

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/2996163>

Spectrum Estimation and Harmonic Analysis

Article *in* Proceedings of the IEEE · October 1982

DOI: 10.1109/PROC.1982.12433 · Source: IEEE Xplore

CITATIONS

2,367

READS

566

1 author:



David James Thomson

Queen's University

168 PUBLICATIONS 5,900 CITATIONS

SEE PROFILE

All content following this page was uploaded by [David James Thomson](#) on 15 May 2015.

The user has requested enhancement of the downloaded file. All in-text references [underlined in blue](#) are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.

Spectrum Estimation and Harmonic Analysis

DAVID J. THOMSON, MEMBER, IEEE

Invited Paper

Abstract—In the choice of an estimator for the spectrum of a stationary time series from a finite sample of the process, the problems of bias control and consistency, or “smoothing,” are dominant.

In this paper we present a new method based on a “local” eigen-expansion to estimate the spectrum in terms of the solution of an integral equation. Computationally this method is equivalent to using the weighted average of a series of direct-spectrum estimates based on orthogonal data windows (discrete prolate spheroidal sequences) to treat both the bias and smoothing problems.

Some of the attractive features of this estimate are: there are no arbitrary windows; it is a small sample theory; it is consistent; it provides an analysis-of-variance test for line components; and it has high resolution.

We also show relations of this estimate to maximum-likelihood estimates, show that the *estimation capacity* of the estimate is high, and show applications to coherence and polyspectrum estimates.

I. INTRODUCTION

A MAJOR PROBLEM in time series analysis is choosing an algorithm to estimate the spectrum from a finite observation of the process in such a way that the estimate is not dominated by bias, is consistent and statistically meaningful, and maintains these properties in the presence of minor variations of assumptions. Our emphasis is on the case where the data available are a finite sample from an almost stationary ergodic process containing relatively few outliers. We assume that the range of the spectrum may be large and that the spectrum may contain line components in addition to a continuous background. In addition, we are interested primarily in *nonparametric* estimates as opposed to those where a specific functional form is assumed. For such cases, the procedures described in Thomson [324], and Kleiner *et al.* [190], [191] work well. It should be noted, however, that these techniques are heuristic and that, despite its long history, the “best” existing solutions to the spectrum estimation problem are still not completely satisfactory. In particular, when the series is short,¹ the spectrum is mixed, or the range of the spectrum is large, problems are likely. If all three of these conditions are true, problems are guaranteed. In the following, a new method is developed which gives a more efficient solution to such problems.

In addition to the basic estimation procedure, we show its applicability to estimating coherence and polyspectra. In this paper we also define *estimation capacity* which is a logarithmic information measure and show that the estimates proposed here have a much higher capacity than do estimates based on the sample autocovariances. We also show that the proposed estimates have a high likelihood and some connections between

these estimates, maximum likelihood, and extrapolation estimates.

In the case where “outliers” or missing values are present, we assume that their influence will be controlled by a “robust filter” in an iterative modeling and filtering approach of the type described in the papers mentioned above. Since this paper is addressed to the modeling aspects of the overall problem, we will assume that the data are nearly Gaussian and are exactly so for variance expressions. We also assume that the data are a finite sequence of samples, equally spaced in time, and that the computations will be done digitally.

A. Existing Nonparametric Estimates

Traditionally, nonparametric spectrum estimates have been divided into two classes, direct and indirect. Of the two, the direct estimates are older, dating to Schuster’s periodogram [293]. Because of the computational burden imposed by direct estimates before the discovery of the fast Fourier transform [79] and also by analogy with classical multivariate statistics, indirect estimates, i.e., those based on estimates of the autocovariances of the series, were commonly used from the work of Bartlett [28], [29], Parzen [246]–[248], Blackman and Tukey [41], and are still occasionally used. However, since the autocovariances may be obtained as the discrete Fourier transform of the extended periodogram, for the purposes of this discussion we assume the following steps: *first*, forming a direct spectrum estimate

$$\hat{S}_D(\omega) = \left| \sum_{n=0}^{N-1} x(n) D_n e^{-i\omega n} \right|^2$$

at radian frequency ω by tapering the current data sequence $x(n)$ (typically either the raw data or the residuals from a robust prewhitening operation) with a *data window* D_n , transforming, squaring; and *second*, because $\hat{S}(\omega)$ is inconsistent in the sense that its variance does not decrease with sample size, *smoothing* it, typically by convolution with a second window $G(\omega)$, giving the smoothed spectrum estimate,

$$\tilde{S}(\omega) = \hat{S}_D(\omega) * G(\omega).$$

Implicit in this equation is the connection with indirect estimates: do the convolution by multiplying the sample autocovariances (the transform of $\hat{S}_D(f)$) by the lag window corresponding to G in the time domain. Since the data window D primarily controls bias while the smoothing primarily effects variance, the two operations are usually considered to be unrelated. (Note that, for a given data window D , the optimum smoothing window G may be obtained by the methods of Papoulis [243].)

Both operations pose problems. In the direct estimate, the use of a data window is essential (Brillinger [58]). If the data

Manuscript received May 17, 1982.

The author is with Bell Laboratories, Whippany, NJ 07981.

¹A “short” series is one where the resolution required is of the same order as the reciprocal series length. If the true spectrum is complex, the number of data points in a “short” series may still be large.

are unwindowed (all $D_n = \text{constant}$) or, equivalently, if $\hat{S}_D(\omega)$ is based on the sample autocorrelations, the estimate is likely to be too badly biased to be useful. Conversely, when a data window is used, bias is reduced but so is the variance efficiency. One may also be distressed by the thought that a data window weights equally valid data differently. This dilemma has created considerable controversy [58], [235], [360].

In a similar way, the smoothing operation is unsatisfactory unless there is reason to believe that the underlying spectrum is smooth. If, however, as appears to be the more typical case, the true spectrum is "mixed," that is, it contains line components on a smooth background, acceptable "smoothers" are nonlinear. Since these smoothers operate on the raw spectrum estimate, phase information present in the original data is not used and, consequently, the line detection operation is much less efficient than it should be.

As a replacement for the two independent estimation stages described above, we propose a unified algorithm having several interesting features: first, it is a small sample theory with the sample size entering explicitly into the methods and performance bounds; second, it justifies the use of data windows; third, the estimate is consistent; fourth, the procedure is data adaptive and, in difficult situations where the range of the spectrum is large, will give more stable estimates in regions where the spectrum is large without being excessively biased where it is low; fifth, it provides an analysis of variance test for line components (including the process mean); and sixth, for multivariate data, it results in new *classes* of estimates. As a particular example of the latter, the technique results in two distinct estimates of coherence, one for line components, one for the continuum. In addition, these estimates are closely related to maximum-likelihood procedures. We also give an example showing their utility for analyzing nonstationary data. In the following sections we define the basic estimation and the adaptive weighting procedures. (Earlier versions of this method appear in [325], [326].)

B. Notation

We assume that the data consist of N contiguous samples, $x(0), x(1), \dots, x(N-1)$, which are an observation from a stationary, real, ergodic, zero-mean, Gaussian time series. The sample size N is supposed to be finite and typically "small." For notational convenience we shall generally write Fourier transforms with the observation epoch centered at the time origin. We assume that the time between successive samples is 1 so that frequency f and radian frequency $\omega = 2\pi f$ are defined on their principal domains $(-\frac{1}{2}, \frac{1}{2})$ and $(-\pi, \pi]$, respectively. Boldface letters are used for vectors and matrices with components given by the corresponding italics, superscript * indicates complex conjugate, online * denotes convolution, superscript † conjugate transpose, and & denotes the expected value operator. We denote the true spectrum of the *sampled* process by S , including possible aliasing effects. For processes intrinsically defined in continuous time, we assume that adequate antialiasing filters and sufficient resolution and sampling rate have been used, so that the spectrum of the sampled process reasonably approximates the original over the Nyquist band.

C. Outline of the Estimation Procedure

We begin with the general Cramér spectral representation for a stationary process

$$x(n) = \int_{-1/2}^{1/2} e^{i2\pi v[n-(N-1)/2]} dZ(v)$$

in which $dZ(f)$ is a zero-mean orthogonal increment process. $dZ(f)$ is related to the spectrum $S(f)$ by definition

$$S(f) df = \mathbb{E}\{|dZ(f)|^2\}.$$

The problem of spectrum analysis is that of estimating the statistical properties, particularly the moments, of $dZ(f)$ from the *finite* sample $x(0), \dots, x(N-1)$.

In the time domain, to say that the sample $\{x(t)\}; t = 0, \dots, N-1$ represents a projection from the infinite sequence $\{x(t)\}$ generated by $dZ(f)$ is trite; in the frequency domain the same expression has some profound implications. Since we are interested in the properties of a frequency-domain entity, it is natural to begin with the Fourier transform $y(f)$ of the observations

$$y(f) = \sum_{n=0}^{N-1} e^{-i2\pi f[n-(N-1)/2]} x(n).$$

Using the spectral representation in place of the data in the finite discrete Fourier transform gives the fundamental equation of spectrum estimation

$$y(f) = \int_{-1/2}^{1/2} \frac{\sin N\pi(f-v)}{\sin \pi(f-v)} dZ(v)$$

as the equation expressing the projection from $dZ(v)$ onto $y(f)$ in the frequency domain. We will treat it as a linear Fredholm integral equation of the first kind.

The problem considered in this paper is the approximate solution of this equation, spectrum estimates based on these approximate solutions, their sampling properties, and their relation to other spectrum estimation procedures.

Using the integral equation approach we adopt a weighted eigenfunction expansion for its "solution" in the locality $(f_0 - W, f_0 + W)$ of some frequency of interest f_0 . The equation and general considerations leading to this decision are discussed in Section II. We also summarize some properties of the eigenfunctions (discrete prolate spheroidal wave functions) which satisfy the integral equation

$$\lambda_k(N, W) \cdot U_k(N, W; f) = \int_{-W}^W \frac{\sin N\pi(f-f')}{\sin \pi(f-f')} U_k(N, W; f') df'.$$

These are described in Section II-A.

Having established these preliminaries and notation, the basic solution technique is given in Section III. This solution results in the local high-resolution estimate

$$\hat{S}(f; f_0) = \left| \sum_{k=0}^{K-1} U_k(N, W; f-f_0) w_k(f_0) y_k(f_0) \right|^2$$

where the expansion coefficients are given by

$$y_k(f_0) = \int_{-1/2}^{1/2} U_k(N, W; f) y(f-f_0) df.$$

These may be simply computed using the fast Fourier transform

$$y_k(f) = \sum_{n=0}^{N-1} x(n) \cdot \frac{v_n^{(k)}(N, W)}{\epsilon_k} e^{-i2\pi f[n-(N-1)/2]}$$

of the data, windowed by the discrete prolate spheroidal sequences, $v_n^{(k)}(N, W)$. Based on the moments of these estimates (Section IV), the coefficient weights $d_k(f)$, necessary to obtain

a convergent solution, are described in Section V. Section VI consists of an example of the estimation procedures discussed in Sections II through V including plots of the data, individual eigenspectrum estimates $|y_k(f)|^2$, weights, and the stabilized estimate

$$\bar{S}(f_0) = \sum_{k=0}^{K-1} |d_k(f_0) \cdot y_k(f_0)|^2.$$

Section VII presents further sampling properties and some efficiency calculations. In Section VIII we show a relation between the eigenspectrum estimates and the periodogram.

Section IX is addressed to the general problem of the efficiency in spectrum estimation. Using mutual information concepts, we define *estimation capacity* and show that the estimates based on prolate spheroidal wave functions are very good in this respect.

Section X presents a new high-resolution estimate based on a free parameter expansion. Section XI is concerned with the characteristics of frequency-translated prolate spheroidal wave functions as basis sets and properties of basis sets suitable for spectrum estimation. In the next Section, XII, we show a close relation between prolate spheroidal wave functions, Karhunen-Loëve expansions, and maximum-likelihood spectrum estimates. We also show a general double orthogonality property and some relations between these and extrapolation estimates.

In Section XIII we discuss some aspects of harmonic analysis for which these estimates are particularly well suited. This includes a new analysis of variance test for line components and some results on resolution.

The subject of Section XIV is coherence and, by similarity, polyspectra. (In both, one attempts to estimate cross moments: in coherence, between different series; in polyspectra, between different frequencies. Both are subject to the same problems resulting from rapid phase changes.) Again, new classes of estimates are obtained. Among these is a technique for identifying related frequency components in nonstationary data. Section XV is a brief summary and a reminder of the place of this theory in the larger problem.

Because the literature applicable to spectrum estimation is so immense, it is very difficult to give complete references. There are many general references: [4], [14], [31], [43], [51], [56], [57], [74], [81], [83], [86], [95], [111], [112], [137], [143], [151], [155], [168], [180], [185], [187], [196], [228], [242], [248], [265], [353].

As a final introductory point we mention the range of the spectrum. I have frequently been told by time series analysts that spectra with ranges of over 40 or 50 dB approach the pathological. In contrast, my personal experience has been that when data are carefully collected and analyzed, spectra from physical origins rarely have less than 50-dB range. I have also experienced some situations when perfectly reasonable communications problems led to a desire to estimate spectra with ranges of from 160 to over 200 dB. In most of these cases, the sampling rates required prohibited digital analysis; however, it is now possible to buy commercial digitizers with analog bandwidths of 1 GHz, and while quantization accuracy is still a limitation for many problems, there are others where the limitation is the range of the algorithm.

II. THE BASIC INTEGRAL EQUATION

The basic motivation for studying the power spectrum of a process is that any stationary process has a Cramér spectral

representation

$$x(t) = \int_{-1/2}^{1/2} e^{i2\pi ft} dZ(f)$$

for all t . The random orthogonal-increments measure $dZ(f)$ has, for zero-mean processes

$$\mathbb{E}\{dZ(f)\} = 0.$$

Its second moment, the *power spectral density* or simply the *spectrum* $S(f)$ of the process is defined by

$$S(f) df = \mathbb{E}\{|dZ(f)|^2\}.$$

This definition defines our problem—estimation of the statistical properties, in particular the moments, of $dZ(f)$.

While details of this representation are available in [82], [95], [193], [286], and [351], it should be recalled that $dZ(v)$ is an *orthogonal increment process*, that is, for distinct frequencies, f and v , $dZ(f)$ and $dZ^*(v)$ are statistically uncorrelated. (Note that uncorrelated does not imply independence as $dZ(f) = dZ^*(-f)$ for real processes.) For notational simplicity, it is convenient to translate the time origin to the center of the observation epoch and, changing the definition of $dZ(f)$ by a phase factor, to write

$$x(t) = \int_{-1/2}^{1/2} e^{i2\pi v[t-(N-1)/2]} dZ(v). \quad (2.1)$$

Since we wish to estimate the statistics of $dZ(f)$ from the sample of N contiguous observations, $x(0), x(1), \dots, x(N-1)$, we transform to the frequency domain using the finite discrete Fourier transform defined, again for notational convenience, in time-centered form

$$y(f) = \sum_{t=0}^{N-1} e^{-i2\pi f[t-(N-1)/2]} x(t). \quad (2.2)$$

In these transforms we consider frequency to be a continuous parameter with principal domain $(-\frac{1}{2}, \frac{1}{2}]$ and functions of frequency to be periodically extended outside this domain. Note carefully that, since $y(f)$ may be inverted to recover the data

$$x(t) = \int_{-1/2}^{1/2} e^{i2\pi f[n-(N-1)/2]} y(f) df$$

it constitutes a trivially sufficient statistic and, hence, no information is lost by the transform operation. Because of this equivalence, we shall use either $\{x(t)\}$ or $y(f)$ interchangeably as "data."

Combining the preceding two equations gives

$$y(f) = \int_{-1/2}^{1/2} \sum_{t=0}^{N-1} e^{i2\pi(v-f)[t-(N-1)/2]} dZ(v)$$

from which, on recognizing the sum as the Dirichlet kernel

$$\frac{\sin N\pi(f-v)}{\sin \pi(f-v)} = \sum_{t=0}^{N-1} e^{i2\pi(v-f)[t-(N-1)/2]}$$

one arrives at the equation

$$y(f) = \int_{-1/2}^{1/2} \frac{\sin N\pi(f-v)}{\sin \pi(f-v)} dZ(v). \quad (2.3)$$

We consider this to be the basic equation of spectrum estimation.

The most obvious interpretation of this equation is as a convolution describing the "window leakage," "smearing," or "frequency mixing," which is a consequence of using the *finite* Fourier transform. As a result of this effect, there is no obvious reason to expect the statistics of $y(f)$ to resemble those of $dZ(f)$. It should also be noted that, unlike $y(f)$, the basic periodogram,² $P_x(f) = |y(f)|^2$, is *not* a sufficient statistic for the data, which implies that the phase information abandoned in periodogram-based estimates is essential [237], [267], [322]. Consequently, the periodogram is a poor choice as a starting point for any serious data analysis technique. While problems with the periodogram are well known [30], [94], [173], [271], etc., the importance of this equation is such that it merits further attention and we make the following observations:

1) The insufficiency of the periodogram is clearly inherited by any estimate based on or equivalent to the periodogram. This obviously includes both smoothed periodograms (and it is irrelevant if the smoothing is done directly on the periodogram, on the log periodogram, or by fitting a spline or rational polynomial to it), and, because the transform of the periodogram is the sample autocovariance function, autoregressions, moving-average representations, and other decompositions based on sample autocovariances. In addition, deconvolution methods based on the periodogram or autocovariances are *intrinsically* more difficult, owing to the eigenvalue behavior of the sinc^2 kernel [129].³

2) The problem has much in common with the classical statistical general linear model [269], [292], [337]

$$y = X'\beta + e$$

where y represents the observations, X the model, β the coefficients to be estimated, and e the error between the hypothesized model and the observations. In the spectrum estimation case, the Fourier transform corresponds to the observations, the model is specified by the Dirichlet kernel, and the model coefficients generate the spectrum estimate. In classical regression and analysis of variance problems, the approach is normally *first*, to solve the equations (either by least squares or some other approximation technique) and *second*, to examine the statistics of the estimated coefficients. Judging from the number of papers published on periodogram-equivalent estimates, it is apparent that, for spectrum estimation problems, a different approach has been fashionable: *first*, square the observations (i.e., compute the periodogram); *second*, ignore the model; and *third*, use the squared observations as the solution, which is then possibly tested for significance. If we specialize the linear model analogy to a simple regression problem, what we have done is equivalent to using the data themselves for the regression line. If the signal-to-noise ratio is high enough, this may be a good approximation to the line but gives no information about the coefficients. Further, on the logarithmic scale necessary for spectrum estimation, the finite sample Dirichlet kernel is very different from a Dirac delta

² Both because it is useful and for historical reasons we reserve use of term *periodogram* to mean the magnitude-squared Fourier transform of the *unwindowed*, or rectangular windowed, function. When a non-uniform data window is involved, we refer to the squared magnitude of the finite Fourier transform of the data times window as either a *windowed periodogram* or as a *direct spectrum estimate*.

³ $\text{sinc } x = \sin \pi x / \pi x$. The sinc kernel is the continuous-time equivalent of the Dirichlet kernel used here.

function (which it approaches asymptotically) and, consequently, the approach based on using the model and solving the resulting equations has some appeal.

In this context one must emphasize that, for processes with spectra typical of those encountered in engineering, the sample size must be extraordinarily large for the periodogram to be reasonably unbiased. While it is not clear what sample size, if any, gives reasonably valid results, in my experience periodogram estimates computed using 1.2 million data points on the WT4 waveguide project, see [1], were too badly biased to be useful. The best that could be said for them is that they were so obviously incorrect as not to be dangerously misleading. In other applications where less is known about the process, such errors may not be so obvious. Thus while the estimates described in this paper are certainly more difficult and expensive to compute than the periodogram, this expense must be weighed against the cost of a wrong answer.

3) The kernel has been occasionally referred to as a Fejer kernel although this terminology is better used for its square.

4) The integral equation (2.3) appears in [56]. In addition similar equations appear in numerous optical and radar inverse problems. Because this kernel serves as the identity element or reproducing kernel in the space of Fourier transforms of index limited sequences, [19], [233], [250] are relevant.

A. An Alternative Viewpoint

A more constructive viewpoint is to regard (2.3) as a linear Fredholm integral equation of the first kind for $dZ(v)$ with the goal of obtaining approximate solutions whose statistical properties are, in some sense, "close" to those of $dZ(f)$. Since this equation is the frequency-domain expression of the projection from the infinite stationary sequence generated by the random orthogonal measure $dZ(f)$ onto the finite sample, it does not have an inverse; hence it is impossible to obtain exact or unique solutions. What we desire are the statistics of those approximate solutions that are both statistically and numerically plausible. Throughout this procedure, one must bear in mind that the essential problem of spectrum analysis is to estimate the statistical properties of $dZ(v)$ as opposed to those of $y(f)$. Despite the effort spent on the statistical properties of $y(f)$, e.g., periodogram-based spectrum estimates, it is not clear that they are often of much interest.

While the indeterminacy of the basic integral equation prohibits exact solutions, several approximate solutions have been defined. For this purpose, numerous methods and criteria have been proposed:

1) Regularization methods, such as Tikhonov's, which add a mean-square-curvature constraint, and other minimum mean-square-error methods [22], [116], [149], [254], [255], [317]. Proust and Goutte [266] use prior information to convert to a Fredholm equation of the second kind.

2) Methods dependent on explicit representations: for example, the sampling theorem [154]; prolate spheroidal wave functions [26], [61], [117], [287], [356]; and on-spline representations, [6], [96], [97]. Sjøntoft [300] uses a Taylor expansion of the reciprocal kernel.

3) Iterative methods, for example [171], [184], [259]; and iterative extrapolation techniques [122], [221] and further references given in Section XII.

4) Methods dependent on specific time-series representations and properties [89], [106], [253], [278], [279].

5) The problem is also closely related to ridge regression problems [105], [236].

6) References [141], [169] and those cited in Section X

contain general information. See also [23], [63], [92], [315] for recent numerical work.

The choice of a technique for computing approximate solutions depends primarily on which characteristics are desired in the solution and, while these are to some extent subjective, they are as follows:

1) The solution should be "local," that is the estimated spectrum at one frequency should not depend strongly on details of the spectrum at "distant" frequencies; see [12], [361]. Philosophically similar ideas are used for subsection moment solutions in electromagnetic theory [150] and in spline theory [49], [85]; see also [305].

2) The solution should be easy to compute. This implies numerical stability and closed-form or convergent iterative procedures. To characterize "easy" see [104].

3) It must be possible to characterize the statistics of the solution.

4) Because spectrum estimates are commonly an intermediate step in a complete analysis which may involve subsequent design of filters or predictors based on the estimated spectrum, both the estimated spectrum *and its logarithm* should be "good" estimates. (Recall, for example, that the one-step prediction variance is $\exp[\int \ln S(f) df]$.) See [175], [251], [252], [343]. Note that this requirement on the logarithm of the spectrum effectively precludes "unbiased" estimates based on equivalent lag windows, as such estimates may be negative.

Of the methods mentioned above, the most successful approaches have been eigenfunction expansions combined with a least squares error criterion. We adopt this approach with a *local* least squares error criterion. Recall that, formally, solutions of integral equations of the type

$$y = K * z$$

where K is a kernel with eigenfunctions

$$\lambda_m \psi_m = K * \psi_m$$

standardized by

$$\int \psi_m \psi_n = \delta_{n,m}$$

is given by

$$\hat{z} = \sum_m \frac{1}{\lambda_m} y_m \psi_m$$

where the sum is over the set of eigenfunctions and the expansion coefficients y_m are given by the usual Fourier-Bessel formula

$$y_m = \int y \psi_m$$

(see [311], [330], [334] for further information). Since the eigenvalues typically decay exponentially [240], such solutions are of no practical use. Consequently, the class of realizable solutions is restricted to those corresponding to "large" eigenvalues. These may be obtained either by simply truncating the sum, or, as will be done in Section V, by weighting the expansion coefficients so that the solution may be written

$$\hat{z} = \sum_{m=0} D(\lambda_m, y_m) \cdot \psi_m$$

where $D(\lambda_m, y_m)$ is a weight function.

In a remarkable series of papers, of which the most recent is Slepian [306], it has been shown that the eigenfunctions of

the Dirichlet kernel (and its continuous time counterpart), known as *prolate spheroidal wave functions*, are fundamental to the study of time- and frequency-limited systems. We thus contemplate "solving" the integral equation in some local interval about f , say $(f - W, f + W)$, using discrete prolate spheroidal wave functions as a basis. We shall refer to this interval as the *local* or *interior* domain and the remainder of the principal frequency domain as *exterior*.

