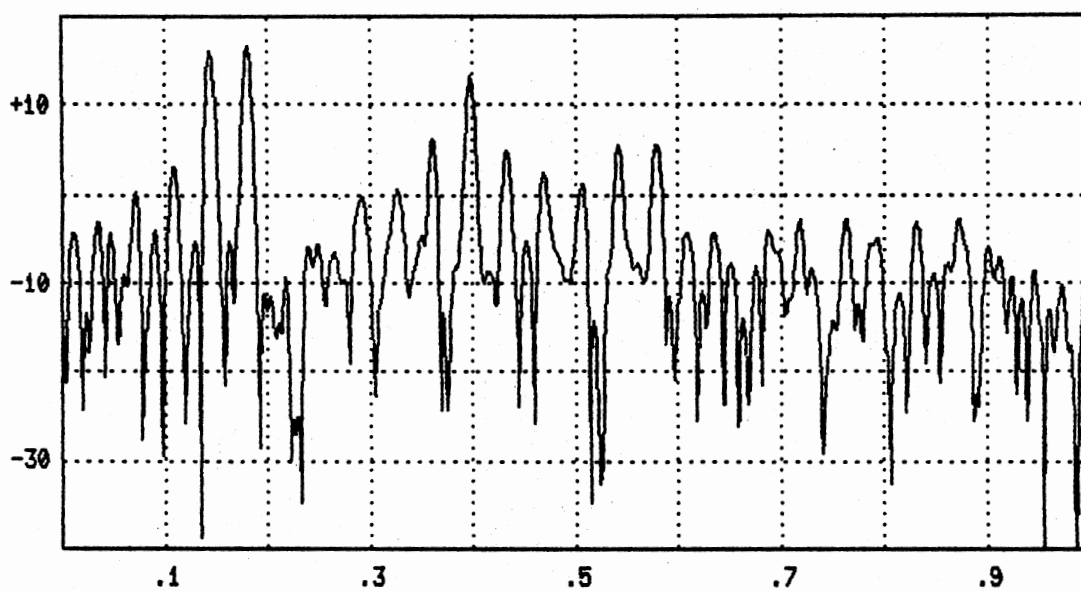
(a) Log Power Spectrum (dB vs Fractional Frequency); Periodogram Estimate of Vowel + White Noise Spectrum; SNR = 10 dB

(b) Log Power Spectrum (dB vs Fractional Frequency); AR(10) Estimate Using Autocorrelation Method; SNR = 10 dB

Figure 37. Vowel Spectrum in White Noise

Figures 38 and 39 display the result of applying various other estimators to the same white noise corrupted data frame. Figure 38 shows results obtained with the smoothed power subtraction rule and Figure 39 shows results obtained with the smoothed magnitude subtraction rule. Part a of each figure shows the result obtained with the noise filtering formulation; the noise filter response functions are displayed in part b. The weighted information estimates, displayed in part c, were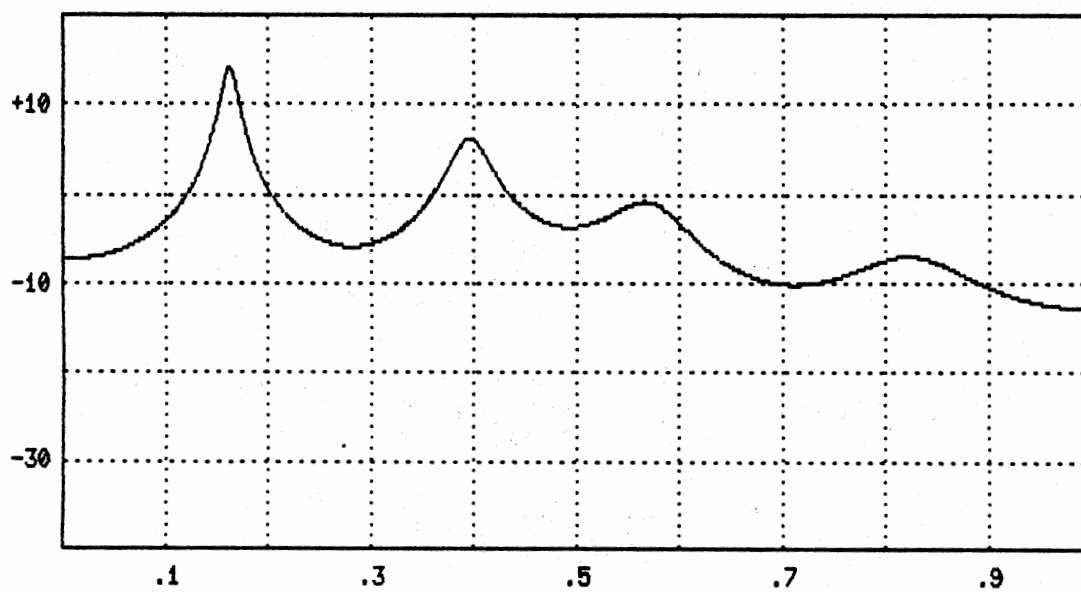 obtained using the modified Newton iteration described in Chapter IV; the weight functions, displayed in part d, were selected as an AR(6) fit to the noise filter response functions displayed in part b.[8]