B. Background: Discrete Prolate Spheroidal Wave Functions and Sequences

In this section we give a short list of formulas and properties of these eigenfunctions from Slepian's paper. The eigenfunctions, denoted by $U_k(N, W; f)$, $k = 0, 1, \dots, N-1$ are known as *discrete prolate spheroidal wave functions* (DPSWF) and are solutions of the equation

$$\int_{-W}^W \frac{\sin N\pi(f-f')}{\sin \pi(f-f')} U_k(N, W; f') df' = \lambda_k(N, W) \cdot U_k(N, W; f) \quad (2.4)$$

where W , $0 < W < \frac{1}{2}$ is the bandwidth defining "local" and here is normally of the order $1/N$. The functions are ordered by their eigenvalues

$$1 > \lambda_0(N, W) > \lambda_1(N, W) > \dots > \lambda_{N-1}(N, W) > 0.$$

The first $2NW$ eigenvalues are extremely close to 1 and, of particular relevance here, *of all functions which are the Fourier transform of an indexlimited sequence the discrete prolate spheroidal wave function, $U_0(N, W; f)$ has the greatest fractional energy concentration in $(-W, W)$* . For small k and large N the degree of this concentration is given by Slepian's asymptotic expression (in slightly different notation)

$$1 - \lambda_k(N, W) \sim \frac{\sqrt{2\pi}}{k!} \left[\frac{8N \sin \pi W}{\cos^2 \pi W} \right]^{k+1/2} \left[\frac{1 - \sin \pi W}{1 + \sin \pi W} \right]^N$$

or, for larger N with $N\pi W = c$, and $k < [2NW]$

$$1 - \lambda_k(N, W) \sim \frac{\sqrt{2\pi}}{k!} (8c)^{k+1/2} e^{-2c}. \quad (2.5)$$

From a conventional spectrum estimation viewpoint, this expression gives the fraction of the total energy of the spectral window outside the main lobe (i.e., outside $(-W, W)$).

The eigenfunctions, $U_k(N, W; f)$, $k = 0, 1, \dots, N-1$ are *doubly orthogonal*, that is they are orthogonal over $(-W, W)$

$$\frac{1}{\lambda_k(N, W)} \int_{-W}^W U_j(N, W; f) \cdot U_k(N, W; f) df = \delta_{j,k} \quad (2.6)$$

and orthonormal over $(-\frac{1}{2}, \frac{1}{2})$.

$$\int_{-1/2}^{1/2} U_j(N, W; f) \cdot U_k(N, W; f) df = \delta_{j,k}. \quad (2.7)$$

The Fourier transforms of the discrete prolate spheroidal wave functions are known as *discrete prolate spheroidal sequences* (DPSS),

$$v_n^{(k)}(N, W) = \frac{1}{\epsilon_k \lambda_k(N, W)} \int_{-W}^W U_k(N, W; f) e^{-i2\pi f[n-(N-1)/2]} df \quad (2.8)$$

valid for $k = 0, 1, \dots, N-1$ and all n , ϵ_k is 1 for k even and

i for k odd. Because of the double orthogonality, there is a second Fourier transform

$$v_n^{(k)}(N, W) = \frac{1}{\epsilon_k} \int_{-1/2}^{1/2} U_k(N, W; f) e^{-i2\pi f[n-(N-1)/2]} df$$

valid for both $n, k = 0, 1, \dots, N-1$. As one would expect, the finite discrete Fourier transform of the prolate spheroidal sequence results in the discrete prolate spheroidal wave functions

$$U_k(N, W; f) = \epsilon_k \sum_{n=0}^{N-1} v_n^{(k)}(N, W) e^{i2\pi f[n-(N-1)/2]}.$$

It should also be noted that the discrete prolate spheroidal sequences satisfy a Toeplitz matrix eigenvalue equation

$$\sum_{m=0}^{N-1} \frac{\sin 2\pi W(n-m)}{\pi(n-m)} \cdot v_m^{(k)}(N, W) = \lambda_k(N, W) \cdot v_n^{(k)}(N, W) \quad (2.9)$$

and so are easily computed for moderate values of N . Like the wave functions the $\{v_n^{(k)}\}$ are doubly orthogonal, being orthogonal on $(-\infty, \infty)$ and orthonormal on $[0, N-1]$. Detailed information on these functions is contained in Slepian's 1978 paper [306] and in [114], [201], [202], [245], [280], [303]-[308], [335], [358]. A method for computing the functions for larger values of N is given in the Appendix.

III. EIGENESTIMATES

Using the background material outlined in the preceding two sections, we now construct estimates of the spectrum from approximate solutions of the integral equation (2.3). We begin by considering the discrete prolate spheroidal wave function Fourier-Bessel expansion coefficients of dZ in $(f - W, f + W)$

$$Z_k(f) = \frac{1}{\sqrt{\lambda_k(N, W)}} \int_{-W}^W U_k(N, W; v) dZ(f+v). \quad (3.1)$$

While unobservable, the $Z_k(f)$ are of considerable analytic interest in that they are the expansion coefficients which would be obtained if the entire sequence were passed through an ideal bandpass filter, from $f - W$ to $f + W$, before truncation to the finite sample size. The normalization implicit in this definition results in $E\{|Z_k(f)|^2\} = S$ when the spectrum S is white.

Now consider an estimate of these coefficients obtained by expanding the Fourier transform of the sample $y(f)$ over the interval $(f - W, f + W)$ on the same set of basis functions

$$y_k(f) = \frac{1}{\lambda_k(N, W)} \int_{-W}^W U_k(N, W; v) y(f+v) dv. \quad (3.2)$$

Using the basic integral equation (2.3) for $y(f)$, which expresses the projection operation from dZ onto y , together with the integral equation defining the discrete prolate spheroidal wave functions, $y_k(f)$ may be expressed in terms of dZ as

$$y_k(f) = \int_{-1/2}^{1/2} U_k(N, W; \xi) dZ(f+\xi). \quad (3.3)$$

Since this expression, used with the properties of the orthogonal increment process $dZ(f+\xi)$, is convenient for evaluating

various moments of the estimates, it should be noted carefully. In addition, the differences between the idealized coefficient (3.1) and the estimated coefficient (3.3) are important. Note particularly that the domain of integration is bandlimited in the idealized coefficient but includes the complete principal domain in the estimate. We call the $\{y_k(f)\}$ the eigencoeficients of the sample.

An alternative form of this estimate may be obtained using the definition of the discrete prolate spheroidal sequences, (2.8), given above and writing the Fourier transform $y(f)$ directly in terms of the data to obtain

$$y_k(f) = \sum_{n=0}^{N-1} x(n) \cdot \frac{v_n^{(k)}(N, W)}{\epsilon_k} e^{-i2\pi f[n-(N-1)/2]}. \quad (3.4)$$

Thus the k th eigencoefficient is the discrete Fourier transform of the data multiplied by the k th discrete prolate spheroidal sequence acting as a data window. We note that, in practice, it is simpler to regard the $v_n^{(k)}(N, W)$'s as real and the $U_k(N, W; f)$'s as complex so that the phase offset factors disappear and the calculations can be done with standard fast Fourier transform algorithms.

In Fig. 1(a)-(e) we show a set of $K = 5$ such data windows, which we will be using for the examples in this paper. In these windows $N = 100$ and $W = 0.04$. In this figure only the order zero window is "typical" of data windows in common use in that it is strictly positive. For the remainder, note that the k th data window has k zeros. It is of interest that windows of this type, i.e., having zeros in the domain of interest, are not new but were proposed by Lord Rayleigh in 1879, [270]. Note, however, that the higher order windows have a larger fraction of their energy near the ends of the interval.

To continue the development of the solution of the integral equation along the lines outlined in the preceding section, we next form the high-resolution estimate $\hat{z}_h(f; f_0)$, which is valid for $f_0 - W < f < f_0 + W$ as

$$\hat{z}_h(f; f_0) = \sum_{k=0}^{K-1} U_k(N, W; f-f_0) \frac{y_k(f_0)}{\lambda_k(N, W)}. \quad (3.5)$$

In this expression we are using, temporarily, unit coefficient weights for the first $K = [2NW]$ terms. Corresponding to this estimate of dZ , there is a high-resolution spectrum estimate

$$\hat{S}_h(f; f_0) = \frac{1}{N} \left| \sum_{k=0}^{K-1} U_k(N, W; f-f_0) \frac{y_k(f_0)}{\lambda_k(N, W)} \right|^2. \quad (3.6)$$

While more will be said about this estimate in later sections, we note that, since it is the absolute square of a complex Gaussian random variable, it is distributed as a multiple of χ^2_2 and, hence, is inconsistent. We thus take its average over the interior frequency domain $(f_0 - W, f_0 + W)$

$$\bar{S}(f_0) = \frac{1}{2W} \int_{f_0-W}^{f_0+W} \hat{S}_h(f; f_0) df$$

which, using the orthogonality properties of the prolate functions, becomes

$$\bar{S}(f_0) = \frac{1}{2NW} \sum_{k=0}^{K-1} \frac{1}{\lambda_k(N, W)} |y_k(f_0)|^2. \quad (3.7)$$

Since the eigencoefficients $y_k(f)$ are computed by transform-

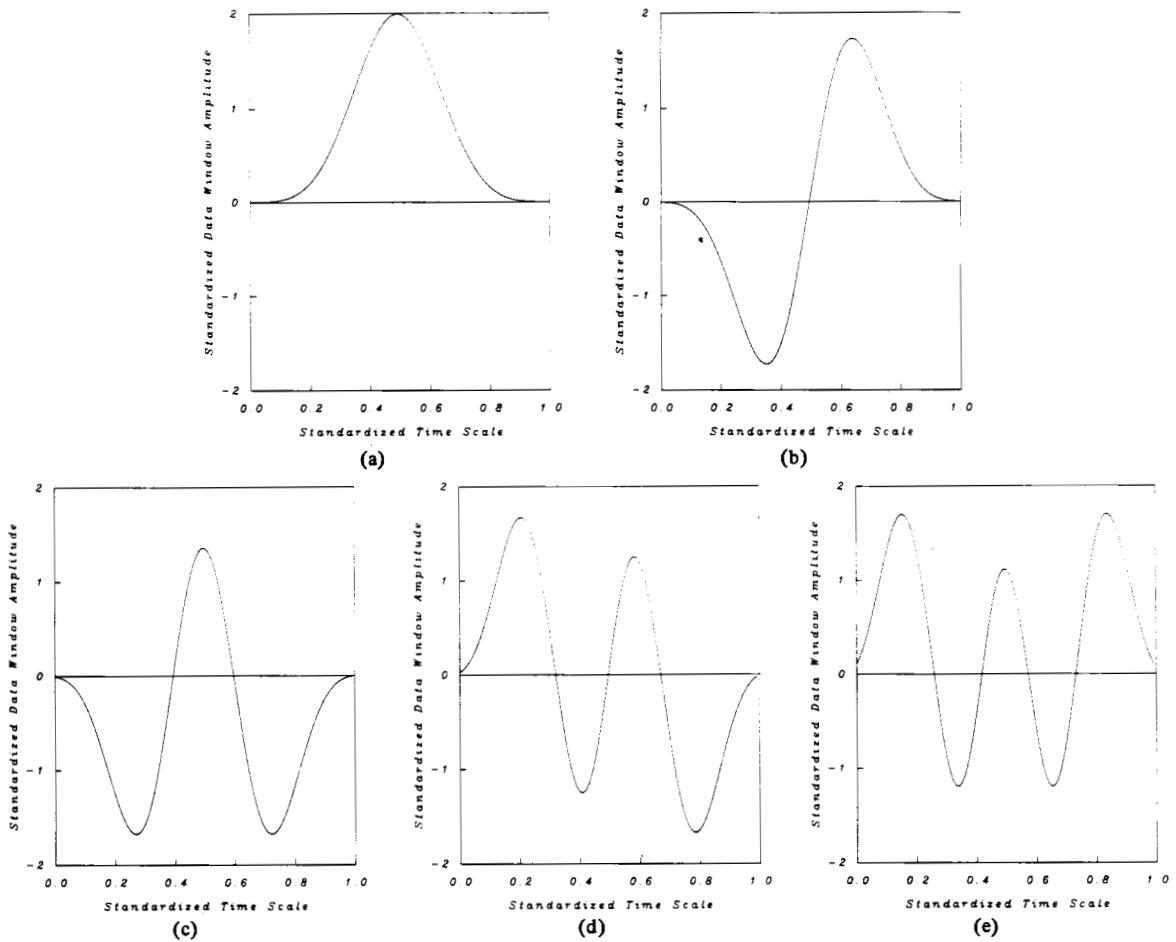


Fig. 1. Prolate spheroidal wave function data windows for a time-bandwidth product = 4 and $k = 0, 1, \dots, 4$. (a) Function 0, $k = 0$. (b) Function 1, $k = 1$. (c) Function 2, $k = 2$. (d) Function 3, $k = 3$. (e) Function 4, $k = 4$.

ing the data multiplied by the k th data window $v_n^{(k)}(N, W)$, their absolute squares

$$\hat{S}_k(f) = |y_k(f)|^2 \quad (3.8)$$

are *individually* direct spectrum estimates [338]. We refer to them as the *prolate eigenspectrum estimate of order k* , or simply the *k th eigenspectrum*. Since the discrete prolate spheroidal sequences are orthonormal on $[0, N - 1]$ the eigenspectra are normalized so that $\{\hat{S}_k(f)\} = S$ when the true spectrum S is white.

We note first that $\hat{S}_0(f)$ is the best known direct estimate of spectrum for a given W [103], [182], [323], [327]. However, when used by itself, as it always has been in the past, this estimate has had to be smoothed to produce a consistent result. This smoothing operation has the undesirable characteristic of increasing the effective bandwidth of the estimate to several times W with the concomitant increase in bias inherent in such smoothing operations. This effect has, of course, been known since Bartlett [28] and has been treated extensively since then in [41], [87], [137], [168], [214], [215], [243], [246]-[248], [264], among many others. In addition to the increased bias and loss of resolution intrinsic to such convolutional smoothing procedures, they are not to be recommended as, although the bias of $\hat{S}_0(f)$ is low, its variance efficiency is also low. While more will be said about this effect in Section VII, it is apparent from the proliferation of special windows, e.g.,

[21], [121], [153], [260], [274] to name but a few, that for a small sample the use of a single data window is generally unsatisfactory; see also [140], [152].

In the estimation procedure proposed here, however, there is an important element missing in all preceding windowed estimates—the additional eigencoefficients $y_1(f), \dots, y_K(f)$. The presence of these terms has four important effects on the estimate:

- 1) As will be shown in Section IV, the different eigencoefficients are almost uncorrelated for locally reasonably white spectra and, as they are individually complex Gaussian, their absolute squares are individually distributed as χ^2_1 , and, consequently, the estimate (3.7) has about $2K$ degrees of freedom. It is thus consistent as, for fixed W , this is equivalent to $4NW$ degrees of freedom.
- 2) This stability is achieved *without* the decrease in resolution associated with convolutional smoothers.
- 3) It will be shown in Section VI that the variance efficiency of the estimate is good, typically better than 80 percent. This is because the information missed by $\hat{S}_0(f)$ alone is largely recovered by the other eigencoefficients.
- 4) Because of the properties of the prolate spheroidal wave functions the unweighted frequency average used to convert the basic high-resolution estimate $\hat{S}_h(f; f_0)$ to the stabilized estimate $\bar{S}(f)$, the effective spectral window associated with $\bar{S}(f)$ approaches an almost “ideal” shape.

IV. SAMPLING PROPERTIES: I—MOMENTS

We next consider the lower order moments of these estimates. Because $\mathbb{E}\{dZ(f)\} = 0$, it is immediate that $\mathbb{E}\{y_k(f)\} = 0$. The second moment

$$\begin{aligned} \mathbb{E}\{y_j(f)y_k^*(f')\} &= \mathbb{E}\left\{\int_{-1/2}^{1/2} \int_{-1/2}^{1/2} U_j(N, W; \xi) \right. \\ &\quad \cdot U_k(N, W; \xi) dZ(f+\xi) dZ^*(f'+\xi)\left.\right\} \end{aligned}$$

becomes, by the properties of the orthogonal increment process,

$$\begin{aligned} \mathbb{E}\{y_j(f)y_k^*(f')\} &= \int_{-1/2}^{1/2} U_j(N, W; f+\xi) S(\xi) U_k(N, W; f'+\xi) d\xi. \quad (4.1) \end{aligned}$$

For the special case $j = k$ one has

$$\mathbb{E}\{\hat{S}_k(f)\} = |U_k(N, W; f)|^2 * S(f). \quad (4.2)$$

Thus the expected value of the k th eigenspectra $\hat{S}_k(f)$ is the convolution of the true spectrum $S(f)$ with the k th spectral window $|U_k(N, W; f)|^2$. Since the prolate spheroidal wave functions have all but a fraction $1 - \lambda_k$ of their total energy in the domain $(-W, W)$, all the lower order eigenspectra are good estimates from a bias viewpoint. Looking ahead to Section VI, Fig. 2(a)–(e) shows the first five spectral windows $|U_k(N, W; f)|^2$ for the case $N = 100$ and $NW = 4$ from which both the low sidelobes and the change in character near $f = W = 0.04$ are apparent. Note that, except for the zeroth case which resembles other windows in common use, the spectral windows have multiple “central lobes” interspersed between the k zeros of $U_k(N, W; f)$ in $(-W, W)$. Also, as one would expect from the behavior of the eigenvalues, the sidelobes of the higher order windows increase with order. The spectral window corresponding to $S(f_0)$

$$\frac{1}{K} \sum_{k=0}^{K-1} |U_k(N, W; f)|^2$$

is shown in Fig. 3. By virtue of (3.4), the individual eigenspectrum estimates $\hat{S}_k(f) = |y_k(f)|^2$ are direct estimates, hence, inconsistent and distributed as a multiple of χ^2 . Their average, however, is distributed approximately as a multiple of χ^2_{NW} and so is consistent.

It is convenient to split this integral into two parts, the first expressing the *local* bias

$$\overleftrightarrow{S}_k(f) = \int_{-W}^W |U_k(N, W; \xi)|^2 S(f+\xi) d\xi \quad (4.3)$$

and the second the *broad-band* bias

$$B_k(f) = \left| \int_{-W}^W U_k(N, W; \xi) dZ(f+\xi) \right|^2 \quad (4.4)$$

where the cut integral is defined as

$$\int_{-W}^W = \int_{-1/2}^{1/2} - \int_{-W}^{-1/2}.$$

The expected value of the broad-band bias is

$$\mathbb{E}\{B_k(f)\} = \int_{-W}^W U_k^2(N, W; \xi) S(f+\xi) d\xi. \quad (4.5)$$

Additionally, the broad-band bias may be bounded by use of the Cauchy inequality

$$B_k(f) \leq \int_{-W}^W |U_k(N, W; f)|^2 df \cdot \int_{-W}^W |dZ(f+\xi)|^2.$$

In this inequality the first integral expresses the energy in the delta window outside $(-W, W)$ and so equals $1 - \lambda_k(N, W)$. The second integral has the expected value

$$\int_{-W}^W S(f+\xi) d\xi \leq \int_{-1/2}^{1/2} S(f+\xi) d\xi = \sigma^2 \quad (4.6)$$

so that the broad-band bias is bounded by

$$\mathbb{E}\{B_k(f)\} \leq (1 - \lambda_k(N, W)) \sigma^2. \quad (4.7)$$

A. Quadratic and Local Bias

In many cases, particularly when working with very short series, it is unrealistic to assume that the spectrum in $f - W, f + W$ varies slowly, and, consequently, the local bias terms may be significant. To evaluate this effect we assume that the spectrum may be expanded in a Taylor series about f_0 so that the expected value of the k th eigenspectrum at this frequency may be written

$$\begin{aligned} \mathbb{E}\{\hat{S}_k(f_0)\} &= S(f_0) + \frac{S'(f_0)}{1!} \int_{-1/2}^{1/2} \xi |U_k(\xi)|^2 d\xi \\ &\quad + \frac{S''(f_0)}{2!} \int_{-1/2}^{1/2} \xi^2 |U_k(\xi)|^2 d\xi + \dots \end{aligned}$$

By symmetry, the coefficients of S' , S''' , etc., vanish, but those of S'' and higher even-order terms do not. Because it dominates the asymptotic bias, we consider only the quadratic term, that is the coefficient of $S''(f_0)$, which we write in normalized form as

$$G_k(N, W) = N^2 \int_{-1/2}^{1/2} |U_k(N, W; \xi)|^2 \xi^2 d\xi.$$

With this definition, the asymptotically dominant term of the local bias of the k th eigenspectra becomes

$$\frac{1}{N^2} \frac{S''(f_0)}{2!} G_k(N, W).$$

Because of the extreme concentration of the prolate functions, the value of this integral is largely a result of the integrand in $(-W, W)$. Second, because the effective spectral windows $|U_k(N, W; \xi)|^2$ widen with increasing k , it is clear that $G_k(N, W)$ should increase with k . Also, the fact that the energy in $(-W, W)$ is nearly 1 may be used to obtain an order-of-magnitude approximation to this integral so that, crudely, $G_k(N, W) \approx (NW)^2/3$. While this approximation is poor for the individual G_k 's, it is quite good for their average as the average spectral window, Fig. 3, approaches rectangular. Since $NW = c/\pi$, it is clear that, asymptotically, the local quadratic bias term must decrease as N^{-2} . For reference, doing the integrals numerically

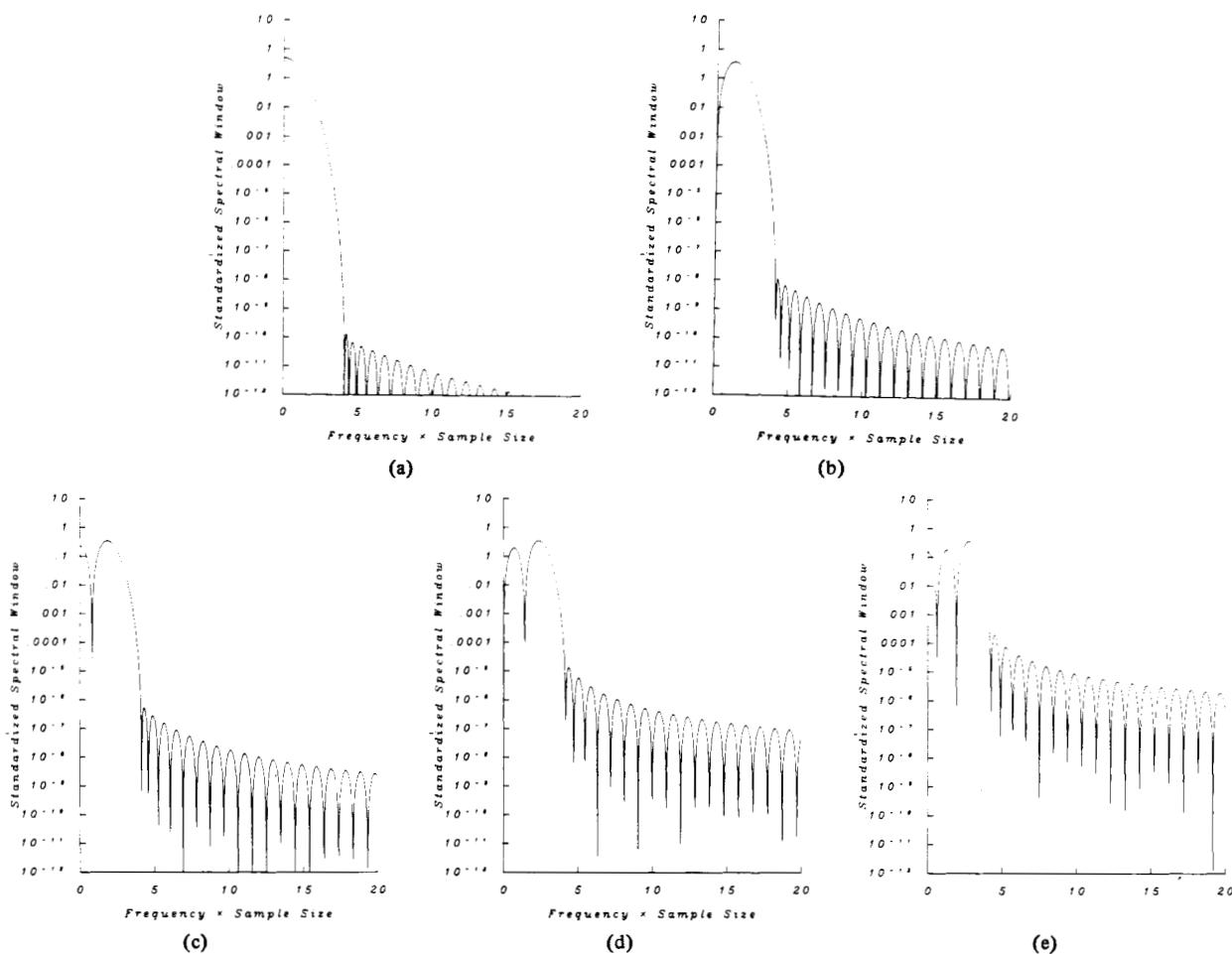


Fig. 2. Spectral windows corresponding to the prolate spheroidal wave function data windows shown in Fig. 1. The abscissa is in units of frequency times sample size. Note the sharp cutoff at $NW = 4$. (a) Function 0, $k = 0$. (b) Function 1, $k = 1$. (c) Function 2, $k = 2$. (d) Function 3, $k = 3$. (e) Function 4, $k = 4$.