Comparison of Figures 38a and 39a to Figure 36b reveals the deficiencies of these noise filtered estimates; in particular, the reader should note the amplitude of the third and fourth (highest frequency) resonance peaks as well as the depth of the valleys near the fractional frequency values of zero and one. These features are partially corrected in Figures 38c and 39c by the weighted information formulation; most notable is the correction of the valley depth near the fractional frequency value of zero. Also worth noting is the improved valley depth near the fractional frequency of one in Figure 38c and the improved amplitude of the fourth resonance peak in Figure 39c.

---

[8]The weight functions need not be selected to have an AR form; however, the author's experience with this iterative method indicates that convergence is more difficult to achieve with more complex weight function forms.
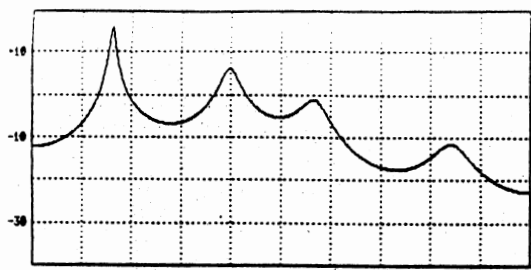
124

(a) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); AR(10) ESTIMATE USING SMOOTHED POWER SUBTRACTION RULE
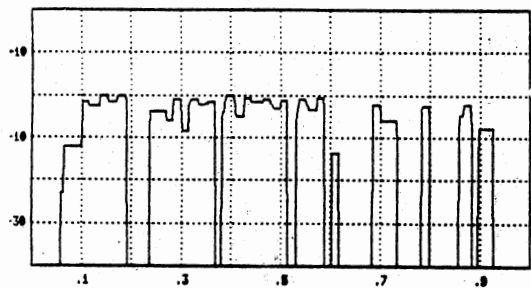
(b) SUPPRESSION FUNCTION (ATTENUATION [dB]) vs FRACTIONAL FREQUENCY; NOISE FILTER RESPONSE--SMOOTHED POWER SUBTRACTION RULE
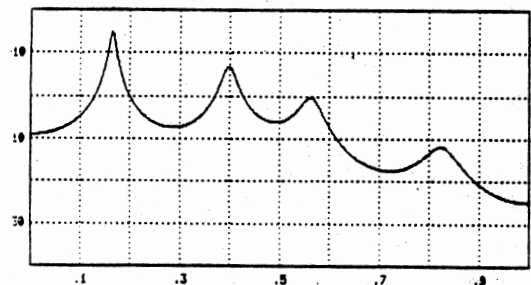
(c) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); AR(10) ESTIMATE USING AR(6) WEIGHTED INFO & SMOOTHED POW. SUB.
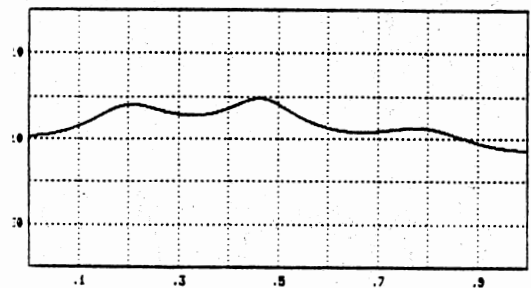
(d) WEIGHT FUNCTION (dB vs FRACTIONAL FREQUENCY); AR(6) FIT TO SMOOTHED POWER SUBTRACTION FILTER RESPONSE