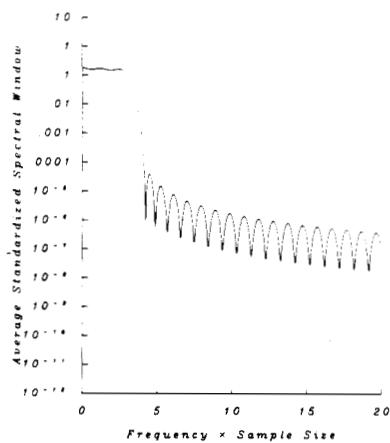


Fig. 3. The equivalent spectral window obtained by taking the arithmetic average of the first five eigenspectrum estimates. Note the nearly rectangular shape for frequencies below $Nf = 4$ and the low sidelobes.

with $c = 4\pi$ gives 0.644, 1.930, 3.225, 4.551, and 6.060 for G_0-G_4 , respectively.

Because G_k increases with k , it is reasonable to wonder if the quadratic bias associated with the eigencoefficient approach might not be greater than that obtained using only $\hat{S}_0(f)$ and

conventional convolutional smoothers. To make such a comparison we fix the degrees of freedom of the estimate and use the Papoulis [243] window (which was optimized with respect to quadratic bias) to smooth the zeroth eigenspectrum estimate. This results in quadratic bias coefficients of 2.64 and 18.5 for 4

and 10 degrees-of-freedom smoothers, respectively. Contrasted with this the corresponding quadratic bias coefficients of the averaged eigenspectrum estimates are only 1.287 and 3.282. Thus the quadratic bias of the eigencoefficient approach is much lower than that of the convolutional smoothers. Also, in anticipation of Sections VII and IX, we note that both the variance efficiency and estimation capacity of the convolutional smoother are poorer than those resulting from the eigenspectrum approach.

B. Distributions of Eigenspectrum Estimates

In addition to its expected value, the distribution of $\hat{S}_k(f)$ is of interest. First, if the process is Gaussian, $\hat{S}_k(f)$ will be distributed as chi-square with two degrees of freedom, χ_2^2 . Moreover, even when the original data are quite non-Gaussian, the additional filtering implicit in the estimation procedure will make the coefficients $y_k(f)$ tend to a complex-normal form, so their squares will be χ_2^2 . This effect is treated at length in [54], [220], [222], [282], [285], among others. Additional sampling properties, including bivariate distributions, of the eigenspectra are inferable from [45], [91], [135], [217], [229], [232], [297], [302], [313], [357], [362], and elsewhere. Further references of analytic interest include [10], [11], [120], [163], [164], [205], [206], [211], [261], [284], [295], [296], [298], [345].

A more interesting class of distributional problems arises out of the split into local and broad-band bias components where the broad-band bias term is dominant. In such cases, the local distribution appears significantly different from those where the local contribution dominates, and, while the distribution is proportional to χ_2^2 in an *ensemble* sense, for a given sample it is more appropriate to model it in terms of a *noncentral* chi-square distribution with a random noncentrality parameter. Identifying the noncentrality parameter with the broad-band bias component, it is clear that, in regions where this term dominates, the relative variability of the estimate will be much lower than normal. This effect is clearly visible in the examples of Section VI, particularly at frequencies around 0.15 cycles in Fig. 6.

A second special case in the general expression (4.1) above is that where $f = f'$ but $j \neq k$

$$\mathbb{E}\{y_j(f)y_k^*(f)\} = \int_{-1/2}^{1/2} U_j(N, W; f + \zeta) S(\zeta) U_k(N, W; f + \zeta) d\zeta. \quad (4.8)$$

Again, using the eigenvalue properties, the contribution from the exterior domain may be bounded to quantities exponentially small in NW and so, in many (but not all) cases of interest, may be ignored. Next, if the spectrum within $(f - W, f + W)$ is reasonably flat, the coefficients will be uncorrelated by the orthogonality properties of the prolate functions, but if the spectrum is highly peaked or changes rapidly in this region, the correlation may be significant. As an example, consider the case where $S(\zeta)$ consists of a unit step discontinuity at $f = \zeta$. Here, at frequency f , we obtain the covariance matrix given in Table I.

The final case we consider is that where $f' = f + \Delta f$ so that the prolate spheroidal wave functions are offset and no longer orthogonal. If we assume that the spectrum is white, we have

$$\Lambda_{jk}(\Delta f) = \int_{-1/2}^{1/2} U_j(N, W; f) U_k(N, W; f + \Delta f) df.$$

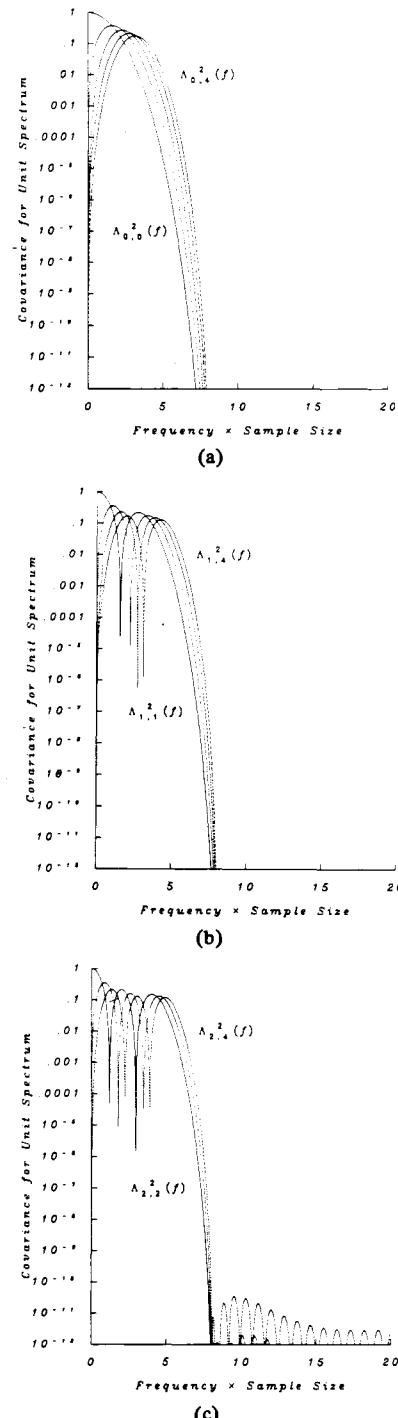


Fig. 4. Frequency offset (or lagged) cross-covariances, $\text{cov } \{\hat{S}_j(g), \hat{S}_k(f+g)\}$ for $S = 1$ and $NW = 4$. For $f > 2W$ the estimates are essentially uncorrelated.

TABLE I
EIGENCOEFFICIENT CORRELATIONS FOR UNIT STEP

$j \backslash k$	0	1	2	3	4
0	.5000				
1	+.3949i	.5000			
2	-.0037	-.2843i	.5000		
3	-.1549i	0	+.3398i	.5000	
4	+.0028	+.0976i	-.0017	-.2940i	.5000

Typical values of the square of this correlation coefficient are shown in Fig. 4. Note particularly the very small values attained for $\Delta f > 2W$.

V. ADAPTIVE WEIGHTING

While the bias properties of the lower order eigenspectra are generally excellent, because the eigenvalues decrease from one as k increases towards $2NW$, the bias characteristics of the successive estimates must degrade. Consequently, in regions where the spectrum is small, the higher order estimates will be less reliable than the lower order eigenspectra and must be downweighted. If one views the contribution to the k th spectrum estimate from the region $(f - W, f + W)$ as "signal," the contributions from the rest of the frequency as "noise," and the order, k , as "frequency," the weighting procedure is analogous to ordinary Wiener filtering, with the only difference being the basis functions. (The ordinary frequency, f , is simply a parameter defining the solution domain.)

We thus introduce a sequence of weight functions, $d_k(f)$, which, like the coefficients they modify, are functions of frequency and defined so that the mean-square error between $Z_k(f)$ and $y_k(f) \cdot d_k(f)$ is minimized. Using definition 3.1 for $Z_k(f)$ and (3.3) for $y_k(f)$ gives

$$\begin{aligned} Z_k(f) - y_k(f)d_k(f) &= \frac{1}{\sqrt{\lambda_k(N, W)}} \\ &\cdot \int_{-W}^W U_k(N, W; \xi) dZ(f + \xi) \\ &- d_k(f) \int_{-1/2}^{1/2} U_k(N, W; \xi) dZ(f + \xi) \end{aligned} \quad (5.1)$$

or, collecting regions of integration,

$$\begin{aligned} &= \left(\frac{1}{\sqrt{\lambda_k(N, W)}} - d_k(f) \right) \cdot \int_{-W}^W U_k(N, W; \xi) dZ(f + \xi) \\ &- d_k(f) \int_{-W}^W U_k(N, W; \xi) dZ(f + \xi). \end{aligned}$$

From this expression, it can be seen that the error consists of the sum of two terms, one defined on $(-W, W)$, the other on the remainder of the principal domain. Because both of these integrals are with respect to the random orthogonal measure dZ , they are independent and consequently the mean-square error is simply the sum of the squares of the two terms. Using the orthogonal increment properties of dZ again, and assuming the spectrum varies slowly over $(-W, W)$, the mean-square value of the first integral is well approximated by

$$\mathbb{E} \left\{ \left| \int_{-W}^W U_k(N, W; \xi) dZ(f + \xi) \right|^2 \right\} \approx \lambda_k(N, W)S(f).$$

The second integral is the broad-band bias, $B_k(f)$, of the k th eigenspectrum estimate defined in Section IV. This function depends on the gross features of the spectrum in the exterior domain. Since estimating the spectrum is the problem, this information must be approximated from the sample, and we use a two step procedure. First, by considering its average value over all frequencies, a fair initial estimate may be obtained

$$\int_{-1/2}^{1/2} \mathbb{E} \{B_k(f)\} df = \sigma^2 (1 - \lambda_k(N, W))$$

where σ^2 is the process variance

$$\sigma^2 = \int_{-1/2}^{1/2} S(f) df.$$

Combining these two integrals, and minimizing the mean-square error with respect to $d_k(f)$, gives the approximate optimum weight

$$d_k(f) \approx \frac{\sqrt{\lambda_k(N, W)}S(f)}{\lambda_k(N, W)S(f) + \mathbb{E} \{B_k(f)\}} \quad (5.2)$$

and the corresponding average of the spectral density function

$$\hat{S}(f) = \frac{\sum_{k=0}^{K-1} |d_k(f)|^2 \hat{S}_k(f)}{\sum_{k=0}^{K-1} |d_k(f)|^2}. \quad (5.3)$$

Since the (obviously unknown) values of the spectrum and broad-band bias appear in the weight expression, we replace them with estimates. The definition is then recursive and the resulting spectrum estimate is a solution of the equation

$$\sum_{k=0}^{K-1} \frac{\lambda_k(\hat{S}(f) - \hat{S}_k(f))}{[\lambda_k \hat{S}(f) + \hat{B}_k(f)]^2} = 0 \quad (5.4)$$

and thus must be in the interval bounded by the smallest and largest of the $\hat{S}_k(f)$'s. In practice the equation has been solved iteratively using the average of the two lowest order estimates as a starting value. Convergence has been rapid; for the examples shown in the following section convergence to the point where successive estimates differed by less than 5 percent required a maximum of 14 iterations and only 2.9 on average.

A useful by-product of this estimation procedure is an estimate of the stability of the estimates

$$v(f) = 2 \sum_{k=0}^{K-1} |d_k(f)|^2 \quad (5.5)$$

the approximate number of degrees of freedom for the estimate $\hat{S}(f)$ as a function of frequency. We note that $v(f)$ is a sensitive function of bias. If the average, over frequency, of $v(f)/2K$ is significantly less than 1, then either the value W is too small, or additional prewhitening should be used. Combined with the variance efficiency coefficient described in the next section \bar{v} provides a useful "stopping rule."

Once this initial estimate has been made, the estimated spectrum can be used to improve the estimate of broad-band bias $B_k(f)$. This may be efficiently computed by transforming the convolution (4.5) in the time domain. For this purpose we define an outer lag window

$$L_k^{(o)}(\tau) = \int e^{i2\pi\tau\xi} U_k(N, W; \xi)^2 d\xi$$

so that

$$\hat{B}_k(f) = \sum_{\tau} e^{-i2\pi f\tau} L_k^{(o)}(\tau) R^{(c)}(\tau) \quad (5.6)$$

where $R^{(c)}(\tau)$ is the autocovariance function corresponding to the spectrum estimate at the beginning of the current iteration. For this operation to be efficiently done using standard fast Fourier transform algorithms, two facts must be noted; first,

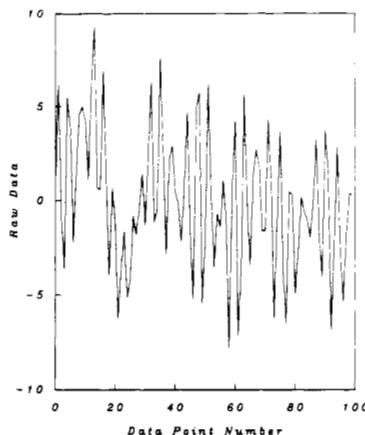


Fig. 5. The realization of the process described by (6.1) used in subsequent examples. $N = 100$.

the estimation procedure typically results in extrapolated autocovariance sequences, that is they are nonzero for lags $|\tau| > N$; second, in common with all numerical operations directly using the autocovariance function, the use of double precision arithmetic is advisable for many spectra of interest; see [86], [183].

An additional refinement is available by using the noncentral distribution suggested for $\hat{S}_k(f)$ in the previous section with $\hat{B}_k(f)$ considered as an estimate of the noncentrality parameter. Estimating $\hat{S}(f)$ by approximate maximum-likelihood results in a formula similar to (5.4). As will be shown in the next section, the effective weights obtained by this technique are somewhat higher than those obtained by the least squares method. However, since the difference in spectrum estimates is seldom more than 1 or 2 percent and the least squares method is much simpler, we omit the details.

VI. AN EXAMPLE

To illustrate some of the uncommon features of the eigen-spectrum estimates, we consider a process whose spectrum is a composite of features of spectra typically found in communications systems. In this process the data consist of a number of subcomponents

$$\begin{aligned} x(t) = & p_1(t) + p_2(t) \cos(\omega_c t) + p_3(t) \sin(\omega_c t) \\ & + 2.4 \sin(\omega_1 t) - 2.6 \cos(\omega_2 t) + n(t) \quad (6.1) \end{aligned}$$

where the processes $p_1(t)$, $p_2(t)$, and $p_3(t)$ are independent with Bessel autocovariance functions of the type described in [20], [77], [118]. (The autocorrelation function is given by $J_0(\tau/\tau_0)$ so that the spectrum of the p_i processes is band-limited and proportional to $\sqrt{1 - (f/B)^2}$ for $|f|$ less than the bandwidth, $B = 0.078125$.) Data for these three processes were created using Karhunen-Loeve expansions, with the eigenvalues and eigenfunctions being computed by a procedure similar to that described in the Appendix. The expansion coefficients, with variances given by the eigenvalues, were generated by a Gaussian random-number generator. The two "line components" are at frequencies $f_1 = 0.2556$ and $f_2 = 0.3242$, neither having an integral number of periods in the sample of $N = 100$ data points. The observational noise is represented by $n(t)$ which is a white noise process of variance 10^{-6} . On the plots of the various estimates, the continuous part of the theoretical spectrum of the composite process is shown as the dashed line and identified as " $S(f)$." The expected value of the amplitude

of the two line components at the working frequency resolution $\frac{1}{512}$ is shown by asterisks and marked "lines."

Fig. 5 shows a typical realization of length $N = 100$ points from such a process. Observe that, although the process is highly structured and predictable, it appears noise-like. The data and corresponding spectral windows used in this example were shown earlier as Figs. 1 and 2. The eigenestimates $\hat{S}_k(f)$ for the data set shown in Fig. 5 above are shown in Fig. 6(a)-(e). In these estimates, a time-bandwidth product, $NW = 4$, has been used so that the estimator has enough dynamic range. The spacing of the two line components, however, $0.3242 - 0.2556 = 0.0686$ is less than $2W$ so that interactions occur. Observe that the estimates $\hat{S}_0(f)$ and $\hat{S}_1(f)$ are similar except for details and that both follow the true spectrum except for the areas immediately adjacent to the band edges. In these regions, they perform as expected. In the regions around the line components (about which more will be said in Section XIII), the details reflect primarily the shape of the respective spectral windows. In both estimates the basic shape of the theoretical spectrum is reproduced and the effect of the finite resolution, $W = 0.040$, is clear. With the next three estimates, $\hat{S}_2(f)$ to $\hat{S}_4(f)$, the effects of the eigenvalues decreasing away from 1 becomes successively more pronounced, particularly in the band between $0.12 < f < 0.19$. Observe also the lower and more systematic variations of the estimates in this band due to the noncentral distributional characteristics induced by the bias compared to the larger variations of the estimates at lower frequencies. Where the spectrum is larger, however, these estimates are still clearly providing useful data as, while the details differ, they reproduce the correct general shape.

The broad-band bias, computed using (5.6), is also shown in Fig. 6(a)-(e) as the lower curve identified as $\hat{B}_k(f)$. This bias component is, as expected, very low for the two lowest order estimates while, in the gap mentioned above, the higher order estimates consist primarily of broad-band bias. Note also that there are considerable variations between the estimated bias and that observed; this is a result of the rather arbitrary division made at W between local and broad-band bias and the fact that here $\hat{B}_u(f)$ is computed from an estimate which has significant local bias.

Using these estimates of the broad-band bias in (5.4) gives the weight functions shown by the solid lines in Fig. 7(a)-(e). In these figures, the least squares weights are shown as the dashed lines while the approximate maximum-likelihood weights are shown as solid lines. Observe that the estimate $\hat{S}_0(f)$ is given

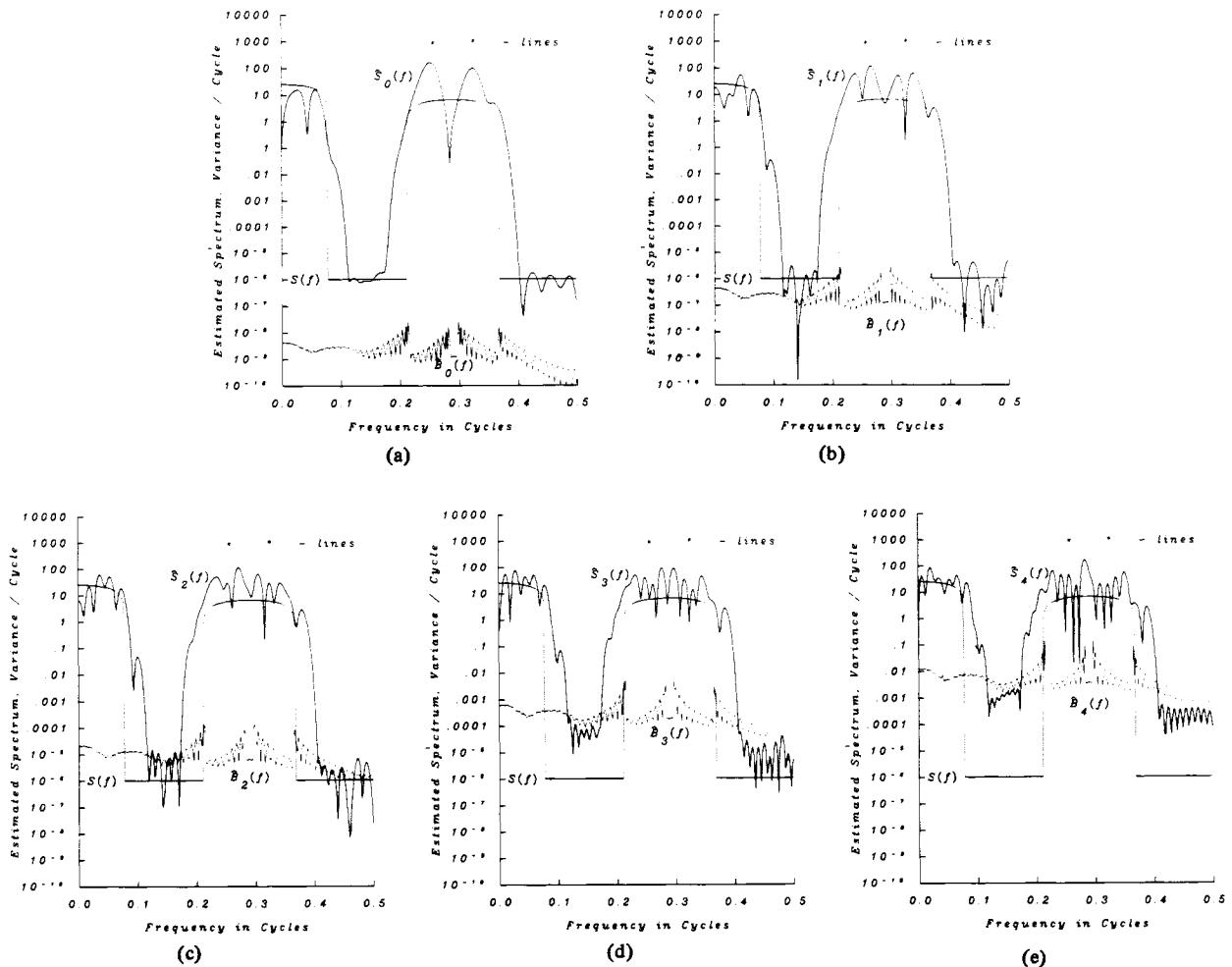


Fig. 6. Estimates of the eigenspectrum $\hat{S}_k(f)$, and the broad-band bias $\hat{B}_k(f)$, for $k = 0, 1, \dots, 4$ and the data shown in Fig. 5. The data windows used were shown in Fig. 1. The true spectrum is shown as the dashed line and the two line components by *. In the case of the line components, the plotted amplitude corresponds to a frequency resolution of 1/512. (a) Eigenspectrum 0, $k = 0$. (b) Eigenspectrum 1, $k = 1$. (c) Eigenspectrum 2, $k = 2$. (d) Eigenspectrum 3, $k = 3$. (e) Eigenspectrum 4, $k = 4$.

nearly full weight and $\hat{S}_1(f)$ almost as much. However, in the regions where the broad-band bias is large, the weights on the higher order estimates have dropped significantly so that in these regions $\hat{S}_4(f)$ is weighted by only $\approx 10^{-7}$. In regions where the spectrum is high, on the other hand, $\hat{S}_4(f)$ is effectively receiving unit weight. The resulting spectrum estimate, computed using (5.4) with $\hat{B}_k(f)$ given by (5.6), is shown in Fig. 8. The overall estimate is clearly quite good except near the line components (again, there will be more about this in Section XIII) where only the two roughly symmetric "bumps" are visible. Also, as expected, the stabilized estimate $\bar{S}(f)$ does not follow the discontinuities at the band edges. (Recall that $\bar{S}(f)$ was obtained in Section III by integrating the basic high-resolution estimate over $(f - W, f + W)$ so that a low-pass characteristic is to be expected.) Because the local bias causes the gaps to be significantly narrowed, an estimate of the innovations variance of the process would be too high. This effect has some serious implications when one plans adaptive pre-whitening or model fitting with a process similar to this one. This example will be continued in Section X when we again consider the high-resolution estimates. In addition, the estimate varies across the band. Where the spec-

trum is large, all the $K = 5$ estimates contribute so that the stability is characterized by 10 degrees of freedom whereas in the lower regions, where only the first two estimates are reasonably unbiased, the stability is only about 3.5 degrees of freedom. The stability, in equivalent degrees of freedom, is plotted in Fig. 9. Its average value, across frequency, is 6.38.

An important lesson to be learned from this example is that handling mixed spectra, particularly where the range of the spectrum is large, is difficult. In the individual eigenspectrum estimates neither line is particularly obvious, compared to other "peaks," even though the line energy to local noise power is about 13 dB. This phenomenon is more common in such cases than when the noise is white, reflecting the lesser mixing and longer persistence of effects in the highly structured processes. In particular, when the duration of the data is less than the length of a reasonable moving-average representation, as happens in this example, these effects can be particularly severe. It is equally common to observe "peaks" well above the apparent background, which reflect nothing more than sampling variation and the simple fact that in most spectrum estimation problems one is looking at large numbers of estimates. Consequently, simulations based on nearly white or white spectra

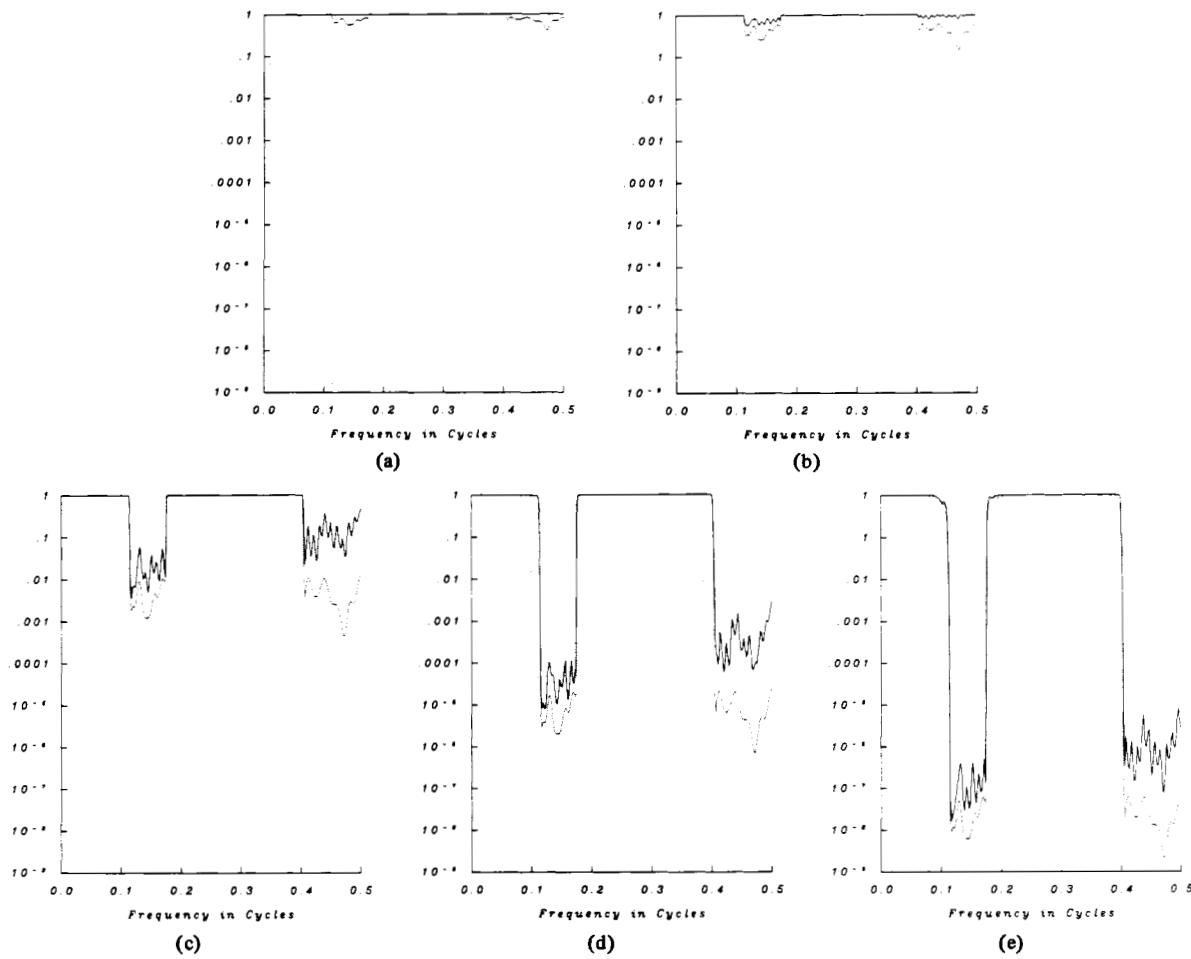


Fig. 7. The least squares (dashed line) and approximate maximum-likelihood (solid line) solution weights, $|d_k(f)|^2$, for $k = 0, \dots, 4$ and the eigenspectrum estimates and bias estimates shown in Fig. 6. For $k = 0$ the approximate maximum-likelihood weight is nearly 1. Note the change in behavior as k increases particularly in regions where the spectrum is low. Also, when the spectrum is "large," all weights are large. (a) $k = 0$. (b) $k = 1$. (c) $k = 2$. (d) $k = 3$. (e) $k = 4$.

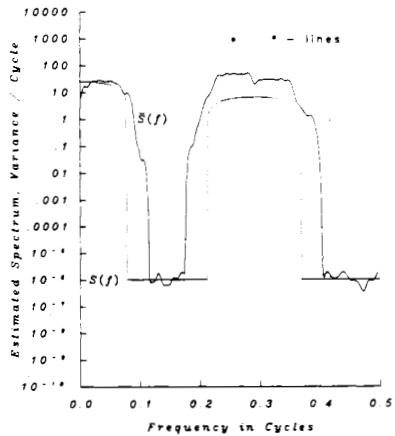


Fig. 8. The weighted average spectrum estimate $\bar{S}(f)$. This estimate is formed by combining the eigenspectrum estimates shown in Fig. 6 using the weights plotted in Fig. 7. As before, the true spectrum is shown by the dashed line. For the line components compare this figure with Figs. 23 and 24.

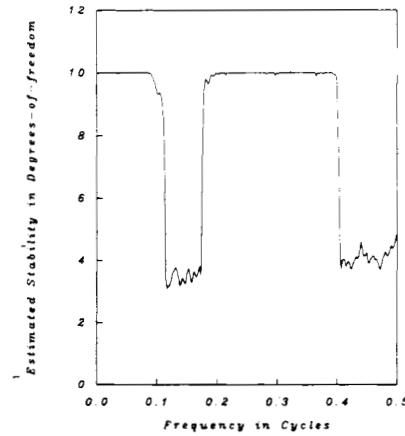


Fig. 9. The estimated stability of the spectrum estimate shown in Fig. 8 plotted in degrees-of-freedom. By using more than the $K = 5$ eigen-spectra used in these examples, it is possible to obtain greater stability in regions where the spectrum is large.

can provide a seriously misleading basis for inferences about the highly colored spectra encountered in nature, particularly when only short records are available.

VII. SAMPLING PROPERTIES: II—VARIANCE AND VARIANCE EFFICIENCY

The characterization of spectral estimates is a difficult subject on which there is little agreement and there are numerous papers; [9], [42], [101], [102], [107], [153], [290] among others. To study the sampling properties of these estimates, we assume that the data are Gaussian and, using the formula for the fourth moments of a Gaussian process, obtain

$$\begin{aligned} \text{cov} \{ \hat{S}_j(f+g), \hat{S}_k(f-g) \} &= \left| \int_{-1/2}^{1/2} S(\xi) U_j(f+g-\xi) U_k(f-g-\xi) d\xi \right|^2 \\ &\quad + \left| \int_{-1/2}^{1/2} S(\xi) U_j(f+g-\xi) U_k(f-g+\xi) d\xi \right|^2. \quad (7.1) \end{aligned}$$

In this expression, the last term is large only near the origin and the Nyquist frequency. The first term, however, is a generally dominant convolution form having several features of interest. First, if the frequency separation $2g$ is larger than $2W$, so that the "central lobes" of the two estimates do not overlap, the range of integration can be split, as above, and the covariance bounded in terms of the eigenvalues. Such correlation bounds are typically small. Second, if the frequency separation is less than $2W$ so that the central lobes overlap, the covariance depends primarily on the spectrum in the domain $(f-g-W, f+g+W)$, and the contribution from spectral components outside this neighborhood can be bounded by the eigenvalue properties in a manner similar to that given above. If the spectrum around f is constant or linear, the correlation at frequency separation $\Delta = 2g$ is

$$\begin{aligned} \Lambda_{jk}^2(\Delta) &= |U_j(\Delta) * U_k(\Delta)|^2 \\ &= \left| \sum_{n=0}^{N-1} v_n^{(j)} v_n^{(k)} e^{in\Delta} \right|^2. \quad (7.2) \end{aligned}$$

In particular, when Δ is zero, the correlation between eigen-spectra is

$$\text{corr} \{ \hat{S}_j(f), \hat{S}_k(f) \} = \left| \int_{-1/2}^{1/2} U_j(\xi) U_k(\xi) d\xi \right|^2$$

which is zero for $j \neq k$. Even for Δ not zero the frequency offset cross correlations are quite small, for example with $c = 4\pi$ one has $\Lambda_{01}^2(1.59/T) = 0.366$ and $\Lambda_{02}^2(2.2/T) = 0.264$ as maximum values. Typical functions were shown in Fig. 4 for this case. Note the extremely small correlations expected for frequency separations more than $2W$.

Intimately related to the variance of an estimator is the notion of efficiency. Clearly the efficiency of a spectrum estimate involves not only the variance of the estimate at a given frequency but also the covariability of estimates at different frequencies, as otherwise the apparent efficiency could be increased by additional smoothing. Carrying this idea to its limit and using the integrated spectrum as an estimate of variance for uncorrelated data, one obtains a simple measure of the overall efficiency of the estimate [172]. For a single eigen-

TABLE II
EIGENVALUES AND VARIANCE EFFICIENCIES

c	K	$1 - \lambda_{K-1}$	Ξ_K
π	1	.0189	.745
	2	.2504	.982
	3	.7564	.934
2π	1	5.725e-05	.509
	2	2.438e-03	.734
	3	4.061e-02	.900
	4	.2783	.992
	5	.7253	.955
3π	1	1.348e-07	.412
	2	9.245e-06	.595
	3	3.850e-04	.731
	4	5.086e-03	.844
	5	5.386e-02	.939
4π	1	2.946e-10	.356
	2	2.768e-08	.515
	3	1.210e-06	.632
	4	4.245e-05	.730
	5	5.899e-04	.814
	6	7.496e-03	.890

spectrum estimate, the variance efficiency is

$$\xi_k = \frac{1}{N \sum_{n=0}^{N-1} [v_n^{(k)}(N, W)]^4}$$

while for the average of the first K estimates one obtains

$$\Xi_K = \frac{1}{N \sum_{n=0}^{N-1} \left[\frac{1}{K} \sum_{k=0}^{K-1} [v_n^{(k)}(N, W)]^2 \right]^2}.$$

Using the prolate spheroidal approximations to the discrete prolate spheroidal wave functions with, as before, $c = N\pi W$ results in the efficiencies shown in Table II. It is apparent from this table that the efficiencies of the eigenspectrum estimates can be high; for example, the estimates with $c = 3\pi$, $K = 4$, and $c = 4\pi$, $K = 5$ are both over 80 percent efficient, easily computed, and yet provide excellent bias protection. When the adaptive weighting described in the previous section is used, the bias of these estimates is much lower owing to the extremely low leakage of their initial eigenspectra. From the table it may also be seen that if more than the first $[2c/\pi]$ estimates are used, not only does bias protection drop rapidly, but the efficiency also drops.

We must emphasize that the idea of variance efficiency should not be taken as a strong criterion in comparing spectrum estimates. First, it is strictly valid only for white noise processes. Second, it ignores bias and its consequences so that, for example, if one judged estimates solely from a variance efficiency viewpoint, one would be left with the periodogram and its variants as the only admissible form.

The most important thing to note is that, if conventional procedures are used, that is the estimate is made by

- 1) multiplying the data by a good data window, $v_n^{(0)}(N, W)$, or a Kaiser [182] approximation,
- 2) transforming the windowed data and squaring,
- 3) using a matched convolutional smoother to stabilize the result,

the efficiency cannot exceed $\Xi_1(c)$ for any smoothing technique. Thus for $c = 3\pi$ and 4π the eigenspectra approach is more than twice as efficient as conventional windowed estimates.

With regards to the efficiency, recall that several "optimum" spectrum estimates based on general quadratic forms, $\hat{S}(f) = \mathbf{X}^\dagger \mathbf{A}(f) \mathbf{X}$, have been proposed, e.g., [137], [209], [214], where knowledge of the spectrum was assumed to compute the optimum weight function, $\mathbf{A}(f)$. Since the unweighted eigenspectrum estimates are also quadratic forms, it is reasonable to question how they can be more efficient than these earlier estimates. There are several reasons. First, $\mathbf{A}(f)$ was usually restricted to depend *only* on the sample autocovariances and not allowed to be general. Second, optimality was generally established only asymptotically. Third, it was assumed that the periodogram was unbiased. Thus since examples of such estimates for $\mathbf{A}(f)$ not restricted to the class of periodogram estimates appear to be unknown, their optimality cannot be taken too seriously. Further, the *weighted* estimate (5.4) is not a simple quadratic form but a rational combination of them. This problem is pursued further in the following section where we show a connection between the eigenspectrum estimates and the periodogram, and in Section IX where we consider the idea of *logarithmic efficiency*. Also, in anticipation of Section XII, we note that while the periodogram corresponds to a maximum of the likelihood function, it is *not* the global maximum.

Combining these results with the stability estimate $v(f)$, (5.5) gives an overall measure of the efficiency

$$\text{eff} \approx \overline{v(f)} \Xi_K(c)N$$

which may be used to compare the effectiveness of prewhitening (which reduces N), varying W and K , etc.

VIII. RELATIONS BETWEEN EIGENESTIMATES AND THE PERIODOGGRAM

There is an interesting relationship between eigenspectrum estimates and the periodogram showing the bias problems of periodogram-based estimates simply in terms of the weighting used in Section V above. We begin by expanding the Dirichlet kernel using a bilinear formula (see [334 ch. 3])

$$\frac{\sin N\pi(f - f')}{\sin \pi(f - f')} = \sum_{k=0}^{N-1} U_k(N, W; f) U_k(N, W; f')$$

or, replacing f and f' by $f - f_0$ and $f' - f_0$,

$$\frac{\sin N\pi(f - f')}{\sin \pi(f - f')} = \sum_{k=0}^{N-1} U_k(N, W; f - f_0) U_k(N, W; f' - f_0).$$

Multiplying both sides of this equation by $dZ(f')$ and integrating gives

$$y(f) = \sum_{k=0}^{N-1} U_k(N, W; f - f_0) y_k(f) \quad (8.1)$$

where the basic integral equation (2.3) has been used to obtain the left side and (3.3) the right. Squaring both sides and multiplying by $1/N$ gives an expression for the periodogram valid for $|f - f_0| < W$

$$I(f) = \frac{1}{N} \left| \sum_{k=0}^{N-1} U_k(N, W; f - f_0) y_k(f) \right|^2. \quad (8.2)$$

Observe that *all* the eigencoefficients appear in this expression with unit weight. If one now uses a uniform smoother of width

$2W$ the result is

$$\bar{I}_W(f_0) = \frac{1}{2NW} \sum_{k=0}^{N-1} \lambda_k(N, W) \hat{S}_k(f_0). \quad (8.3)$$

This shows the smoothed periodogram to be equivalent to a weighted average of the eigenspectra with weights independent of frequency. In addition the average is over *all* N eigenspectra, not just the $2NW$ whose large eigenvalues imply that the information contained in these coefficients is of local origin.

Comparing the effective weights applied to the different eigenspectra, $\lambda_k(N, W)$, to those from the simple least squares approach

$$\frac{\lambda_k(N, W) S^2(f)}{[\lambda_k(N, W) S(f) + \sigma^2(1 - \lambda_k(N, W))]^2}$$

shows that the two are equal *only* for white spectra $S(f) \equiv \sigma^2$, and that otherwise the higher order eigenspectra contribute significant bias to the periodogram.

IX. SAMPLING PROPERTIES: III—LOGARITHMIC EFFICIENCY

A problem in spectrum estimation is that of characterizing a "good" estimate. The idea of variance efficiency discussed in the previous section is useful in a restricted sense but fails in that it does not penalize bias. In addition, the variances computed are polynomial functions of the spectra as opposed to the logarithmic or ratio functions one would expect from the Itakura-Saito [166], [167] or similar measures; see also [170], [179], [189], [316].

As an alternative, consider the spectrum estimation process as part of a communications channel, or more correctly, as multiple channels closely spaced in frequency. As such, it would be natural to ask about the capacity of these channels, about interference between them, and similar questions. Specifically we consider the average information gained by computing a second estimate beyond that available from the initial estimate.

This second estimate may be either one at a different frequency, or as with the eigencoefficient expansion, an additional estimate at the same nominal frequency made using a different filter. To do this we still impose the constraint that any filters used must be restricted to the N data points available, but consider these N points as a "sliding block" [131]. Thus instead of having the eigencoefficients $\{y_k(f)\}$, we consider them to be an ensemble of stationary sequences $\{y_k(f|\tau)\}$, where τ indexes the origin or block position. The information we measure will then be between the sequences $\{y_k(f_0|\tau)\}$ and $\{y_k(f_1|\tau)\}$ in the case of different center frequencies, f_0 and f_1 , or between the stationary sequences $\{y_i(f|\tau)\}$ and $\{y_k(f|\tau)\}$ for two filters being used at the same center frequency. Again, note that the assumed stationarity is with respect to the block position parameter τ . The basic idea we use is expressed by the formula for the average mutual information between random events "0" and "1" [32], [197]

$$M(0; 1) = I(1) - I(1|0)$$

where $I(1)$ is the differential entropy associated with the probability distribution of event "1" and $I(1|0)$ the differential entropy of event "1" conditioned on event "0." We rewrite this formula as

$$I(1|0) = I(1) - M(1; 0)$$

interpret "1" as the second estimator, "0" as the initial estimator, and the result as the entropy rate of the second estimator sequence above that inferable from the first sequence. The argument can also be done in the opposite order and, since the order in which the estimates are made should be irrelevant, define the *mutual estimation capacity* of the pair of filters as

$$EC_{01} = \frac{1}{2} \{J(0) + J(1)\} - M(1; 0).$$

To evaluate this function we consider the two sequences to be Gaussian and separated by filters H_0 and H_1 . It is tempting

$$I(1|0) = \frac{1}{2} \int_{-1/2}^{1/2} \ln \left[\frac{S(f) \{ |H_1(f)|^2 N_0(f) + |H_0(f)|^2 N_1(f) \} + N_1(f) N_0(f)}{N_1(f) \{ S(f) |H_0(f)|^2 + N_1(f) \}} \right] df.$$

to try to use the simple differential entropy rate powers

$$h_m = \int_{-1/2}^{1/2} \ln \{ |H_m(f)|^2 S(f) \} df$$

for this purpose, but the differential entropy has several disadvantages: first, it is the sum of the process and filter differential entropies independent of the estimation frequency; second, differential entropies are not coordinate invariant; and third, the estimation process is not noise free in that estimates at a given frequency are contaminated, not only with leakage from all other frequencies (which provides the mutual information), but also with measurement and roundoff noise. Thus we define the noise spectrum at a frequency f associated with the estimation sequence produced by filter H_m as $N_m(f)$ resulting in the entropy rate of ([256, eq. (10.2.4)])

$$I_m = \frac{1}{2} \int_{-1/2}^{1/2} \ln \left\{ 1 + \frac{|H_m(f)|^2 S(f)}{N_m(f)} \right\} df.$$

This formula expresses the mutual information between the noiseless filtered sequence and the observed noisy sequence; as such, it is a form of Shannon's [294] equation for channel capacity.⁴

Similarly, it is possible to characterize the filter leakage between estimates at two different frequencies in terms of the mutual information

$$M(0; 1) = -\frac{1}{2} \int_{-1/2}^{1/2} \ln (1 - |C(f)|^2) df$$

where $C(f)$ is the coherence between the two processes. If the filters were ideal bandpass and nonoverlapping, their outputs would be incoherent, $M(0; 1)$ zero, and consequently there would be no filter leakage penalty. Alternatively, if the processes were highly coherent, as they would be with severe filter leakage, the mutual information, $M(0; 1)$ would be high. To apply this concept here one must remember that if one assumes nonphysical noiseless computations, then, since both filtered sequences are subordinate to the input series [193],

⁴Specializing to give an idealized bandpass filter of width W and constant signal and noise spectra within the band the formula becomes

$$C = W \ln \left(1 + \frac{S}{N} \right).$$

the outputs of the two filters will be completely coherent and the result meaningless. Thus we assume that the two output processes have spectra

$$S_m(f) = |H_m(f)|^2 \cdot S(f) + N_m(f)$$

and cross spectra

$$S_{mk}(f) = H_m(f) H_k^*(f) \cdot S(f)$$

so that the entropy rate from filter "1" above that from filter "0" is

If one now assumes that the signal-to-processing-noise ratio is large and that the noise spectra $N_0(f)$ and $N_1(f)$ are identical, then one obtains the particularly simple form for $I(1|0)$

$$I(1|0) = \frac{1}{2} \int_{-1/2}^{1/2} \ln \left[1 + \frac{|H_1(f)|^2}{|H_0(f)|^2} \right] df$$

the additional information derived from filter "1" above that already available from filter "0." Note that this equation is again similar to Shannon's channel capacity formula. Here the "signal spectrum" is given by $|H_1(f)|^2 S(f)$, the "noise spectrum" by $|H_0(f)|^2 S(f)$, the result independent of the spectrum and depending only on the estimation filters. Thus the formula provides a quantitative measure of the intuitive notion that filters with high sidelobe "leakage" give poorer estimates than those with low leakage. Further, because we assumed that the processing noises were independent, we note that making a second estimate with otherwise identical filtering gives an estimation capacity of $\frac{1}{2} \ln 2$, a 1-bit improvement in variance!