Figure 38. Vowel Spectrum
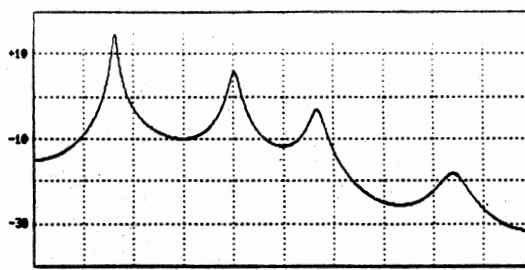
(a) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); AR(10) ESTIMATE USING SMOOTHED MAGNITUDE SUBTRACTION RULE
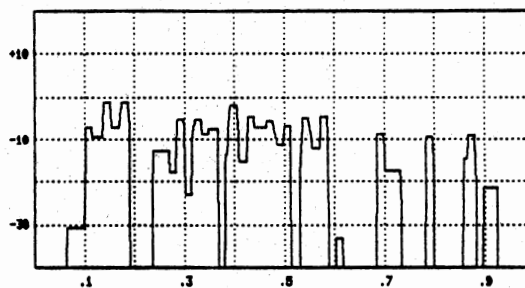
(b) SUPPRESSION FUNCTION (ATTENUATION [dB]) vs FRACTIONAL FREQUENCY; NOISE FILTER RESPONSE--SMOOTHED MAGNITUDE SUBTRACTION RULE
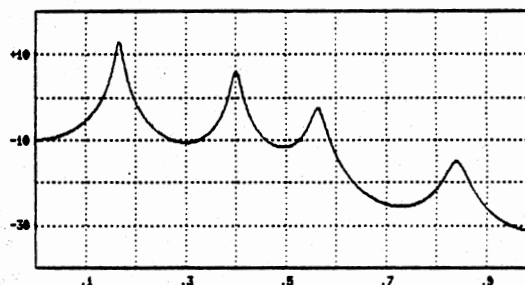
(c) LOG POWER SPECTRUM (dB vs FRACTIONAL FREQUENCY); AR(10) ESTIMATE USING AR(6) WEIGHTED INFO & SMOOTHED MAG. SUB.
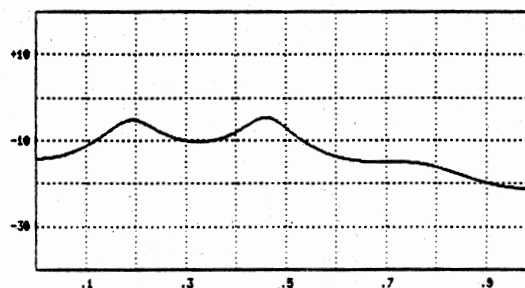
(d) WEIGHT FUNCTION (dB vs FRACTIONAL FREQUENCY); AR(6) FIT TO SMOOTHED MAGNITUDE SUBTRACTION FILTER RESPONSE

Figure 39. Vowel Spectrum

CHAPTER VI

CONCLUSION

A new method of spectral estimation has been presented. The method addresses the problem of noise corruption to the time series measurements and assumes knowledge of the noise power spectral density.[1] The method has been demonstrated to yield superior performance, in terms of reduced estimation error, and has been suggested for use in speech analysis applications.

Although the Gaussian assumption is invoked for the theoretical development of the method, examples have been provided that show the method yields superior performance for other signals as well. Similarly, the author also expects the method to be fairly robust with respect to the other assumptions.[2] It is worth noting that while the AR signal model has been assumed throughout, this assumption is by no means necessary to the theoretical development so that

---

[1]Actually, only knowledge of the frequency response of a filter designed to eliminate the noise is assumed. Knowledge of the noise power spectral density merely leads to one common method of designing such a filter.

[2]A possible exception is the assumption of independence between the signal and noise processes for it is this assumption that leads to the model of additive signal and noise power spectral densities.

other (e.g. ARMA, Pisarenko, etc.) models may also be considered.[3]

Computational procedures relevant to the problem of AR model estimation (using the weighted information formulation) have been explored. An algebraic method, applicable when the weight function assumes an AR(M) form, has been discussed; when $M \leq 4$, this method will obtain the solution using an algorithm of reasonable complexity for many applications. Iterative techniques have been discussed that obtain the solution while permitting an extremely flexible class of weight functions; the price of this greater flexibility is a considerable increase in complexity as well as the need for much user interaction. Several methods of coefficient evaluation were presented; one was implemented and used to obtain the simulation results.

The problem of AR model detection (vector quantization) requires the evaluation of two integrals for each model in the finite collection. Evaluation of the first integral is accomplished by Equation (4.58a); this equation requires the same number of additions, multiplications, and (read-only) storage locations as is required by the usual (unweighted)

---

[3]The new formulation would still require minimization of $I_W(\hat{H}f,g)$ and the analogy leading to Equation (3.20) still applies. The only difference is in the selection of a parametric signal model and the system of equations that follows. Uniqueness questions would need to be addressed separately but one may hope to find that similar convexity arguments would apply. Of course, the computational procedures discussed earlier may no longer be appropriate.

methods of vector quantization. The second integral is evaluated as a constant (independent of the data but depending upon the model) by the usual (unweighted) methods of vector quantization; Equation (4.81) is advocated for evaluation of the second integral with the weighted information formulation. With about a dozen terms, as suggested for speech analysis applications, evaluation of the second integral using Equation (4.81) is about equivalent in complexity to evaluation of the first integral.

Suggestions for Future Research

There are numerous ways to extend and refine the ideas and methods presented here. The following suggestions, offered in no particular order, are thought to be worthwhile.

- Extension to other spectral models. As mentioned earlier, the AR model form is not necessary; moreover, for some applications it may not even be appropriate.

- Assuming an AR model, determine the conditions for (and a proof of) existence. Empirical evidence for existence is strong; it is thought that the conditions are quite mild from a practical viewpoint (e.g. that the weight function is bounded). While the question of existence is mostly of theoretical interest by itself; the methods used to prove existence (and the precise

conditions for existence) should have practical value. For example, a proof based upon a contraction map is likely to yield a highly effective iterative solution procedure as well.

- Further investigation of methods of coefficient evaluation. These should be studied in close relation to the specific application in order to select a design offering a reasonable tradeoff between computational effort and performance.

- Investigation of numerical methods for solution of the ideal formulation. It is thought that the ideal formulation should yield still better performance, particularly at very low signal to noise ratios; it is expected that these methods will be very computationally expensive.

- Development of related formulations assuming a correlated noise model. The cross-spectrum (between the signal and noise processes) may be known, say, as a function of the unknown signal model spectrum and the known noise spectrum in some applications; this may occur, for example, if additive independent signal and noise processes were passed through a known nonlinear system prior to observation.

● Further investigation of computational methods appropriate for the AR weight function model; investigation of computational methods appropriate for other parametric weight function models. While the uniqueness result guarantees that only one product model, $C_{P+M}(z)$, satisfying Equations (4.13) and (4.15) has all its "additional" reflection coefficients $\{k_{P+1}, k_{P+2}, \ldots, k_{P+M}\}$ inside the interval $(-1,1)$ it is not known if the other product models satisfying these equations have all their "additional" reflection coefficients outside this interval (of course, they must have at least some of their "additional" reflection coefficients outside this interval); if this were true, the development of an efficient algorithm for higher order AR weight function models would be greatly facilitated. In general, the author believes parametric weight function models provide the greatest hope for procedures yielding a flexible choice of weight function together with an efficient solution algorithm.

● Investigation of the appropriate selection of "shape functions" in connection with use of the weighted information formulation for vector quantization, see Equations (4.79), (4.80), and (4.81). For speech analysis applications, the author envisions each shape

function as the power spectral response function of a
bandpass filter with response characteristics similar
to those filters found in "channel vocoder" systems.

- Performance evaluation in specific (speech analysis and
  other) applications using (global) measures appropriate
  to the particular application. In a voice communi-
  cations system an appropriate measure may be the result
  of some formal subjective listening test. In a recog-
  nition system the recognition error rate may be an
  appropriate measure. Systems that predict stock market
  activity might measure overall investment performance.

- Extension of the formulation to problems of multi-
  dimensional spectral estimation.

- Use of the basic concepts/ideas of the weighted infor-
  mation formulation to develop a procedure treating the
  issues of limited data and noise corruption simultane-
  ously, perhaps in combination with notions of Kalman
  filtering and the Burg algorithm.

CITED REFERENCES

1. Robinson, E.A., "A Historical Perspective of Spectrum Estimation," IEEE Proc., Special Issue on Spectral Estimation, 70, 9, September, 1982.

2. Schuster, A., "On the Investigation of Hidden Periodicities with Application to a Supposed 26-Day Period of Meterological Phenomena," Terr. Magnet., 3, 1898.