Because it is immaterial which estimate is used for conditioning, we use the average of $I(0|1)$ and $I(1|0)$

$$EC_{01} = \frac{1}{4} \int_{-1/2}^{1/2} \ln \frac{1}{4} \left[\frac{|H_0(f)|^2}{|H_1(f)|^2} + 2 + \frac{|H_1(f)|^2}{|H_0(f)|^2} \right] df$$

which, by analogy with channel capacity, we call the *mutual estimation capacity* of the pair of filters. Also, in this definition, the zero offset for identical noise spectra has been subtracted so that the mutual estimation capacity reduces to zero for identical filters. When the two filters are the same except for different center frequencies, $H_1(f) = H_0(f + \Delta f)$, we write the expression as $EC_h(\Delta f)$ and refer to it as the estimation capacity function of the filter h . When a single number is required to characterize a spectrum estimation procedure, we use the *average estimation capacity*.

$$\overline{EC}_h = \int_{-1/2}^{1/2} EC_h(f) df.$$

As defined, the estimation capacity is a proper distance or metric between the two filters; it is zero when they coincide, symmetric, and obeys the triangle inequality. The metric is clearly similar to Gray and Markel's [130] "cosh" distortion measure

$$d_{\text{cosh}} = \int_{-1/2}^{1/2} \left(\cosh \left\{ \ln \left[\frac{|H_1(f)|^2}{|H_0(f)|^2} \right] \right\} - 1 \right) df$$

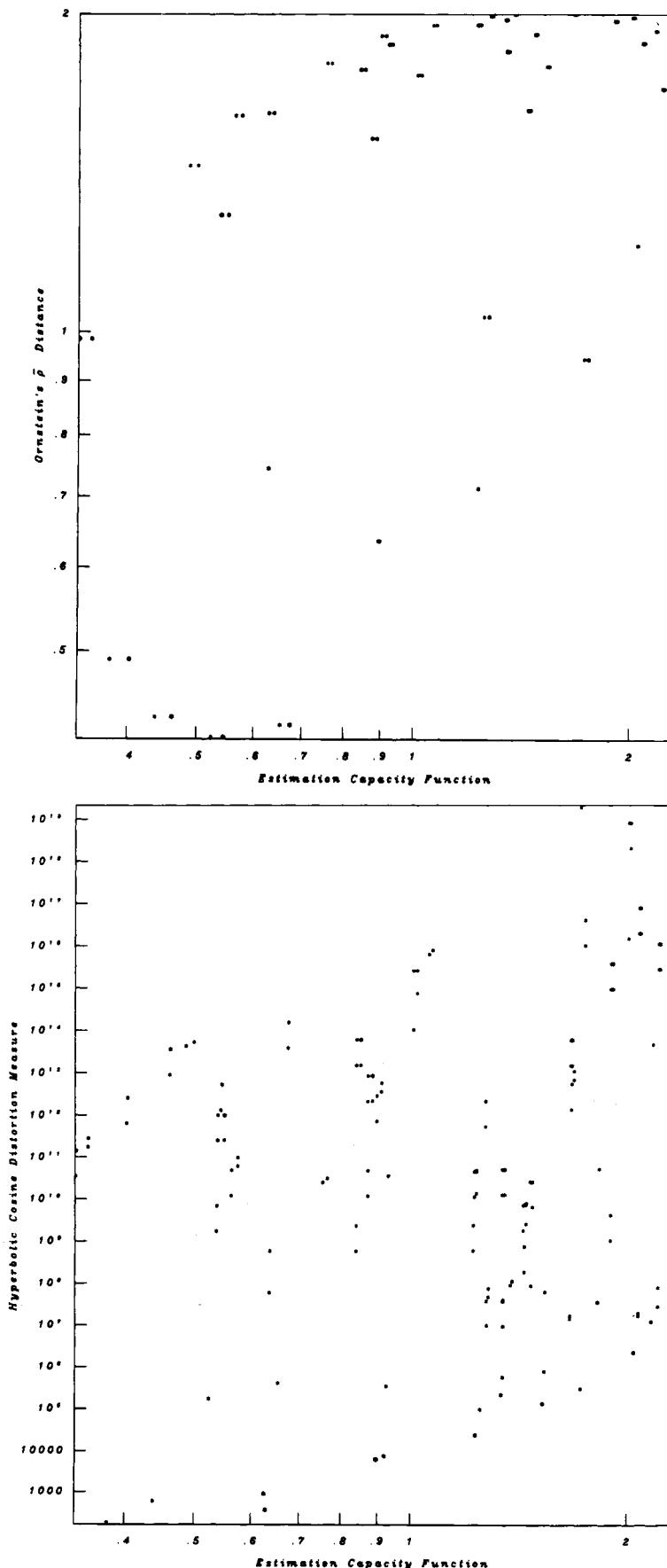


Fig. 10. Scatter plots comparing Ornstein's p distance, Gray and Markel's d_{\cosh} , and the estimation capacity function defined here. In this case, the process spectra being compared are those of the frequency translated basis functions of Section XI.

but is closer to the $\bar{\rho}$ metric described in [132]–[134]

$$\bar{\rho} = \int_{-1/2}^{1/2} \{ |H_1(f)| - |H_0(f)| \}^2 df.$$

Although these measures are superficially similar, they are not particularly well correlated with each other (see Fig. 10). This figure shows scatter plots of the different distance measures applied to all different pairs in the set of basis functions consisting of frequency-translated subsets of prolate spheroidal wave functions to be described in Section XII. Because, in this application, we are using the distance measures on functions supposed to be distinct, large numbers are obtained.

As an example of the differences between estimators Fig. 11 shows N times the estimation capacity function of the periodogram, and the discrete prolate spheroidal wave functions $U_0(100, 0.02; \cdot)$ and $U_0(100, 0.04; \cdot)$. The capacity of the prolate estimators is clearly much higher than that of the periodogram. A further question is raised by the “breaks” in the curves for the prolate estimators near $\Delta f \approx 2W$. Since their spectral windows are wider than the central lobe of the periodogram window, is the increased capacity of the prolate windows just an artifact of their greater bandwidth compensated for by fewer effectively independent estimates? Since the number of estimates available at a frequency spacing of Δf is roughly $1/\Delta f$, Fig. 12 shows the estimation capacity functions times $1/\Delta f$. Note that the answer is unchanged except that it is now obvious that the greatest effect is achieved with the prolate estimates spaced W apart.

Except for some special cases, the interpretation of the numerical values of the capacity function is unclear except in a relative sense. First, the relative estimation capacity $EC(\Delta f)/\Delta f$ is about equal to the first sidelobe level (in nepers) of the spectral window. Second, in harmonic analysis problems and similar parametric estimation problems, the capacity function appears to play a role similar to channel capacity in that it defines the average number of significant bits, using base 2 logarithms, attainable by the estimate. Such numbers must be treated carefully. Third, in the purely nondeterministic spectrum estimation problem, the estimation capacities imply that the logarithm of a prolate estimate with $c = 4\pi$ is valid to perhaps 1.5 bits, and that of a periodogram only to about 0.2 or 0.3 bit on average. This is in general agreement with experience and the low significances attached to simple spectrum estimates of purely random phenomena.

Also note that if convolutional smoothers based on the use of lag windows are used, the estimation capacity is decreased. This is because the sidelobes of the spectral windows corresponding to lag window smoothers are typically orders of magnitude higher than those from the lower order prolate windows and frequently worse than the moderately high-order prolate windows. Typically, compared on the basis of their spectral windows, the use of lag window smoothers entails an order of magnitude loss in estimation capacity. If the disastrous numerical properties of such smoothers were included, the result would be even worse.

Despain and Bell [93] have shown an interesting connection between channel capacity and resolution under the assumption of strict bandlimiting. In the spectrum estimation problem the information available from filter “0” in the sample is $N \cdot EC_0$. If these data are reprocessed, for example by extrapolation, to have a longer effective length, say N' , with a higher resolution $\Delta f'$, then the effective signal-to-noise ratio and capacity EC'

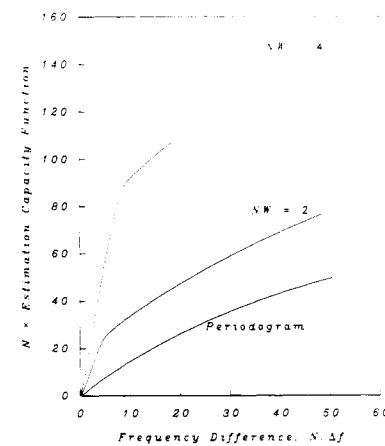


Fig. 11. Comparison of the estimation capacity function of the periodogram with those for the order-zero prolate estimates with $NW = 2$ and $NW = 4$. In all cases the sample size $N = 100$.

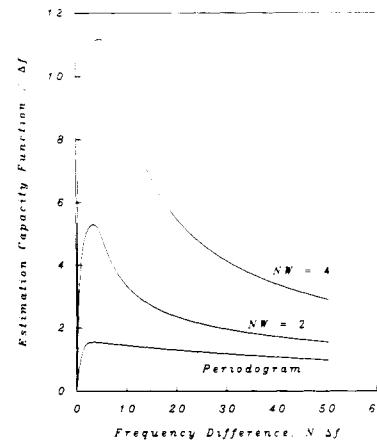


Fig. 12. A second comparison of the estimation capacity functions of the periodogram and order-zero prolate spectrum estimates. The data plotted here are the estimation capacity function divided by the frequency separation of the estimators.

are lower. Thus the reprocessed information content is $N'EC'$ or $EC'/\Delta f'$. Since information can at best be conserved, we have

$$\Delta f' \geq \frac{EC'}{N EC_0}$$

or, writing both information measures in terms of generalized signal-to-noise power ratios,

$$\Delta f' \geq \frac{1}{N} \frac{\ln(1 + \text{SNR}_{\text{output}})}{\ln(1 + \text{SNR}_{\text{input}})}.$$

From this formula it may be seen that superresolution requires exceptionally high input signal-to-noise ratios for the result to be statistically meaningful. This implies that “superresolution” requires “super” filters and agrees with the logarithmic continuity to be mentioned in Section X. Finally, for the resolution to be meaningful, the original data must have low enough measurement and aliasing noise to have equivalent Shannon capacities.

For the case when groups of estimators are used at each frequency, the situation becomes slightly more complex and less satisfactory. Following Pinsker, we replace the coherence de-

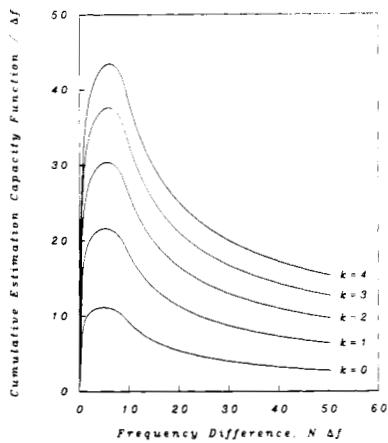


Fig. 13. The cumulative estimation capacity functions of two sets of eigenspectrum estimates as a function of maximum order and frequency separation of the sets. In this case, a white spectrum and a noise level of 10^{-11} is assumed.

pendent terms, $\ln(1 - |C|^2)$ by

$$\sum_{k=0}^{K-1} \ln(1 - |C_k|^2)$$

where the C_k 's are the *canonical coherences* between the two groups of estimators [57], [231]. Since the squared canonical correlations are the eigenvalues of a matrix of the form $M = \Sigma_{00}^{-1} \Sigma_{01} \Sigma_{11}^{-1} \Sigma_{10}$ where the Σ 's are the spectral density matrices between the two sets of filter outputs, the sum in the preceding equation becomes $\log |I - M|$. Using ALTRAN [60] to do the algebra, one arrives at a particularly simple result: assuming identical noise spectra, $N(f)$, the mutual information between the two groups of filters, $M(0; 1)$, is given by

$$M(\{H_{0,\cdot}\}; \{H_{1,\cdot}\}) = \frac{1}{2} \int_{-1/2}^{1/2} \ln \left[\frac{(N(f) + |H_{0,\cdot}(f)|^2 S(f))(N(f) + |H_{1,\cdot}(f)|^2 S(f))}{N(f)(N(f) + |H_{0,\cdot}(f)|^2 S(f) + |H_{1,\cdot}(f)|^2 S(f))} \right] df$$

where the index “.” denotes the usual summation convention

$$|H_{m,\cdot}(f)|^2 = \sum_{k=0}^{K-1} |H_{m,k}(f)|^2.$$

To compute the mutual information rate between the process and the output of a group of filters, a similar method is used. However, because the outputs of all the filters in the set are subordinate to the original process, the partitioning of the spectral matrix is particularly simple in that Σ_{11} is a scalar and Σ_{12} and Σ_{21} vectors resulting in

$$I(\{H_{0,\cdot}\}) = \frac{1}{2} \int_{-1/2}^{1/2} \left[1 + \frac{|H_{0,\cdot}(f)|^2 S(f)}{N(f)} \right] df.$$

Using these formulas to compute the estimation capacity function in the limiting case of zero noise gives a result identical to that obtained above, with the transfer functions of the individual filters replaced by the sums of power transfer functions. Thus as would be expected from the relation between the eigenestimates and the periodogram shown in Section VIII, the estimation capacity of simple sums of eigenspectra is *lower* than the capacity of the zeroth estimate alone. This is a result of the sidelobes of the spectral window of the unweighted sum (shown earlier in Fig. 3) being much higher than those of the order-0 window. If, on the other hand, one computes the addi-

tional information obtained on the assumption that there is a definite noise floor (due either to processing or measurement noise), then the estimation capacity function *increases* with the number of eigenspectrum estimates roughly until the average sidelobes of the window exceed the noise floor.

As a compromise between these extremes Fig. 13 shows cumulative estimation capacities, defined as

$$CEC_k(\Delta f) = CEC_{k-1}(\Delta f) + EC'_k(\Delta f) - EC'_k(0)$$

where $EC'_k(\Delta f)$ is given by

$$\frac{1}{4\Delta f}$$

$$\cdot \int_{-1/2}^{1/2} \ln \left[\frac{(N(f) + |H_{c,k}(f)|^2 + |H_{c,k}(f + \Delta f)|^2)^2}{(N(f) + |H_{c,k}(f)|^2)(N(f) + |H_{c,k}(f + \Delta f)|^2)} \right] df$$

for $k = 0, 1, \dots, 4$, $NW = 4$, and the noise spectrum, $N(f) = 10^{-11}$. The cumulative spectral windows, $|H_{c,k}(f)|^2$, are simply the sum of the spectral windows from 0 through $k-1$. Note that, with noise included, the effective capacity of the order-0 window is slightly lower than was the case shown in Fig. 12 for zero noise and that both the order-1 and -2 windows contribute almost as much as the order-0 window. This contribution is achieved because, while the sidelobes of the windows are somewhat above the noise spectrum, the window width is somewhat wider, and the additional information is obtained. In this regard, recall the analogy with Shannon capacity and note that the cumulative spectral window approaches rectangular, i.e., an ideal bandpass, *within* $(-W, W)$ as k increases. However, as k is increased further, the improvement of the in-band filter shape no longer compensates for the

decreased sidelobe level so that the capacity drops rapidly. Thus the combined contribution of the order-3 and -4 windows is only slightly more than that of the order-0 window. Again, we emphasize that all the filters are constrained to the domain of observation, in this example 100 data points. In addition, we must emphasize that this calculation is for a *white* spectrum. If the spectrum is highly colored, the capacities depend strongly on both estimation frequencies and, as one would expect from the general equations, much larger differences between the windows are seen.

X. HIGH-RESOLUTION AND FREE-PARAMETER EXPANSIONS

We recall the high-resolution estimate (3.4) written using the weights found in Section V

$$\hat{S}_h(f; f_0) = \frac{2}{v(f)} \left| \sum_{k=0}^{K-1} U_k(N, W; f - f_0) d_k(f) y_k(f_0) \right|^2. \quad (10.1)$$

Since this estimate is based directly on an approximate solution of the basic integral equation, it has the highest attainable resolution of any estimate based *exclusively* on the data.

It must be emphasized that if *additional a priori* information is available, such as knowledge of line components or a valid parametric model for the data, estimates with higher *apparent* resolution may be made. In the more common case, however,

where little is known about the data such *a priori* assumptions can be exceedingly dangerous and lead to seriously misleading conclusions. In the general case where such additional information is unavailable, the problem has been thoroughly studied [34], [35], [230], [291] in addition to those referenced in Section II. These studies divide the general problem into two distinct problems, extrapolation and reconstruction, which differ markedly in their convergence properties. In the extrapolation problem the solution exhibits Holder continuity with respect to the input signal-to-noise ratio so that good extrapolation is possible at reasonable noise levels. Specifically, if we denote the accuracy of the data by ϵ , Holder continuity implies that the accuracy of the extrapolation can be proportional to ϵ^γ with $0 < \gamma < 1$. The reconstruction, or inverse problem, on the other hand, converges only logarithmically in terms of the input signal-to-noise ratio. Here the accuracy of the reconstruction is at best proportional to $1/|\ln \epsilon|$ (see also [141], [169]). Thus the attainability of resolution significantly higher than the Rayleigh limit requires unrealistically high signal-to-noise ratios. Consequently, the existence of a well-behaved extrapolation does not guarantee a high-quality reconstruction of the original spectrum.

There are, however, common circumstances where the basic estimates may be combined into a *free parameter expansion* whose apparent resolution is exceptionally high. This procedure is useful for studying behavior near discontinuities in the spectrum, for example the rolloff characteristics of a filter or nearly bandpass process. Here the apparent superresolution is obtained because the spectrum in an adjacent *region* is low. As an exploratory device, the high-resolution estimate is useful for investigating the structure of low-level signals in a highly colored background and the "skirts" of near-bandpass spectra.

Recall that the high-resolution estimate $\hat{S}_h(f; f_0)$ is valid for $f_0 - W < f < f_0 + W$. Since, for a fixed frequency f , any of the continuum of estimates $\hat{S}_h(f; f_0)$ with $f - W \leq f_0 \leq f + W$ is applicable, the high-resolution estimate has a *free parameter* f_0 . Moreover, these estimates do not all give the same result; indeed, we exploit the fact that they can differ by many orders of magnitude in constructing the composite, or *free parameter expansion*. It must be emphasized that the differences between the estimates must not be taken as an indictment of the procedure; the basic equation is a projection operator which does not have a unique inverse [124], [159], [314]. The procedure is simply to try to choose the statistically most reasonable approximate solution.

We note further that situations of this type are not restricted to solutions of inverse problems but occur widely. For example, the expansions of probability density functions of likelihood ratio tests [50] depend on an unspecified parameter ρ . Fields [108] uses "free parameter" asymptotic expansions for hypergeometric functions. To use a spectrum estimation example, autoregressive spectrum estimates can vary widely as a function of the order. To resolve this ambiguity two methods are in common use: either use a "stopping rule" [33] to choose one of the estimates or, alternatively, the reciprocal averaging technique [62] to convert the autoregressive estimates to the Capon [65], [66] maximum-likelihood form.

There are several ways to determine the initially free parameter: selection of the estimate having the highest entropy; weighted linear averages; or weighted harmonic averages; to name but a few possibilities. To fix the parameters we assume that the final estimate will be a weighted average of the high-resolution estimates over the range $f - W < f_0 < f + W$. We

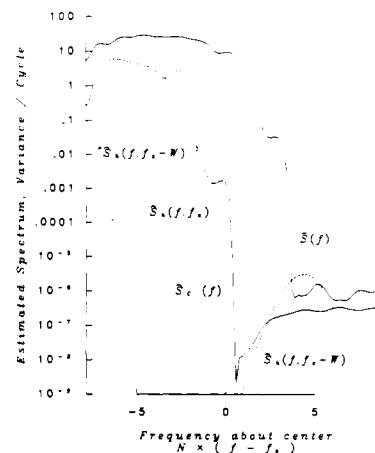


Fig. 14. Components of the free parameter estimate in the vicinity of the band edge at $f = 0.078125$ cycles for the data shown in Fig. 5. The figure shows the basic high-resolution estimates $\hat{S}_h(f; f_0)$ at center frequencies of $f_x - W$, f_x , and $f_x + W$. Also shown are the composite estimate $\hat{S}_C(f)$ and the average estimate $\bar{S}(f)$ shown earlier in Fig. 8. The composite estimate is shown over the full frequency range in Fig. 24.

choose for weights the estimated Fisher information

$$w(f_0) = \frac{v(f_0)}{\bar{S}(f_0)^2} \quad (10.2)$$

so that the composite estimate $\hat{S}_C(f)$ is given by

$$\hat{S}_C(f) = \frac{\int_{f-W}^{f+W} w(f_0) \hat{S}_h(f; f_0) df_0}{\int_{f-W}^{f+W} w(f_0) df_0}. \quad (10.3)$$

This choice of weight imposes the constraint on W that $\bar{S}(f)$ should have sufficient degrees of freedom for $w(f_0)$ to have a reasonable distribution.

A. The Section VI Example, Continued

As an example, we reconsider the problem of Section VI with emphasis on the band edge at $f_x = 0.078125$. The smoothed estimate $\bar{S}(f)$ and the associated degrees of freedom $v(f)$ used to generate the weights for this example were shown earlier in Figs. 8 and 9. Fig. 14 shows three of the high-resolution estimates for center frequencies of $f_x - W$, f_x , and $f_x + W$, the smoothed estimate $\bar{S}(f)$, and the combined estimate $\hat{S}_C(f)$. Looking at the high-resolution estimates, one can see that the one centered at f_x tries to follow the step but lacks range. Noting, however, that orthogonal systems generally have difficulty fitting discontinuities this estimate, if plotted on a linear power scale, is relatively good. This effect is clearer in the contour plot, Fig. 15, of the common logarithm of the high-resolution estimates $\hat{S}_h(f; f_0)$. In this, plot frequency f is plotted on the horizontal axis and the center frequency of the expansion domain f_0 on the vertical, both scaled to units of $N \cdot (f - f_x)$. The plotted contours are at integer values of $\log_{10} \hat{S}_h(f; f_0)$ so that the thin dashed lines on the edges of the expansion domain represent values of 10^{-9} while the heavy solid contour roughly overlapping the diagonal in quadrant 3 represents points where $\hat{S}_h(f; f_0) = 10$. Note the characteristics in quadrant 1 for fixed values of f as a function of f_0 . Returning to Fig. 14, consider the estimate centered at $f_x + W$.

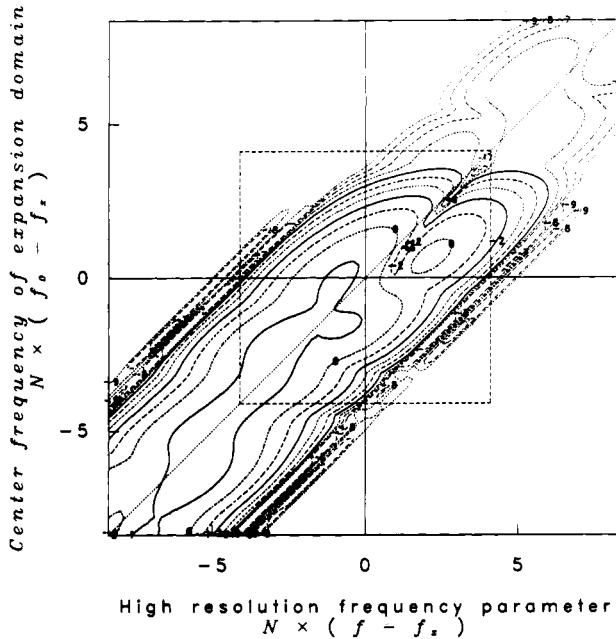


Fig. 15. Contours of the high-resolution spectrum estimate $\log_{10} \hat{S}_h(f)$ for the frequency region of Fig. 14. The center of the expansion domain f_0 is plotted on the ordinate with the ordinary frequency f being plotted on the abscissas.

With this center frequency and the extremely sharp spectral window associated with the prolate windows, Fig. 2, this estimate is largely unaffected by the higher spectral levels present at frequencies below f_x .

The combined estimate shows almost the correct break frequency; on the higher portions of the spectrum, the shape is reasonable and in the low white noise portion, the estimate is quite flat. Again, we emphasize that this resolution has been obtained from a data set of only 100 points. One is led at this point to wonder if the estimate could find additional normally obscured "features" of the spectrum. As a specific case, consider the problem of detecting a low-level line component, say of apparent magnitude 10^{-3} , "hidden" close under the band edge, perhaps at $f_x + 1/T$. While such detection is frequently possible, a glance at the weight function shows that most of the contribution to the combined estimate comes from high-resolution components centered above $f_x + W$. As one would expect for such an extreme case, the estimated stability of the high-resolution estimates centered in the frequency region around $f_x + W$ is only a few degrees of freedom and consequently the statistics of the detection process are not good (but still much better than those available from other methods!).

In practice, it is unadvisable to extend the expansions to $|f - f_0| = W$ but rather to stop near 0.8 to 0.9W. Further, in regions where only a few degrees of freedom are available, it is advisable to rescale $\hat{S}_h(f; f_0)$ by dividing by a factor proportional to

$$\sum_{k=0}^{K-1} |w_k(f_0) U_k(N, W; f - f_0)|^2$$

compensating, in part, for the extremely rapid rolloff of the lower order prolate windows near W .

Before leaving this section we mention that the composite estimate is plotted over the full frequency range in Section XIV with additional processing for the line components. A second example is given in Section XII. We leave as an open

question the problem of combining the high-resolution estimates subject to a maximum-entropy constraint.

XI. CHOICE OF BASIS FUNCTIONS

Of the numerous areas where the philosophy expressed in this paper differs from that currently in vogue, one of the most important is in the choice of basis functions. In this context we note that most of the successful physical descriptions of wave phenomena have two features in common: use of a coordinate system matching (or close to matching) the geometry of the problem; and second, expansions based on the eigenfunctions, or natural modes, appropriate to that coordinate system. Although we are interested in a harmonic decomposition and Fourier transforms, the finite Fourier series provides a poor set of basis functions. Instead, they are best regarded as providing the coordinate system so that the space is that of the finite discrete Fourier transforms of index-limited functions. For the frequency range $(-W, +W)$ of this space, the eigenfunctions are the discrete prolate spheroidal wave functions. From usage in the preceding sections, however, rather than using all N discrete prolate spheroidal wave functions, we have taken the set of $2NW$ functions having eigenvalues close to 1 and implicitly augmented this set with frequency-translated versions of itself. Thus the method may be regarded as a frequency-domain dual of Welch's [346], [347] method and the techniques used in [13], [84]. Unfortunately, this augmented set is defective in that the basis functions are not orthogonal, and, more seriously, because frequency is a continuous parameter, there are potentially an infinity of functions so that the set is indeterminate. We thus consider the following criteria for a basis set for spectrum estimation and inference problems:

1) *Completeness*. The set of N basis functions should be linearly independent so that the N data points are recoverable from the expansion coefficients. This implies that the expansion coefficients will form a trivially sufficient set.