3. Yule, G.U., "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers," Phil. Trans. Roy. Soc. London, Series A, 226, 1927.

4. Walker, G., "On Periodicity in Series of Related Terms," Proc. Roy. Soc. London, Series A, 131, 1931.

5. Blackman, R.B. and Tukey, J.W., The Measurement of Power Spectra From the Point of View of Communications Engineering, reprint of the two 1958 articles (B.S.T.J., 37), New York: Dover, 1959.

6. Fourier, Jean Baptiste Joseph, The Analytical Theory of Heat, reprint of the 1878 translation from his celebrated treatise (Theorie Analytique de la Chaleur. Paris: Didot, 1822.) elaborating his controversial 1807 manuscript, New York: Dover, 1955.

7. Pisarenko, V.F., "The Retrieval of Harmonics from a Covariance Function," Geophys. J. Roy. Astron. Soc., 33, 1973.

8. Jaynes, E.T., "Information Theory and Statistical Mechanics - Part I", Phys. Rev., 106, 1957.

9. _____, "Information Theory and Statistical Mechanics - Part II", Phys. Rev., 108, 1957.

10. _____, "On the Rationale of Maximum-Entropy Methods," IEEE Proc., Special Issue on Spectral Estimation, 70, 9, September, 1982.

11. Burg, J.P., "Maximum Entropy Spectral Analysis," reprint from the Proc. 37th meeting of the Soc. of Exploration Geophys. (1967), in Modern Spectrum Analysis, Childers, D.G., Ed., New York: IEEE Press, 1978.

12. Steinhardt, A.O., "An Optimization Theoretic Framework for Spectral Estimation," Ph.D. Thesis, U. of Colorado, 1983.

13. Shore, J.E., and Johnson, R.W., "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," IEEE Trans. on IT, 26, 1, January, 1980.

14. Kullback, S., Information Theory and Statistics, New York: Wiley, 1959.

15. Shore, J.E., and Johnson, R.W., "Properties of Cross-Entropy Minimization," IEEE Trans. on IT, 27, 4, July, 1982.

16. Jeffreys, H., "An Invariant Form for the Prior Probability in Estimation Problems," Proc. Roy. Soc., A186, 1946.

17. Shore, J.E., "Minimum Cross-Entropy Spectral Analysis," IEEE Trans. on ASSP, 29, 2, April, 1981.

18. Akaike, H., "A New Look at the Statistical Model Identification," reprint from IEEE Trans. on AC, 19, 1974, in Modern Spectrum Analysis, Childers, D.G., Ed., New York: IEEE Press, 1978.

19. Parzen, E., "Some Recent Advances in Time Series Modeling," reprint from IEEE Trans. on AC, 19, 1974 in Modern Spectrum Analysis, Childers, D.G., Ed., New York: IEEE Press, 1978.

20. _____, "Time Series Model Identification by Estimating Information," Texas A&M Research Foundation, Tech. Rept. N-35, November, 1982.

21. Markel, J.D., and Gray, A.H., Jr., Linear Prediction of Speech, New York: Springer-Verlag, 1976.

22. Makhoul, J., "Stable and Efficient Lattice Methods for Linear Prediction," reprinted from IEEE Trans. on ASSP, 25, 1977 in Modern Spectrum Analysis, Childers, D.G., Ed., New York: IEEE Press, 1978.

23. Burg, J.P., "A New Analysis Technique for Time Series Data," reprinted from a 1968 presentation at the NATO Advanced Study Institute on Signal Processing

in Modern Spectrum Analysis, Childers, D.G., Ed.,
New York: IEEE Press, 1978.

24. Itakura, F. and Saito, S., "Analysis Synthesis Tele-
phony Based Upon the Maximum Likelihood Method,"
Reports of the Sixth Int. Cong. on Acoust., Y.
Kohasi, Ed., Tokyo, 1968.

25. Robinson, E.A., and Treitel, S., Geophysical Signal
Analysis, Englewood Cliffs: Prentice-Hall, 1980.

26. Kay, S.M., "Recursive Maximum Likelihood Estimation of
Autoregressive Processes," IEEE Trans. on ASSP,
31, 1, February, 1983.

27. Pinsker, M.S., Information and Information Stability of
Random Variables and Processes, translated from
the original published in Moscow (1960), San
Francisco: Holden Day, 1964.