2) *Numerical Conditioning*. As a practical consideration, the basis functions must be not only linearly independent but also nearly orthogonal. If they are not, effective sufficiency may be lost to numerical instability.

3) *Perturbation Insensitivity*. A useful set of basis functions should not depend excessively on certain conditions being exactly met. For example, the orthogonality of the Fourier series is sensitive to a nonuniform weight. Similarly, estimation procedures based on using frequencies spaced $1/N$ apart and dependent on the zeroes of the Dirichlet or sinc kernels fail this requirement; see [156].

4) *Low Spectral Mixing*. Since we are interested in spectrum estimation, it is natural to request that the *power spectral densities* of the basis functions be as distinct as possible. Since we are working in a time-limited domain, the spectrum of the basis functions cannot vanish over any interval and, consequently, some mixing is unavoidable; see [36], [142], [224].

The first two conditions are easy and are satisfied by any orthogonal set. However, by considering the trivial basis functions

$$b_n(t) = \delta_{n,t}, \quad n, t = 0, 1, \dots, N-1$$

it follows that, since their spectra are all identically 1, orthogonality is not enough to satisfy the last condition.

As a result, we consider a measure of the difference between the different basis spectra. This problem is similar to that occurring in speech processing but with the opposite emphasis:

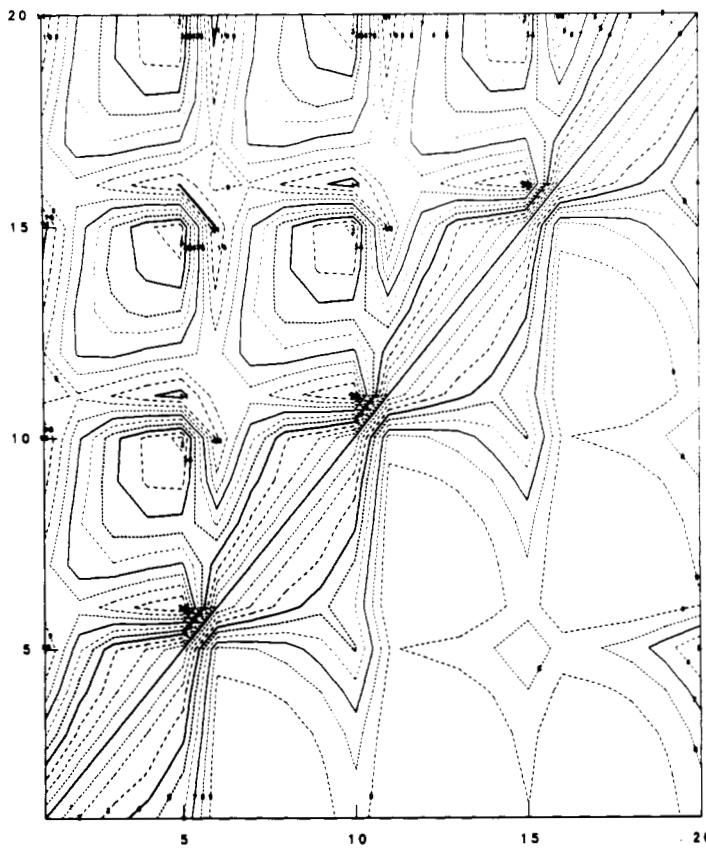


Fig. 16. Two measures of the distance between the spectra of the basis set consisting of frequency-translated prolate spheroidal wave functions. The data above the main diagonal are the estimation capacity functions between the two basis spectra; below the main diagonal the values represent Ornstein's measure. In this case, four bands and five prolate functions are used in each band. The block structure is apparent. The abscissa and ordinate labels give the basis indices. Scaling on the contours is arbitrary. (See Fig. 10 for numerical values.)

while measures such as Ikatura-Saito test different spectra for similarity, the desire here is to have the spectra of the basis functions as dissimilar as possible.

Thus as an alternative to the continuous frequency translation assumed above, we use the set of functions consisting of the first $K \approx [2NW]$ discrete prolate spheroidal wave functions, and these functions translated in frequency by multiples of

$$\Delta f = \frac{K}{N} \approx 2W$$

so that there are N/K distinct frequency bands, each containing K functions. Note that while this convention requires that N be factorable, the constraint is not serious as the bands are permitted to overlap slightly. The idea is that the spectra of the basis functions be reasonably distinct, but the basis functions are not necessarily orthogonal. In the time domain the basis functions $Q_t^{(k)}$ are given by

$$Q_t^{(k)} = e^{-i2\pi t m \Delta f} v_t^{(j)}(N, W)$$

where

$$k = j + Km.$$

for $j = 0, 1, \dots, K-1$, $m = 0, 1, \dots, N/K-1$, and $t = 0, 1, \dots, N-1$. Because of the orthogonality of the prolate functions within any band, i.e., m constant, the basis functions will be orthogonal. Between bands, however, the inner product between the basis functions will not generally be zero. Defining

$$h_{k_1, k_2} = \sum_{t=0}^{N-1} Q_t^{(k_1)} Q_t^{*(k_2)},$$

These products are most easily bounded in the frequency domain using the transforms

$$Q_k(f) = \sum_{n=0}^{N-1} e^{-i2\pi f[n-(N-1)/2]} Q_t^{(k)} = U_j(N, W; f + m \Delta f)$$

and Parseval's theorem to give

$$h_{k_1, k_2} = \int_{-1/2}^{1/2} U_{j_1}(N, W; f + m_1 \Delta f) U_{j_2}(N, W; f + m_2 \Delta f) df.$$

This is the function $\Lambda_{j_1, j_2}((m_1 - m_2)\Delta f)$ encountered in Sections IV and VII for a white spectrum. It follows that H has block identity matrices of size $K \times K$ along its diagonal and small elements elsewhere and so is almost an identity matrix. Because the discrete prolate spheroidal wave functions are real, H is symmetric.

The effective sufficiency of the proposed basis (which determines if and how well the original data can be recovered from the expansion coefficients) is given by the *condition number* of the basis matrix. For this problem the condition number is $\sqrt{\gamma_{\max}/\gamma_{\min}}$ where γ_{\max} and γ_{\min} , the largest and smallest eigenvalues of HH^\dagger , have been typically less than 2.

We now consider the differences between the spectra of the

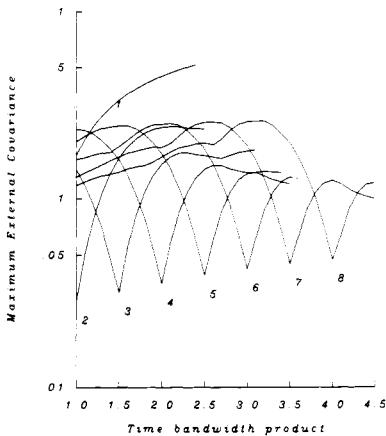


Fig. 17. The maximum off-diagonal values of covariance $\Lambda_{jk}^2(\Delta f)$ occurring in the matrix HH^\top for $\Delta f > k/N$. The abscissa is the time-bandwidth product NW of the basis functions and the curves are parameterized by the number of basic prolate spheroidal wave functions used. For example, the curve labeled "3" applies to sets of basis functions consisting of 3 discrete prolate spheroidal wave functions and their frequency-translated replications.

basis functions and recall that obtaining basis functions having distinguishable spectra was the original goal of this section. The problem here is the selection of a suitable measure of the differences, particularly when one adds the additional constraint of localization. By this we mean, loosely, that if two basis spectra are similar in a compact subset of the frequency domain, the effect is less serious than the same net similarity spread over the entire frequency domain.

The region above the main diagonal of Fig. 16 is a contour plot of the estimation capacity function defined in Section IX plotted against the indices k_1 and k_2 of the basis functions. The values below the main diagonal are for Ornstein's $\bar{\rho}$ distance. Note, first, that the block structure is apparent, and second, that the basis functions are separated by significant distances. In Fig. 17 we show plots of the maximum value of $\Lambda_{jk}^2(\Delta f)$ attained for $\Delta f \geq k/N$ as the basic prolate parameter W varies. Note that, at the natural design points, the maximum covariances are low and the minima are not excessively sharp.

XII. RELATIONS TO MAXIMUM LIKELIHOOD

In current spectrum estimation terminology a "maximum-likelihood" estimate is likely to refer either to a Capon [65], [66] estimate (see [62], [185] for further details) or to a parametric ARMA estimate where the parameters are estimated using maximum likelihood. In this section we make some observations on a *nonparametric* maximum-likelihood procedure based on the Karhunen-Loëve expansion. This is not done without significant reservations. First, maximum likelihood is not the only estimation technique, and it is possible that others are better suited for spectrum estimation problems; second, maximum-likelihood estimates are known to act poorly when the number of parameters to be estimated is a reasonable fraction of the sample size while the spectrum is a function whose parameterization is potentially infinite. Third, data near the ends of the observation epoch are related as much to or, particularly if the series is very short, more to unobserved data than the remainder of the observations. This is analogous to the limited information [15]-[17] models and somewhat to Kalman filters. Also, while prediction and inferential problems from a finite sample have been treated proba-

bistically [99], corresponding spectrum estimation procedures appear to be unknown.

Compensating for these reservations we have is the fact that the double orthogonality property of the prolate spheroidal wave functions is shown to be common to all Karhunen-Loëve eigensystems. This implies some interesting connections between maximum-likelihood and extrapolation estimates. Second, from asymptotic arguments, periodograms are frequently associated with maximum-likelihood estimates [100], [167], [349]; see also [27], [51], [90], [223], [239]. We give reasons and an example which show that for small samples the prolate-based estimates have a much higher likelihood than the periodogram.

While the general properties of Karhunen-Loëve expansions are well known [213], [242], [341], we give the following summary to establish notation. If one is given a sample of size N from a time series with a known autocovariance function

$$R(\tau) = \mathbb{E}\{x_t x_{t+\tau}\}$$

then the discrete Karhunen-Loëve expansion is given by

$$\mathbf{X} = \Psi \mathbf{C}$$

where \mathbf{C} is the coefficient vector

$$\mathbf{C} = \Psi^\dagger \mathbf{X}$$

and Ψ is the matrix whose m th column is the eigenvector associated with the eigenvalue θ_n in the matrix equation

$$\Psi \Theta = R \Psi.$$

In this equation Θ is a diagonal matrix of the eigenvalues θ_n , and R is the Toeplitz matrix of autocovariances. Written explicitly the eigenvectors are defined by

$$\theta_n \psi_n(t) = \sum_{u=0}^{N-1} R(t-u) \psi_n(u) \quad (12.1)$$

with the normalization constraints

$$\sum_{t=0}^{N-1} \psi_n(t) \psi_k(t) = \delta_{n,k}. \quad (12.2)$$

The expansion coefficients c_n

$$c_n = \sum_{t=0}^{N-1} \psi_n(t) x(t)$$

have expected value $\mathbb{E}\{c_n\} = 0$ and covariances

$$\mathbb{E}\{c_n c_j\} = \theta_n \delta_{n,j}$$

so that the likelihood of the observation may be written

$$L(\{x\}) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\theta_n}} e^{-1/2(c_n^2/\theta_n)}.$$

At this point one could attempt to directly maximize the likelihood as a function of the autocovariances R (which enter both through the eigenvalues and the coefficients) but the perturbation equations are moderately complex. Solving the perturbation equations for a stationary point gives the conditions $\hat{c}_n^2 = \hat{\theta}_n$. See [39] for some insight into the inverse algebraic eigenvalue problem, also [332].

As an alternative we consider the Karhunen-Loëve eigen-equation written in the frequency domain. Using the Wiener-

Khintchin relation

$$R(n) = \int_{-1/2}^{+1/2} S(v) e^{i2\pi nv} dv$$

and defining the discrete Fourier transform of the time domain eigenvectors, the "eifunctions"⁵

$$\Phi_n(f) = \sum_{t=0}^{N-1} \psi_n(t) e^{-i2\pi f[t-(N-1)/2]}$$

the Fourier transform of (12.1) becomes

$$\theta_n \Phi_n(f) = \int_{-1/2}^{1/2} S(\xi) \frac{\sin N\pi(f - \xi)}{\sin \pi(f - \xi)} \Phi_n(\xi) d\xi. \quad (12.3)$$

Note that this equation expresses the reverse operation of the usual reduction of an integral equation to an algebraic eigenvalue equation: here we have elevated the algebraic equation to an integral equation. This equation, however, shows several features of the expansion process not obvious in the time-domain matrix equation.

First, there is the obvious similarity between the kernel of this integral equation, the basic integral equation, and the integral equation defining the discrete prolate spheroidal wave functions. No such apparent similarity exists between the corresponding time-domain equations (2.9) and (12.1). This similarity will be used later.

Second, we have the result:

Theorem: All Karhunen-Loève eifunction systems are doubly orthogonal; first with respect to unit weight

$$\int_{-1/2}^{1/2} \Phi_n(f) \Phi_k^*(f) df = \delta_{n,k} \quad (12.4)$$

and second with the spectrum as weight.

$$\int_{-1/2}^{1/2} \Phi_n(f) \Phi_k^*(f) S(f) df = \theta_n \delta_{n,k}. \quad (12.5)$$

Proof of the orthonormality relation is obvious from (12.2) and Parseval's theorem. The second orthogonality equation is easily proven by multiplying both sides of (12.3) by $\Phi_k^*(f)$ and integrating. Since the Dirichlet kernel serves as the identity element in the space of Fourier transforms of index-limited functions

$$\int_{-1/2}^{1/2} \Phi_k^*(f) \frac{\sin N\pi(f - \xi)}{\sin \pi(f - \xi)} df = \Phi_k^*(\xi)$$

the result follows.

The first implication of this theorem is that the curious double orthogonality property of the prolate spheroidal wave functions is not unique. In particular when $S(f) = 1$ for $|f| < W$ and 0 elsewhere, (12.3) is identical to (2.4).

This double orthogonality property implies some interesting relations between the Karhunen-Loève eigenrepresentation, parametric autoregressive and moving average representations, and extrapolation techniques. First, the hierarchy is obvious:

⁵We use the name "eifunction" to distinguish the frequency-domain eigenfunction defined as the Fourier transform of the finite time-domain eigenfunction.

Karhunen-Loève is a double orthogonalization as opposed to autoregressive representations which may be obtained by simple Gram-Schmidt orthogonalizations of the functions $e^{i2\pi nf}$ with respect to the spectrum as weight; see [186], [321].

Consider the expansion of the Karhunen-Loève eifunctions in the basis set consisting of the frequency-translated DPSWF's considered in Section XI. Our motivation for considering such an expansion is the well-known Szegő theorem [136], [138], [157], [158], [234], [319], [320]. According to this theorem, the eigenvalues will be asymptotically equal to the spectrum at frequencies spaced $1/T$ apart. This suggests that the corresponding eifunction is localized about these frequencies and consequently might be well approximated by prolate spheroidal wave functions translated to the same center frequency. Note that, since the set of frequency-translated DPSWF's is complete in our space, there is no question of their ability to do the expansion; the hope is simply that they will provide a "better" expansion than some other possible sets. While "better" is subjective, two quantities are desired: first, greater understanding of the behavior of the eigensystem for finite sample sizes; second, improved numerical techniques for computing the eigenfunctions from a sample spectrum. The motivation for the latter is provided by Gershgorin's theorem, see [165], which defines regions containing the eigenvalues. Applied directly to the Toeplitz system, the radius of these regions is approximately

$$\sum_{\tau=1}^{N-1} |R(\tau)|$$

which is typically large. Contrasted to this poorly conditioned system, the frequency-domain eigenvalue equation will be seen to have its eigenvalues localized close to that expected from the Szegő theorem by a *Gershgorin radius of the same order*. We begin by expanding the eigenfunctions in the set of basis functions described in the previous section

$$\Phi_n(f) = \sum_{j=0}^{N-1} g_{n,j} Q_j(f)$$

so that the integral equation becomes

$$\theta_n \sum_{j=0}^{N-1} g_{n,j} Q_j(f) = \sum_{j=0}^{N-1} g_{n,j} \int_{-1/2}^{1/2} S(\xi) \frac{\sin N\pi(f - \xi)}{\sin \pi(f - \xi)} Q_j(\xi) d\xi. \quad (12.6)$$

Multiplying by $Q_l^*(f)$, integrating over $(-\frac{1}{2}, \frac{1}{2})$, and again using the Dirichlet kernel as the identity element, we obtain the algebraic eigenvalue equation

$$\theta_n \sum_{j=0}^{N-1} g_{n,j} h_{j,l} = \sum_{j=0}^{N-1} g_{n,j} \int_{-1/2}^{1/2} S(\xi) Q_j(\xi) Q_l^*(\xi) d\xi.$$

This equation is independent of the properties of the prolate functions and works with any set of basis functions $Q_j(f)$, $j = 0, \dots, N-1$. In particular we note the set $Q_j(f) = \hat{\Phi}_j(f)$, the set of Karhunen-Loève eifunctions estimated in a previous iteration. Defining the matrix Q , having as elements the integrals

$$Q_{j,l} = \int_{-1/2}^{1/2} S(\xi) Q_j(\xi) Q_l^*(\xi) d\xi$$

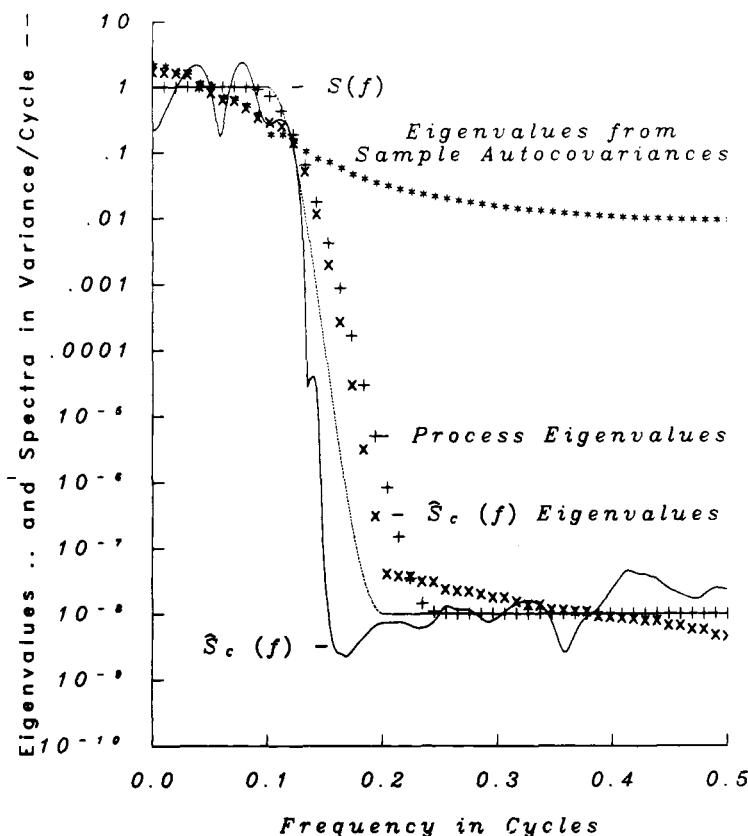


Fig. 18. A comparison of the eigenvalues for the Karhunen-Loëve expansion corresponding to the periodogram and the composite estimate. In this example, the true spectrum is shown by the dashed line, and the composite estimate by the solid line. The eigenvalues corresponding to the true spectrum are plotted with +, and those to $S_c(f)$ by x. The third set of eigenvalues, shown by *, are computed using the sample autocovariances. Here $N = 50$.

we have the generalized eigenvalue equation

$$\theta_n \sum_{j=0}^{N-1} g_{n,j} h_{l,j} = \sum_{j=0}^{N-1} g_{n,j} Q_{l,j}$$

or

$$H\Theta = QG. \quad (12.7)$$

Recalling that the eigenfunctions are orthonormal with respect to uniform weight and orthogonal with respect to the spectrum as weight, an iterative scheme will clearly drive Q to Θ , and, after the first iteration, $H = I$; see [309] for further details on such iterated Galerkin methods.

In the case where the Q_j 's are the shifted DPSWF's described in the preceding section, there is a second Galerkin method. We begin with (12.6) and, as above, multiply by the shifted DPSWF $Q_{j'}(f)$. Here, instead of integrating over $(-\frac{1}{2}, \frac{1}{2})$, we integrate over the principal domain of the shifted function $Q_{j'}(f)$, that is $(m'\Delta f - W, m'\Delta f + W)$. Use of the prolate integral equation (2.4) gives

$$\theta_n \sum_{j=0}^{N-1} g_{n,j} h'_{j,j'} = \sum_{j=0}^{N-1} g_{n,j} \lambda_{k'} \int_{m'\Delta f - W}^{m'\Delta f + W} S(\xi) Q_j(\xi) Q_{j'}^*(\xi) d\xi.$$

In this equation $h'_{j,j'}$ is

$$h'_{j,j'} = \int_{-W}^W U_k(N, W; \xi + (m - m')\Delta f) U_{k'}(N, W; \xi) d\xi$$

where $j = k + mK$ and $j' = k' + m'K$. Writing Λ for the repeated block diagonal matrix of prolate eigenvalues we have

$$H'G\Theta = Q\Lambda G.$$

This equation is similar to (12.7) above; the differences are the replacement of h by h' and the presence of the prolate eigenvalues $\lambda_{k'}$. Comparing the two gives

$$H'H^{-1} = \Lambda.$$

The difference between the two Galerkin methods is the domain of integration of the test functions. In the first, the integration was over the entire frequency domain; in the second, the integration was over a local domain of length $2W$. Thus the difference between the two equations reflects the contribution of frequencies in the exterior domain of the test function. Because the prolate eigenvalues asymptotically become exponentially close to 1 in NW for $k < 2NW$, the implication is that the Karhunen-Loëve eigenvalues are determined to a large degree by the local prolate moments of the spectrum.

In these algebraic eigenvalue equations the matrix Q is, from Section IV,

$$Q_{j,j'} = \{y_k(m\Delta f) y_{k'}^*(m'\Delta f)\}$$

or simply the covariance matrix of the eigencoefficients. Thus an alternative measure of the likelihood is in terms of the multivariate complex-normal distribution of the eigencoefficients. Similarly, the Karhunen-Loëve eigenvalues are closely approximated by those of the eigencoefficient matrix. Note,

however, that in the case where less than N series are available the matrix $\mathbf{T}\mathbf{T}^\dagger$, $\mathbf{T}^\dagger = (y_0^*(0), y_1^*(0), \dots, y_{N-1}^*(0), y_N^*(\Delta f), y_1^*(\Delta f), \dots)$, is singular. Direct evaluation of the $Q_{j,k}$ using the sample spectrum gives a stable result, as only N autocovariances are required to specify a Toeplitz matrix as opposed to the N^2 required for a general covariance matrix.

The difference between this technique and that based on computing the eigencoefficients directly from the Toeplitz matrix can be extremely large. In one case with $N = 24$, a process spectrum similar to that shown in the following example, and computations done in single precision⁶ the following differences were obtained for the two computations of the 14th eigenvalue (true value $\approx 8.665654 \times 10^{-6}$). Using the direct Toeplitz matrix, a Gershgorin radius of 1.35 and eigenvalue 1.386×10^{-3} were obtained; with the iterated Galerkin method described here, the Gershgorin radius at the end of the initial iteration was 6.33×10^{-9} , and the eigenvalue obtained was 8.66570×10^{-6} . Observe that, with the direct Toeplitz matrix calculation, the numerical ill-conditioning of the problem alone is enough for the computed eigenvalue to be in error by a factor of 160.

As a detailed example, consider a process having the spectrum shown by the smooth dashed curve in Fig. 18. The entropy of this spectrum

$$\exp \left\{ \int_{-1/2}^{1/2} \ln S(f) df \right\}$$

is $e^{-12.916} = 2.458 \times 10^{-6}$. On it are superimposed estimates of spectra and eigenvalues as follows:

1) The first set of eigenvalues, shown by "+," was computed from the known spectrum using the iterated Galerkin technique described above. In these calculations double precision⁶ versions of Singleton's FFT [115] and EISPACK [310] matrix routines were used. The corresponding eigenfunctions were used to generate data. For these typical values for the normalized log-likelihood function

$$L = \sum_{k=0}^{N-1} \ln \left(\frac{1}{\theta_k} \right) - \frac{c_k^2}{\theta_k}$$

were about +546 for the 50-point sets used.

2) A free parameter spectrum estimate, Section X, was computed and is shown by the wiggly solid line. For these spectral estimates the entropies range from 5.1×10^{-7} to 1.4×10^{-6} without the bias corrections normally used, [175].

3) This spectrum estimate was used to compute eigenvalues and estimates of the log-likelihood. For this case the eigenvalues are shown by X. These eigenvalues, which appear to decay smoothly because they have been sorted before plotting, match the correct spectrum and are reasonably close to the true eigenvalues. Expanding the sample in the corresponding eigenfunctions gave values of the log-likelihood from a minimum of 364 to 547 with a median value of 523. The explanation for the rather wide range appears to be that while the sample eigenfunctions have approximately the correct frequency content, that is they are becoming localized, they do not match exactly so that there are typically a few for which $\hat{c}_k^2/\hat{\theta}_k$ is large.

⁶On a Honeywell DPS 8/70 36 bit processor. The single and double precision mantissas are 28 and 64 bits, there is an 8-bit binary exponent, and the floating-point registers carry 8 guard bits with rounding on storage.

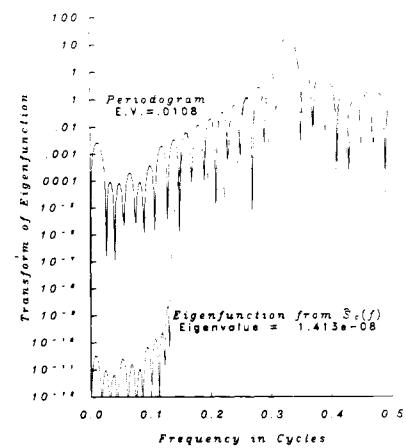


Fig. 19. Spectra of the 17th eigenfunction for the estimates based on the sample autocovariances and $S_C(f)$ of Fig. 18. The solid curve shows the energy distribution for the eigenfunction computed using the sample autocovariance function. Note that large values occur in the low-frequency region where the spectrum (Fig. 18) is large. By contrast, the dashed curve shows the energy distribution with frequency of the eigenfunction computed from $S_C(f)$. Because the influence of the low-frequency region is suppressed, expansion coefficients based on this function will be dominated by frequency components in the proper region.

4) To compute the last set of eigenvalues, shown by "*", the data were first transformed to autocovariances, using the positive-definite form common in speech processing

$$\hat{r}_\tau = \frac{1}{N} \sum_{n=0}^{N-1-\tau} x_n x_{n+\tau}$$

and the eigenvalues of the Toeplitz matrix were computed directly, again using double precision. Here the estimated log-likelihoods were remarkably consistent, most being clustered about 216, but with the minimum observed being 146. The explanation for this behavior is that the eigenfunctions are so badly biased that they all respond to the large values of the spectrum at low frequencies. Thus with estimates derived from the periodogram or sample autocovariances, a high value of the log-likelihood does not imply a reasonable model for the data. Note that there is little resemblance between this set and those computed from the known spectrum.

To continue this comparison, Fig. 19 shows the periodogram of two eigenfunctions (or the squared magnitudes of the eigenfunctions). Both correspond to the 17th smallest eigenvalue in their respective sets (and were chosen because the frequency of their maxima are nearly coincident), and both were estimated from the same set of 50 data points. The difference is that the one shown by the solid curve was computed from the sample autocovariances, the one shown by the dashed curve was computed from the composite estimate $\hat{S}_C(f)$. Note particularly the difference in the low-frequency region where the process spectrum is large and the bias in the covariance-based eigenfunction is also large. In the time domain, the difference is not so striking, but, generally speaking, the eigenfunctions computed from the sample autocovariances resemble sines and cosines in appearance whereas the eigenfunctions computed from $\hat{S}_C(f)$ typically have envelopes resembling the prolate functions. In addition, attempts to perturb the spectrum estimate from the eigenbehavior lead to further observations. First, the perturbations must be made carefully if the estimate is to be improved. Second, if the perturbations are not gentle, the estimate tends to switch to the periodogram. Thus the peri-

odogram represents a broad *local* maximum of the likelihood function whereas the *global* maximum is localized and sharp. Moreover, using the corresponding expansion coefficients "reasonable" values of $\hat{c}_n^2/\hat{\theta}_n$ are usually obtained so that the periodogram must be categorized as a consistently self-deceiving estimate. Third, it appears that asymptotic behavior of eigenvalues and eigenfunctions occur at different rates. For the simple spectrum used in this example the eigenvalues are already close to the spectrum with only 50 data points whereas the eigenfunctions bear little resemblance to sines and cosines.

As a final point we touch on a subject mentioned in the Introduction, namely that the methods described in this paper are not intended to be a complete spectrum estimation procedure but rather to be embedded in a larger package. If the initial estimate $\hat{S}_C(f)$ is used to generate an autoregressive pre-whitening filter and a prewhitened estimate is computed, the corresponding log-likelihoods have a median value of 543, close to the original median value of 552. The range is also low, extreme values observed being 518 to 567. Thus while these estimates do not satisfy the exact maximum-likelihood conditions $\hat{c}_n^2 = \hat{\theta}_n$, the observed values of $\hat{c}_n/\hat{\theta}_n^{1/2}$ do appear to have approximately a normal $N(0, 1)$ distribution. Judging from the appearances of different estimates, it appears that those coming closer to having the $N(0, 1)$ distribution may be "better" than those coming closer to satisfying the strict maximum-likelihood conditions.

A. A Relation with Moving Average Processes

A technique used to improve the resolution of spectrum estimates is extrapolation; one makes a predictor from the data and uses it to "extend" the original data. This "extended" data are then used to recompute the spectrum and a new predictor, and so on, iteratively. Details of this procedure are available in [64], [75], [113], [178], [181], [244], [289] with additional relevant material being [37], [38], [89], [350]. These formulations typically involve prolate spheroidal wave functions and some form of bandlimiting assumption. We thus consider the case when an approximate moving average representation $\beta(k)$, $k = -\infty, \dots, -1, 0, 1, \dots$ exists with transfer function

$$B(f) = \sum_{t=-\infty}^{\infty} \beta(t) e^{i2\pi ft}.$$

Since stationary processes are time reversible, we assume that the moving average representation is real, symmetric, and normalized by

$$S(f) = |B(f)|^2.$$

Using this representation in the second orthogonality relation, one has

$$\int_{-1/2}^{1/2} \Phi_n(f) B(f) \cdot \Phi_k^*(f) B(f) df = \theta_n \delta_{n,k}.$$

Converting to the time domain, one has the functions

$$\phi_n(t) = \psi_n(t) * \beta(t)$$

which are orthogonal on $(-\infty, \infty)$ and provide a natural extrapolation of the observations as the prolate functions are used to extrapolate a bandlimited sequence

$$\hat{x}_t = \sum_{n=0}^{N-1} \frac{c_n}{\theta_n} \phi_n(t).$$

A common requirement on such procedures is that the extrapolated data and the original data agree on $[0, N-1]$, and this will not generally be satisfied unless the matrices B and R commute, B being the Toeplitz matrix formed from the moving average coefficients. Note that the prolate spheroidal wave functions are unique in the sense that for the strictly band-limited case the symmetric moving average representation is identical to the autocovariance function.

There are several possible solutions to this dilemma: accept the discrepancy, which in the cases tried so far is small; use a weight in the time domain as the information at the ends of the series is related to unknown information; or, perhaps best, use a simultaneous orthogonalization procedure and replace the Karhunen-Loeve expansion with its extension by solving

$$R\Psi'\Theta' = B\Psi'$$

as a generalized eigenvalue problem; see [119], [176], [177]. If the process is nonsingular, this procedure will asymptotically result in two sets of eigenvalues, one set asymptotically being the square of the other. For analogies with the spheroidal functions see [272], [273].

XIII. HARMONIC ANALYSIS

Generally speaking, *harmonic analysis* has come to mean the study of the line components in a spectrum without regard to whether or not they are at multiples of a common frequency. Unfortunately, since the techniques used for harmonic analysis have been virtually identical to those used for general spectrum estimation, the two names have been used almost interchangeably. Nonetheless, there are many references where the emphasis is on harmonic analysis: [48], [73], [109], [110], [126], [139], [144]-[148], [160], [192], [198], [203], [207], [208], [218], [219], [249], [257], [262], [263], [277], [331], [344], [348], [355], [359]. In addition, many of the papers contained in [2]-[4], [74], [155] are relevant.

To make sense of harmonic analysis it is essential to recognize that the assumption of "pure" line components is a convenient fiction; while often good for a few cycles, it is rarely supportable over extended periods of time. This tends to further divide the subject by series length. In short series, *detection*, [268], [331], and resolution of line components are major problems; for longer series the problems of interest typically concern the *structure* of the line [24], [40], [47], [76], [98], [210], [288], [336]. For such investigations, the use of *structure functions* advocated by Lindsey and Chie [212] deserves wider attention.

In this paper we present a new approach to the problem of "mixed" spectra; that is the case where line components are embedded in stationary background noise with a continuous spectrum. At present, the best solutions to the harmonic analysis problems are probably given by eigenvalue decompositions, with the projection methods of Kumaresan and Tufts [199], [200] being among the most promising. We note that there are two "pure" eigenvalue decompositions common in time-series analysis: at one extreme we have Pisarenko's [258] "all-signal" form; on the other extreme of a purely non-deterministic signal, we have the Karhunen-Loeve expansion. Between these extremes it will probably become necessary to use Slepian's [301] techniques to combine the two methods for effective harmonic analysis in highly colored noise. The method we describe takes a step in that direction by doing a decomposition into "signal" and "noise" components. As such it is largely complimentary to the eigenvalue decomposition and linear prediction methods of harmonic analysis.

With this approach, which consists simply of applying regression techniques to the eigencoefficients, the two problems of spectrum estimation and harmonic analysis are distinct. The distinction between the two problems is that *spectrum analysis* is the study of the *second*, and higher, moments of $dZ(v)$ while the emphasis in *harmonic analysis* is on the *first* moment of $dZ(v)$.

In such cases, the process is usually described as having a nonzero mean-value function consisting of a number of sinusoidal terms at various frequencies, plus perhaps a polynomial trend, plus a stationary random process of the type we have been considering. In terms of the spectral representation this amounts, in practice, to having the extended representation

$$\mathbb{E}\{dZ(f)\} = \sum \mu_m \delta(f - f_m) \quad (13.1)$$

in place of the usual assumption $\mathbb{E}\{dZ(f)\} = 0$. With this definition, the continuous portion of the spectrum is the second absolute central moment of $dZ(f)$.

To demonstrate the approach we assume the simplest case of a single line component at frequency f_0 so that the eigencoefficients have a nonzero expected value

$$\mathbb{E}\{y_k(f)\} = \mu U_k(N, W; f - f_0). \quad (13.2)$$

Again, making the assumption that the continuous component of the spectrum near f_0 is slowly varying or "locally white," it was shown in Section IV that

$$\text{cov}\{y_k(f), y_j^*(f)\} \approx S(f) \cdot \delta_{j,k} \quad (13.3)$$

where the spectrum $S(f)$ is the *continuous* spectrum and does not include the line power.

There are two obvious limiting methods to estimate μ : "point regression" at f_0 , and "integral regression" in the neighborhood of f_0 with the obvious changes to either for both coefficient weighting and truncation. In the first case, one uses only the data at f_0 where one has the obvious relation

$$\mathbb{E}\{y_k(f_0)\} = \mu U_k(N, W; 0) \quad (13.4)$$

and can estimate the mean μ , by standard regression methods, [227]

$$\hat{\mu}(f) = \frac{\sum_{k=0}^{K-1} U_k(0) y_k(f)}{\sum_{k=0}^{K-1} U_k^2(0)}. \quad (13.5)$$

This estimate is the high-resolution estimate described in Section X at $f = f_0$ before squaring. As the eigencoefficients are combined linearly, we may write

$$\hat{\mu}(f) = \sum_{n=0}^{N-1} h_n(N, W) x(n) e^{-i2\pi f[n-(N-1)/2]} \quad (13.6)$$

where the effective harmonic analysis data window $h_n(N, W)$ is given by

$$h_n(N, W) = \frac{\sum_{k=0}^{K-1} U_k(0) v_n^{(k)}(N, W)}{\sum_{k=0}^{K-1} U_k^2(0)}. \quad (13.7)$$

This window and the corresponding spectral window are plotted in Figs. 20 and 21; note the apparent similarity between this data window and that shown in [238, Fig. 7.8]. Also, in Fig.

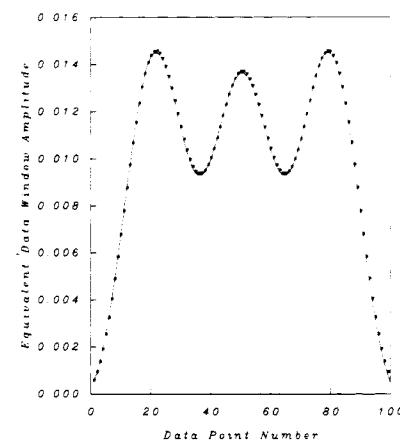


Fig. 20. The equivalent data window for harmonic analysis using single-frequency regression and the first 5 eigencoefficients. As before $NW = 4, N = 100$.

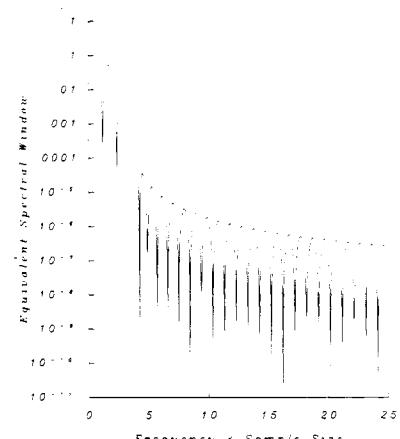


Fig. 21. The spectral window corresponding to the data window of Fig. 20. Note the low sidelobes for frequency spacings greater than W .

21, it may be seen that the central lobe has a half-width of $1.15/N$, only slightly wider than that of the periodogram, and low sidelobes outside the bandwidth W .

The "integral regression" method, which uses information not only from the y_k 's at f_0 but also from the neighborhood of f_0 is a solution of minimizing the sum of integrals

$$\sum_{k=0}^{K-1} \int_{f_0-W}^{f_0+W} |y_k(f) - \mu U_k(N, W; f - f_0)|^2 df \quad (13.8)$$

with respect to μ . The result is a form of matched filter applied to the different eigencoefficients. As above, it possesses an equivalent harmonic window, but here it is formed of convolutions of prolate functions and, consequently, has exceedingly low sidelobes. The drawback is that the method is more complex and, as it uses information from a wider bandwidth, is also subject to noise from the same bandwidth. Because the details are similar, we will not give detailed formulas for this method.

Returning to the first method, the variance of the estimated mean depends on the local continuous part of the spectrum

$$\text{var}\{\hat{\mu}(f)\} = \frac{S(f)}{\sum_{k=0}^{K-1} U_k^2(0)} \quad (13.9)$$

which is only slightly larger than $S(f)/N$. Subtracting their

estimated means from the eigencoefficients gives an estimate of the continuous spectrum. Comparing this value of the background spectra with the power in the line component results in an F variance-ratio test (see [112], [225], [292]) with 2 and $2K - 2$ degrees of freedom for the significance of the estimated line component

$$F(f) = \frac{(K-1)|\hat{\mu}(f)|^2 \sum_{k=0}^{K-1} U_k(N, W; 0)^2}{\sum_{k=0}^{K-1} |y_k(f) - \hat{\mu}(f) U_k(N, W; 0)|^2}. \quad (13.10)$$

If the F test is significant, it is advantageous to "reshape" the spectrum around f_0 (which can be estimated from the location of the maximum F) to give a better estimate of the overall local spectrum

$$\begin{aligned} \bar{S}_r(f) &= |\hat{\mu}(f_0)|^2 \delta(f - f_0) \\ &+ \frac{1}{K} \sum_{k=0}^{K-1} |y_k(f) - \hat{\mu}(f_0) U_k(f - f_0)|^2. \end{aligned} \quad (13.11)$$

This operation must be done with care so that power is conserved numerically. In practice, the eigencoefficients are computed by using a FFT algorithm with the data "padded" with zeros. With standard FFT's, the frequencies form a discrete mesh with frequency increment $\Delta f = 1/M$, M being the length of the transformed arrays and so determined by the number of zeros appended. To avoid circular correlations M is always taken at least $2N$ and frequently much larger. Since spectrum estimates are used both for additional processing stages and are also plotted, we impose two constraints on the reshaping operation: first, that it be done in such a way that power is conserved; second, that when plotted the width of the peaks does not imply greater resolution than exists. Cramér-Rao bounds have been given by [276].

A. Estimates of the Mean

A particularly interesting example is that of estimating the process mean [18]. It is common experience that, when one estimates and subtracts the sample average before estimating the spectrum, the periodogram estimate of $S(0)$ is 0, and better estimates are typically biased low. Here one obtains a slightly better estimate for the mean while the estimated spectrum near the origin is much more accurate.

B. Continuation 2, Section VI Example

To exhibit these features we continue with the example introduced in Section VI. Fig. 22 shows a plot of the F variance ratio statistic (actually, since the plot is of $\log F$, we have effectively plotted Fisher's z statistic), in which the presence of the line components is clearly visible. The peak locations coincide with the true locations of the lines and the values of F are significant at levels >1 percent for both lines. Note also that the lines are sharp; at $F = 5$ the half-width of the lines is about 0.048, which is less than $1/2N$. The reshaped spectrum, Fig. 23, is identical to $\bar{S}(f)$ except that the estimated means have been subtracted from the eigencoefficients. Note that the background spectrum is reproduced accurately, again except for the band edges, and also that the "dip" commonly found at the origin after subtracting the mean is missing. Fig. 24 is a plot of the spectrum estimate obtained by applying the high-resolution composite estimate $\hat{S}_C(f)$ to the eigencoefficients, less their estimated mean value functions, in the same way.

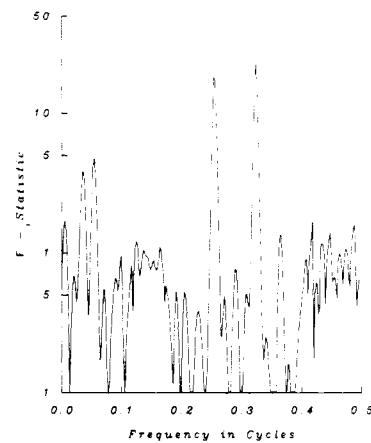


Fig. 22. The analysis of variance "F" test for the presence of harmonic line components using the data shown in Fig. 5. In this case the test has 2 and 8 degrees of freedom. The peaks near 0.05 cycle are thus insignificant while those at the line frequencies are significant above the 99.9-percent point.

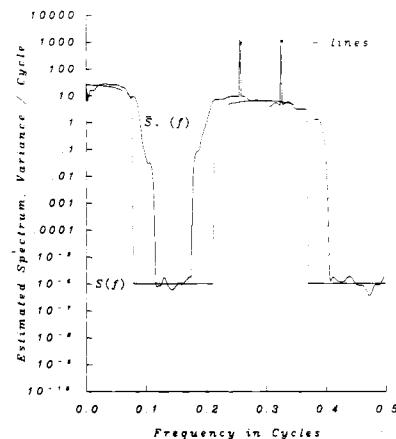


Fig. 23. The reshaped spectrum estimate. In this estimate the effects of the estimated line component amplitudes have been subtracted from the eigencoefficients before squaring and combining and the power replaced at the estimated line frequencies. This figure should be compared with Fig. 8.

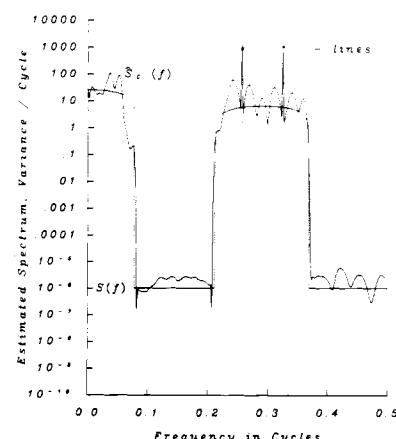


Fig. 24. A composite high-resolution estimate computed by the free parameter expansion of Section X from the reduced eigencoefficients. As in Fig. 23, the line power has been replaced.

Here the band edges are reproduced exceptionally well, but there is an obvious increase in variability in the rest of the spectrum. As mentioned in Sections IX and X, such behavior is to be expected in high-resolution estimates. However, even

with the extreme handicap of a relative discontinuity of $>10^6$, the estimated band edge is in error by less than 0.0038 cycle, about a factor of 3 better than $1/N$ and about a factor of 12 better than achievable with simple windowed estimates.

C. The Problem of Resolution

A controversial problem in time series analysis is that of resolution, that is of distinguishing closely spaced line components in a sample spectrum. We consider the case where there are two line components separated by a frequency Δ in a locally white-noise background of variance σ^2 and attempt to estimate the power in each component. The line spacing Δ is assumed to be less than W and, typically, less than the Rayleigh resolution $1/N$. If, in the presence of the background noise, we are able to estimate the two powers with enough accuracy to confidently assert their presence, we claim that the lines have been resolved. For a given estimation algorithm such assertions must be based *both* on the properties of the data and also on the *condition number* of the algorithm (see below).

The solution to the general (ordinary) least squares problem

$$\mathbf{Y} = \mathbf{X}\beta \quad (13.12)$$

where \mathbf{Y} is the n vector of observations, β the p vector of parameters to be estimated, and \mathbf{X} is the $n \times p$ matrix of coefficients, is well known and given by the *normal equations*,

$$\mathbf{X}^T \mathbf{X}\beta = \mathbf{X}^T \mathbf{Y} \quad (13.13)$$

conveniently expressed in terms of the *Moore-Penrose generalized inverse* or *pseudoinverse* of \mathbf{X}

$$\mathbf{X}^\dagger = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \quad (13.14)$$

as

$$\beta = \mathbf{X}^\dagger \mathbf{Y}. \quad (13.15)$$

The relative sensitivity of this solution to uncertainties in the observations may be expressed in terms of the *condition number* $\kappa^2(\mathbf{X})$ of the coefficient matrix and the relative norms of the uncertainties as

$$\frac{\|\delta\beta\|^2}{\|\beta\|^2} \leq \kappa^2(\mathbf{X}) \frac{\|\delta\mathbf{Y}\|^2}{\|\mathbf{Y}\|^2} \quad (13.16)$$

where the Euclidean norm $\|\mathbf{Y}\|^2 = \mathbf{Y}^T \mathbf{Y}$ has been used for vectors and the condition number defined using the spectral norm of the matrix

$$\kappa^2(\mathbf{X}) = \frac{\gamma_{\max}}{\gamma_{\min}} \quad (13.17)$$

where γ_{\min} and γ_{\max} are the smallest and largest eigenvalues of $\mathbf{X}^T \mathbf{X}$, respectively.

Using this notation and continuing with the harmonic analysis problem as expressed in terms of the eigencoefficients, we identify the following. First, the vector of observations \mathbf{Y} consists of the complex eigencoefficients at the two frequencies under consideration,

$$\mathbf{Y}^T = (y_0(f_0), y_1(f_0), \dots, y_K(f_0), y_0(f_1), y_1(f_1), \dots, y_K(f_1)). \quad (13.18)$$

Associated with \mathbf{Y} are the squared norms $\|\mathbf{Y}\|^2$ and $\|\delta\mathbf{Y}\|^2$. For the harmonic analysis problem the first, $\|\mathbf{Y}\|^2$, is the raw portion of the raw eigenspectrum $\bar{S}(f)$ from the line components. For reasonably strong lines it is approximately $\bar{S}(f)$. The second, $\|\delta\mathbf{Y}\|^2$, is the portion of the eigenspectrum near

f_0 and f_1 (which are supposed to be too close to be directly resolved in $\bar{S}(f)$) from the nondeterministic or background spectrum. $\|\mathbf{Y}\|^2/\|\delta\mathbf{Y}\|^2$ is interpretable as the input "signal-to-noise" power ratio.

Second, the coefficient matrix \mathbf{X} consists of discrete prolate spheroidal wave functions of arguments 0 and $\pm\Delta$

$$\mathbf{X} = \begin{bmatrix} U_0(0) & U_0(+\Delta) \\ U_1(0) & U_1(+\Delta) \\ \vdots & \vdots \\ U_K(0) & U_K(+\Delta) \\ U_0(-\Delta) & U_0(0) \\ U_1(-\Delta) & U_1(0) \\ \vdots & \vdots \\ U_K(-\Delta) & U_K(0) \end{bmatrix}. \quad (13.19)$$

Finally, the vector β consists of the estimated amplitudes of the two line components, $\beta^T = (\hat{\mu}(f_0), \hat{\mu}(f_1))$, with the total power given by the squared norm $\|\beta\|^2$. Also, as before, one can interpret $\|\beta\|^2/\|\delta\beta\|^2$ as the *output* signal-to-noise power ratio.