28. Parzen, E., "Time Series Analysis for Models of Signals
Plus White Noise," in Spectral Analysis of Time
Series, Harris, B., Ed., New York: Wiley, 1966.

29. Pagano, M., "Estimation of Models of Autoregressive
Signal Plus White Noise," Ann. of Stat., 2, 1,
1974.

30. Wiener, N., Extrapolation, Interpolation, and Smoothing
of Stationary Time Series, New York: Wiley, 1949.

31. Kalman, R.E., "A New Approach to Linear Filtering and
Prediction Problems," Trans. of ASME, J. of Basic
Eng., March, 1960.

32. Lim, J.S., and Oppenheim, A.V., "Enhancement and Band-
width Compression of Noisy Speech," reprinted from
Proc. IEEE, 67, 12, 1979 in Speech Enhancement,
Lim, J.S., Ed., Englewood Cliffs: Prentice-Hall,
1983.

33. Preuss, R.D., "A Frequency Domain Noise Cancelling
Preprocessor for Narrowband Speech Communication
Systems," Proc. IEEE Int. Conf. on Acoustics,
Speech, and Signal Processing, April, 1979.

34. Berouti, M., Schwartz, R., and Makhoul, J., "Enhance-
ment of Speech Corrupted by Acoustic Noise," re-
printed from Proc. IEEE Int. Conf. on Acoustics,
Speech, and Signal Processing, April, 1979 in
Speech Enhancement, Lim, J.S., Ed., Englewood
Cliffs: Prentice-Hall, 1983.

35. Boll, S.F., "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," reprinted from IEEE Trans. on ASSP, 27, 2, 1979 in Speech Enhancement, Lim, J.S., Ed., Englewood Cliffs: Prentice-Hall, 1983.

36. McAulay, R.J. and Malpass, M.L., "Speech Enhancement Using a Soft-Decision Noise Suppression Filter," reprinted from IEEE Trans. on ASSP, 28, 2, 1980 in Speech Enhancement, Lim, J.S., Ed., Englewood Cliffs: Prentice-Hall, 1983.

37. Widrow, B., et al., "Adaptive Noise Cancelling: Principles and Applications," reprinted from Proc. IEEE, 63, 12, 1975 in Speech Enhancement, Lim, J.S., Ed., Englewood Cliffs: Prentice-Hall, 1983.

38. Sambur, M.R., "Adaptive Noise Cancelling for Speech Signals," IEEE Trans. on ASSP, 26, 5, 1978.

39. Boll, S.F., and Pulsipher, D.C., "Suppression of Acoustic Noise in Speech Using Two Microphone Adaptive Noise Cancellation," reprinted from IEEE Trans. on ASSP, 28, 6, 1980 in Speech Enhancement, Lim, J.S., Ed., Englewood Cliffs: Prentice-Hall, 1983.

40. Rife, D.C. and Boorstyn, R.R., "Multiple tone parameter estimation from discrete time observations," Bell Syst. Tech. J., 55, 9, 1976.

41. Tufts, D.W. and Kumaresan, R., "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood," Proc. IEEE, Special Issue on Spectral Estimation, 70, 9, September, 1982.

42. Cadzow, J.A., Baseghi, B., and Hsu, T., "Singular-value decomposition approach to time series modelling," IEE Proc., Part F., Special issue on Spectral Analysis, 130, 3, 1983.

43. Cramer, H., Mathematical Methods of Statistics, Princeton: Princeton University Press, 1946.

44. Kaveh, M., and Bruzzone, S.P., "Statistical efficiency of correlation-based methods for ARMA spectral estimation," IEE Proc., Part F., Special issue on Spectral Analysis, 130, 3, 1983.

45. Quirk, M. and Liu, B., "Improving Resolution For Autoregressive Spectral Estimation By Decimation," IEEE Trans. on ASSP, 31, 3, June, 1983.

46. Nehorai, A. and Morf, M., "Enhancement of Sinusoids in Colored Noise and Whitening Performance of Exact Least Squares Predictors," IEEE Trans. on ASSP, 30, 3, June, 1982.

47. Wong, D.Y., Juang, B.H., and Gray, A.H., "An 800 bit/s Vector Quantization LPC Vocoder," IEEE Trans. on ASSP, 30, 5, 1982.

48. Abut, H., Gray, R.M., and Rebolledo, G., "Vector Quantization of Speech and Speech-Like Waveforms," IEEE Trans. on ASSP, 30, 3, 1982.

49. Gray, R.M., Gray, A.H., Rebolledo, G., and Shore, J.E., "Rate-distortion speech coding with a minimum discrimination information distortion measure," IEEE Trans. on IT, 27, 6, 1981.

50. Buzo, A., Gray, A.H., Gray, R.M., and Markel, J.D., "Speech Coding Based Upon Vector Quantization," IEEE Trans. on ASSP, 28, 5, 1980.

51. Linde, Y., Buzo, A., and Gray, R.M., "An Algorithm for Vector Quantizer Design," IEEE Trans. on COM, 28, 1, 1980.

52. Buzo de la Pena, L.A., "Optimal Vector Quantization for Linear Predictive Coded Speech," Ph.D. thesis, Stanford University, 1978.

53. Suresh Babu, B.N., Volz, B.E., Washburn, S.J., and Preuss, R.D., "Software for Vector Quantization," MTR-8414, Bedford, Mass.: The MITRE Corp., 1981.

54. Clapp, R.A.J., "Acoustic Segments in Natural Speech: Analysis and Statistics," MTR-8405, Bedford, Mass." The MITRE Corp., 1981.

55. Chu, P.L., "Frequency Weighted Linear Predictive Coding of Speech," Ph.D. Thesis, University of Calif., Berkeley, July, 1981.

56. _____, and Messerschmitt, D.G., "A Frequency Weighted Itakura-Saito Spectral Distance Measure," IEEE Trans. On ASSP, 30, 4, 1982.

57. Levinson, N., "The Wiener RMS Error Criteria in Filter Design and Prediction," J. Math. Phys., 25, pp. 261-78, 1947.

58. Durbin, J., "The Fitting of Time-Series Models," Rev. Inst. Int. Statist., 28, 3, pp. 233-43, 1960.

59. Burnside, W.S. and Panton, A.W., _The Theory of Equations_, 7th Ed., London: Longmans, Green, & Co., 1912. Also available as a reprint, New York: Dover, 1960

60. Uspensky, J.V., _Theory of Equations_, New York: McGraw-Hill, 1948.

61. Collatz, L., _Functional Analysis and Numerical Mathematics_, translated from the German by Oser, H., New York: Academic Press, 1966.

62. Hille, E., _Analytic Function Theory_, Vol. 1, Boston: Ginn and Company, 1959.

λ

VITA

Robert David Preuss

Candidate for the Degree of

Doctor of Philosophy

Thesis: AUTOREGRESSIVE SPECTRAL ESTIMATION IN NOISE WITH APPLICATION TO SPEECH ANALYSIS

Major Field: Electrical Engineering

Biographical:

Personal Data: Born in Teaneck, New Jersey, December 29, 1953, the son of Harold Paul Preuss and Phyllis Burnap Brooks.

Education: Graduated from Fairmont West High School, Kettering, Ohio, in June, 1971; received Bachelor of Arts Degree, Cum Laude, in Mathematics from Boston University in May, 1975; received Master of Science Degree in Applied Mechanics from Boston University in September, 1977; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in December, 1983.

Professional Experience: Teaching Assistant, Department of Mathematics, Boston University, September, 1974 to May, 1975; Resident Tutor, Boston University, September, 1975 to May, 1976; Teaching Assistant, Department of Aerospace and Mechanical Engineering, Boston University, September, 1976 to May, 1977; Research Assistant, College of Engineering, Boston University, July, 1975 to April, 1978; Technical Consultant, Aerospace Systems Inc., Burlington, Massachusetts, August, 1977 to April, 1978; Project Leader/Member Technical Staff, The MITRE Corporation, Bedford, Massachusetts, April, 1978 to August, 1982; Teaching Assistant, School of Electrical and Computer Engineering, Oklahoma State University, August, 1982, to May, 1983; Research Associate, School of Electrical and Computer Engineering, Oklahoma State University, August, 1982 to August, 1983.

Awards/Affiliations: National Aeronautics and Space
     Administration Certificate of Recognition, March,
     1982; Member American Institute of Aeronautics and
     Astronautics since December, 1977; Member Eta
     Kappa Nu (Omega Chapter) since March, 1983;
     Student Member Institute of Electrical and
     Electronic Engineers (Acoustics, Speech, and
     Signal Processing Group and Information Theory
     Group) since October, 1982.