If one computes the condition number $\kappa^2(\mathbf{X})$ one finds that it is not very different from the condition number obtained for the problem of resolving two lines in a white-noise background

$$\kappa^2(\mathbf{C}) = \left[\frac{N \sin \pi\Delta + \sin N\pi\Delta}{N \sin \pi\Delta - \sin N\pi\Delta} \right]^{\pm 1} \quad (13.20)$$

which, for $N\Delta < 1$ is approximately $12/(N\pi\Delta)^2$. Thus for example, if one attempts to estimate the magnitudes of two lines at known frequencies separated by $0.01/N$, one must contend with $\kappa^2(\mathbf{X}) \approx 12\,000$, and so requires input signal-to-noise powers of at least 40 dB to even ensure that the estimated line magnitudes will equal the noise level. Similarly, at such a spacing, one cannot claim that an estimated line power apparently 40 dB above the noise represents anything except the intrinsic ill-conditioning of the problem. Clearly, when the line frequencies are unknown, the conditioning of the problem is unlikely to improve! It is also necessary to remember, particularly when working with miniprecision computers (32-bit or less word length), that the condition number of a *program*⁷ can be amazingly different from the condition number of the *algorithm* it was intended to implement.

The preceding paragraph applies equally to ordinary least squares and to the eigencoefficient approach when the background spectrum is strictly white. For this case, the two algorithms have approximately equal performance. However, since this case is almost never encountered in practice, it is necessary to compare the condition numbers of the two algorithms when the continuous component of the spectrum is colored.

When the spectrum is colored, it is necessary to replace ordinary least squares with generalized least squares, [7]. Written in the time domain, this is most easily done by regressing the coefficients of the Karhunen-Loeve data expansion against those of the line components as in [301]. This expansion requires the covariance matrix of the background process. Since this covariance matrix is unknown, complications multiply and

⁷The condition number of a process measures the relative sensitivity of results of the process to small relative changes in the input data. The input data for a program include machine precision, floating-point hardware implementation, library algorithms, the compiler, etc., in addition to the quantized time series samples.

one typically obtains an *implicit* estimate specified by relations between the data coefficients and eigenvalues of the "sample Karhunen-Loève expansion" corresponding to the spectrum estimate discussed in Section XII.

If the spectrum is *locally* white, and at a level not dominated by bias, the eigencoefficients are approximately uncorrelated, hence the setup is exactly the same as when the noise is white and the condition number of the normal equations the same. The difference is in the *input* noise power $\|\delta\beta\|^2$, which now is a function of the *local* background spectrum, the sample variance σ^2 , and the maximum coefficient order used

$$S(f) \leq \|\delta\beta\|^2 \leq S(f) + (1 - \lambda_K)\sigma^2. \quad (13.21)$$

Since, for a given value of NW , K is usually chosen so that bias is not dominant, the last term is typically small and the input noise power is given by the *local* spectrum.

Finally, we note that while the condition number just described implies uncomfortably large signal-to-noise ratios for high-resolution applications, they are modest compared to those described for the general reconstruction problem in Section X. This difference is a result of the prior knowledge implied here, i.e., that line components may exist with only their magnitude and phase unknown.

D. The Kay and Marple Example

To provide a basis for the comparison to these techniques to other methods in use we take the data set published in [185, table III] by Kay and Marple. Briefly, these data consist of 64 samples from a process composed of three sinusoids at frequencies of 0.1, 0.2, and 0.21 cycles, plus a continuum component with its maximum spectrum at 0.35 cycle. Of the 11 methods they compare (see [185, fig. 16]), none give completely satisfactory performance. All the methods detect the maximum near 0.2 cycle and four of them show the double line. Performance on the line at 0.1 cycle (20 dB weaker than the pair) is much poorer as the majority of the methods miss it completely, and its amplitude is reasonably estimated by only one of the methods presented. Similarly, performance on the continuum component is generally unsatisfactory and one must conclude that the example is difficult.

Using the methods presented in this paper embedded in the overall estimation structure mentioned in the Introduction (and described in detail in [324]) gives startlingly good results. Specifically, the estimation procedure consisted of the following steps:

- 1) Taking $NW = 3.5$ (a more appropriate value in view of the low dynamic range and expected complexity), eigencoefficients and weights were computed.
- 2) Using (13.10), an F statistic for line components was computed. This resulted in a very sharp peak at 0.1 cycle and a broad peak significant above the 99 percent point at 0.2 cycle. In addition, there is a lesser peak, significant at about the 95 percent level at 0.31 cycle. While a peak at this frequency is not mentioned in Kay and Marple, it is possibly an artifact of nonlinear interactions of the three known line components resulting from quantization. For a similar effect see the variable star example in [43]. The most likely cause, however, is that in a sample of size 64 one expects an upper extreme at roughly the 100 ($1 - 1/64$) percent point and, consequently, no action was taken at this frequency.
- 3) In the absence of an actual line component, a peak in the F statistic of the magnitude observed at 0.1 cycle occurs with probability $< 10^{-8}$ and consequently a line component with

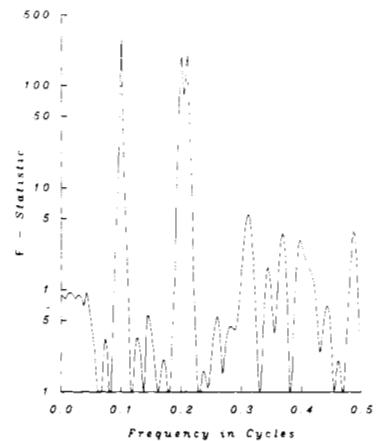


Fig. 25. The analysis of variance tests for the Kay and Marple example. At 0.1 cycle the result is that of a single frequency "F" test while near 0.2 cycle the result is dominated by the projection of a double line test. The peak near 0.3 cycle is of intermediate significance.

the estimated amplitude $0.0392 - 0.0271 i$ was removed from the data.

4) Because of the width and significance of the peak around 0.2 cycle (not to mention the information given in the reference!) a double line F statistic was computed for frequencies within the width of the peak. In this case the "integrated regression" procedure with coefficient weighting was used. This F statistic was plotted by projecting $\max_{f_2} F(f_1, f_2)$ onto the f_1 axis and is shown as Fig. 25. Note, first, that all three line components are resolved and also the exceptional sharpness of the line at 0.1 cycle. Here, for frequencies of 0.098, 0.100, and 0.102 cycle, values of F of 82.7, 292.8, and 57.3, respectively, are obtained so that frequency resolution much better than $\frac{1}{64} = 0.0156$ is obtained.

Similarly, at the double peaks at 0.20 and 0.21 cycle a value of $F = 195$ (on 4 and ≈ 12 degrees of freedom) is obtained. Again, the probability of such a value occurring without the actual line components being present is ridiculously low, and the values of F drop rapidly as the frequencies are changed from their correct values. In this case the estimated line amplitudes were $0.1458 - 0.4591 i$ and $0.1436 - 0.4651 i$ at 0.200 and 0.210 cycle, respectively. As before, their effects were subtracted. The combination of the three line components gives a good approximation to the data and even though the two strong components largely cancel each other at the upper end of the data set, the sample variance is reduced from 0.9253 to 0.1193 by this simple action.

5) Following removal of the estimated line components the spectrum of the residuals was computed. This was done using my preferred method, that is prewhitening using an autoregressive prediction error filter followed by estimation of the residual spectrum. In this case, an autoregression of order 5 was used resulting in a residual variance of 0.0073. Since the range of the spectrum is greatly reduced by this procedure the weights are all large and a mean stability of 9.77 degrees of freedom was obtained for the residual spectrum.

6) Correcting the residual spectrum for the prewhitening filter and replacing the line components at their estimated frequencies gives the spectrum estimate shown in Fig. 26.

In Fig. 26 the spectrum is plotted in units of "variance per cycle" and the amplitudes of the line are such that their numerical power content is correct with their widths corresponding to the estimated frequency resolution. As such, the scaling does not correspond exactly to the plots given in Kay and

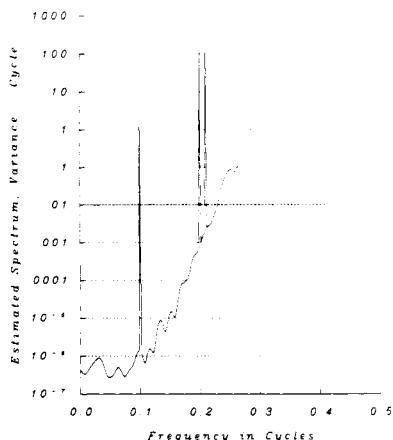


Fig. 26. The estimated spectrum for the Kay and Marple example using reshaping and free parameter expansions.

Marple. Also, in making comparisons between this estimate and those in Kay and Marple, observe that their figures only cover the top 40 dB of the spectrum and that the portions falling roughly below the heavier grid line are not shown in their figures. With regard to details, the estimated power of the lines in the pair are within 2 percent of each other. Also, as expected, the line at 0.1 cycle is 20 dB lower in power. Comparison with [185, fig. 16] of Kay and Marple shows that only the "Special Prony via Hildebrand approach" apparently does as well in this regard. This figure also explains the exceptional sharpness of the peak in the F statistic associated with this line, the local signal-to-noise ratio is better than 50 dB! Finally, the shape of the background spectrum agrees well with published information.

XIV. ESTIMATION OF COHERENCE AND POLYSPECTRA

So far we have been concerned with the simplest problem in the analysis of a stationary time series, namely estimating the ordinary spectrum from a finite realization of a single Gaussian process. In this section we expand the methods introduced so far to the next two problems: estimation of cross-spectra and coherences in multivariate series; and the estimation of higher order moment or cumulant spectra, or polyspectra (which are required to characterize non-Gaussian processes) and an extension to nonstationary processes. These problems are treated together as, typically, the estimators involve averages of products of Fourier transforms of one or more series, at the same or different frequencies.

Of the two, estimates of coherence and cross spectra have received the wider treatment [8], [78], [80], [125], [127], [128], [147], [174], [194], [195], [241], [281], [328], [329]. Polyspectra are treated in [46], [53]–[55], [59], [283], [340] with [41] being a good introduction; see also [123], [312]. Naive estimators of these functions (i.e., those corresponding to the periodogram or simple use of ensemble equations for estimators) are typically badly biased and depend strongly on data preprocessing and the relative group delay between the various components involved in the estimate. Excepting the case where there are large amounts of data available so that a variant of Welch's [347] method of overlapped segments can be used ([69], [70], [72]), several complex ad hoc frequency averages are used so that coherence estimates are usually regarded with extreme skepticism. Again, we claim that the eigencoefficient approach gives a much more satisfactory answer than the usual "frequency averaging" methods.

For this multivariate case we assume that the data consist of N contiguous samples taken concurrently on each of P series $x_m(t)$ for $m = 1, 2, \dots, P$ and $t = 0, 1, \dots, N - 1$. Components of the spectral density matrix, i.e., the spectra and cross-spectra between the different component series, are marked $S_{l,m}(f)$. We denote the k th eigencoefficient of the m th series by $y_{m,k}(f)$.

Paralleling the development in the univariate case, the eigenestimate of the cross-spectrum between series l and m is

$$\bar{S}_{l,m}(f) = \sum_{k=0}^{K-1} y_{l,k}(f) y_{m,k}^*(f) \quad (14.1)$$

where the superscript "*" denotes complex conjugate. Because the different eigencoefficients of each series are nearly uncorrelated from each other, again under the assumption that the local variations in the spectrum are not too extreme, the estimated squared coherence

$$\hat{C}_{l,m}^2(f) = \frac{|\bar{S}_{l,m}(f)|^2}{\bar{S}_{l,l}(f) \bar{S}_{m,m}(f)} \quad (14.2)$$

will have almost the same distribution as is usually assigned to squared coherence estimates for K independent replications of the pair of series. This distribution has been extensively studied in [25], [67]–[72], [125], [204], [216].

In addition, one may use composite methods, so that if enough data are available to form T nearly independent subsets, then one can form the cross-spectral estimates

$$\tilde{S}_{l,m}(f) = \frac{1}{T} \sum_{\tau=1}^T \sum_{k=0}^{K-1} y_{l,k}(f|\tau) y_{m,k}^*(f|\tau)$$

where $y_{l,k}(f|\tau)$ is the k th eigencoefficient from the τ th subset of the l th series.

$$\tilde{C}_{l,m}^2(f) = \frac{|\tilde{S}_{l,m}(f)|^2}{\tilde{S}_{l,l}(f) \tilde{S}_{m,m}(f)}.$$

If the segments are independent, the statistics of the coherence function are the same as those normally associated with KT overlapped segments. The variance of coherence estimates is typically worst for values of the true coherence of about 0.36 and may be approximated by

$$\frac{KT-1}{KT(KT+1)} \left[\frac{1}{KT} + 2 \frac{KT-2}{KT+2} C - \dots \right]$$

and taking $T = 8$ and $K = 6$ as typical values, the variance decreases from 0.054 which would be obtained using conventional approaches to 0.014. If, on the other hand, the segments are not independent but are considerably overlapped subsets of a single longer data set, then the improvement is not as dramatic. This is because overlapping windows of the $v_f^{(0)}(N, W)$ type exhibit reasonable variance efficiencies if the overlap is large enough [324, sec. 3.3].

When the spectrum of the processes contains lines in addition to background components, the interpretation of the usual coherence estimate becomes much more difficult, particularly if the power in the line component is commensurate with that in the local background spectrum. These difficulties arise simply because the two components may have, and often do have, independent physical origins. Thus cases commonly occur where both the background noise process and the line components are independently coherent, but, because their cross-spectra have different phases, their sum appears inco-

herent. The facility of the eigenspectrum approach to estimate the complex amplitudes of the line components, and to separate them from the background, gives a solution to this problem.

The frequency-dependent means are estimated for each series by the method used in the section on harmonic analysis

$$\hat{\mu}_m(f) = \frac{\sum_{k=0}^{K-1} U_k(0) y_{m,k}(f)}{\sum_{k=0}^{K-1} U_k^2(0)}$$

and tested for significance individually. If the component exists in *both* series with high probability, it is reasonable to claim a coherent mean at that frequency; see [218], [219]. Removing mean components found to be significant in either or both series leaves the dispersion components

$$y'_{m,k}(f) = y_{m,k}(f) - \sum_h \hat{\mu}_m(f_h) U_k(N, W; f - f_h)$$

from which cross-spectra and coherences may be computed as above. In frequency regions where the complex mean has been estimated and the mean-value function subtracted there are two fewer degrees of freedom than initially existed.

A. Coherence Estimates with Highly Colored Spectra

In addition to dependencies induced by undetected line components in the processes under study, coherence estimates will become biased in regions where the background spectra are changing rapidly. Using the same methods as in Section IV, one may show that the correlation between the eigencoefficients $y_{m,j}(f)$ and $y_{n,k}(f)$ is given by

$$\begin{aligned} C_{m,n}^{(j,k)}(f) &= \mathbb{E}\{y_{m,j}(f) \cdot y_{n,k}^*(f)\} \\ &= \int_{-1/2}^{1/2} U_j(\xi - f) U_k(\xi - f) S_{m,n}(f) df \end{aligned}$$

which, if $S_{m,n}(f)$ departs radically from a constant in the interval $(f - W, f + W)$, may be significant. Because the eigencoefficients associated with a given series may be highly correlated with each other, simple estimates such as (14.2), appropriate for spectra nearly white near f , may be misleading. Thus when there are enough independent replications, we consider the canonical correlations between the two sets of eigencoefficients. These are defined in terms of the roots of the general matrix eigenvalue problem

$$(C_{n,m} C_{m,m}^{-1} C_{m,n} - \gamma C_{n,n}) L = 0$$

where the largest of the eigenvalues, γ , gives the closest association between the two series; see [88], [161], [231], [354].

B. Examples of Coherence and Polyspectra Estimates

1) *Example 1:* We consider a bivariate case of the process studied in Section VI with emphasis on the region of the central "bump." In this region, the continuous part of the spectrum is characterized by the terms

$$x_{c,1}(t) = p_1(t) + p_2(t) \cos(\omega_c t) + p_3(t) \sin(\omega_c t)$$

and a second series containing similar terms

$$x_{c,2}(t) = p_4(t) - p_3(t) \cos(\omega_c t) + p_2(t) \sin(\omega_c t)$$

where $p_1(t)$ through $p_4(t)$ are independent stationary processes having identical autocovariance functions $R_p(\tau)$. With this

definition both $x_{c,1}(t)$ and $x_{c,2}(t)$ are stationary and have autocovariance functions

$$R_x(\tau) = R_p(\tau)(1 + \cos(\omega_c \tau)).$$

In addition, they are *jointly* stationary with lagged cross correlation

$$R_{12}(\tau) = \mathbb{E}\{x_{c,1}(t)x_{c,2}(t + \tau)\} = R_p(\tau) \sin(\omega_c \tau).$$

The common spectrum of $x_{c,1}(t)$ and $x_{c,2}(t)$ is simply given by the frequency translated sum

$$S_x(f) = S_p(f) + S_p(f - f_c) + S_p(f + f_c)$$

and the cross spectrum by

$$S_{12}(f) = \frac{i}{2} (S_p(f - f_c) - S_p(f + f_c))$$

and it follows that the coherence in the central "bump" is high.

We take $x_1(t)$ to be the series defined in Section VI, that is $x_1(t)$ is $x_{c,1}(t)$ plus the two line components, and also define an additional series $x_2(t)$ given by

$$x_2(t) = x_{c,2}(t) + 2.4 \cos(\omega_1 t) + 2.6 \sin(\omega_2 t).$$

Using the two series, $x_1(t)$ and $x_2(t)$, and computing the coherence directly using (14.2) without consideration of the line components gives the results shown in Fig. 27. Coherence is evident around 0.2 and 0.38 cycles, but between 0.22 and 0.36 the values obtained do not show the close relation between the two series. The low values obtained are a result of cancellations between the continuum and line components as they have different phase relationships.

In Fig. 28 the data have been reprocessed using both the estimated first and second moments, and the improvement is dramatic. In this figure, the solid curve shows the estimated coherence between the continuous parts of the process. Note that the estimated coherence is high where expected, and reasonably low in the regions between the bands where only uncorrelated noise exists. In addition, the apparently significant "spike" near the band edge at $f = 0.07$ is an example of spurious correlation. Recall that, in this example, the two low-frequency components are independent, the apparent coherence is a result of the correlation induced within each set of eigencoefficients by the rapidly changing spectra. The dashed curve is an estimate of the "first moment coherence" and is simply the product of the cumulative distribution functions for the two F tests considered independently. Again, the only highly significant values occur at the two line frequencies.

2) *Example 2:* Now consider a similar pair of series where the definition of the second series is changed to

$$x_3(t) = p_4(t) + p_3(t) \cos(\omega_c t) + p_2(t) \sin(\omega_c t)$$

the only difference between the $x_{c,2}$ and x_3 series being the sign of the $p_3(t)$ term. Consider the problem of estimating the dependencies between $x_{c,1}(t)$ and $x_3(t)$. Clearly both series contain frequency-translated copies of the basic $p_2(t)$ and $p_3(t)$ sequences and so are not independent; the difference between this example and the previous one is that in the $x_3(t)$ sequence the frequency components in the band about ω_c have been reversed. As such the coherence, as commonly defined, vanishes. This is shown in Fig. 29 where only the spurious correlations at the band edges and a residue at the carrier frequency have moderately significant values.

The problem in this example is that, while as before, $x_{c,1}$ and x_3 are *individually* stationary but are now *not jointly* stationary.

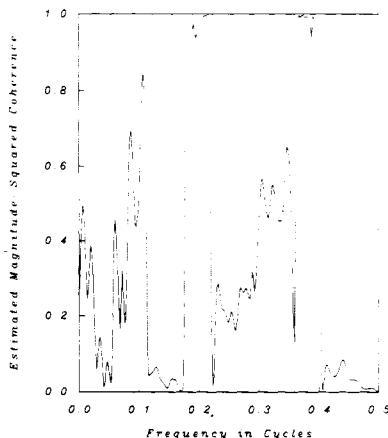


Fig. 27. The estimated magnitude-squared coherence for two processes of the type described in Section VI. and computed directly by (14.1).

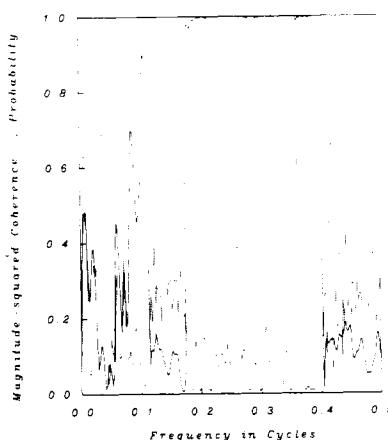


Fig. 28. An estimate of the magnitude-squared coherence applied to the same data as used in Fig. 27 but with estimation and removal of line components before use of (14.1). Here the estimated coherence between 0.2 and 0.4 cycle is close to the true value. The dashed line shows the probability of simultaneous line components. Note that here both line and continuum components are found to be coherent as opposed to the cancellation effects observed in Fig. 27. In both cases, the peak at 0.1 cycle is an artifact induced by the discontinuity in the spectrum.

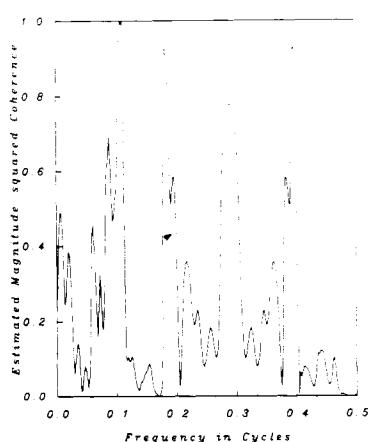


Fig. 29. A second example where simple coherence estimates fail. In this case the frequency components in the 0.2 to 0.4 cycle band of one series have been reversed and the result appears incoherent.

ary, having as lagged cross covariance

$$\{x_{c,1}(t)x_3(t+\tau)\} = R_p(\tau) \sin(\omega_c(2t+\tau)).$$

When analyzing data, however, one does not know the the-

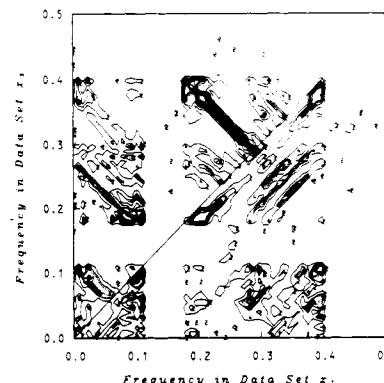


Fig. 30. A generalized coherence estimation procedure applied to the data used to generate Fig. 29. In this figure, the region above the main diagonal shows frequency-reversed coherences, the region below the diagonal shows coherences between frequency-translated coherences. Note that the strong coherence missed by the simple estimate is shown by the band with slope -1 in the upper triangle.

retical covariance structure and, therefore, we assume that short series of processes $x_{c,1}(t)$ and $x_3(t)$ are available and are to be analyzed. We also note: first, processes of the type used in these examples are common in engineering practice (see any recent issue of the IEEE TRANSACTIONS ON COMMUNICATIONS for examples), and also occur naturally because of doppler shifts, etc.; second, that the processes are only *jointly* nonstationary implies that the condition will probably not be noticed unless large amounts of data are available; third, the change in model specification is so apparently insignificant that any practical technique should be robust against it.

We thus propose a "second-order bispectrum" in which the normal rules governing the possible frequency combinations in polyspectra of stationary series are ignored. Thus to examine the structural relationship between the processes at two frequencies, f and f' , we propose an estimate of the forward cross spectrum

$$\tilde{C}_{n,m}(f, f') = \sum_{k=0}^{K-1} \tilde{y}_{m,k}(f) \tilde{y}_{n,k}^*(f')$$

and the reverse cross spectrum

$$\tilde{C}_{m,n}(f, f') = \sum_{k=0}^{K-1} \tilde{y}_{m,k}(f) \tilde{y}_{n,k}(f').$$

The "forward" estimate refers to the case where only a frequency translation has occurred while the "reverse" applies to the case in which a band of frequencies has been reversed as well as translated. The notation $\tilde{y}_{n,k}(f)$ refers to the k th eigencoefficient from the n th series normalized by

$$\tilde{y}_{n,k}(f) = \frac{d_{n,k}(f) y_{n,k}(f)}{\sum_{k=0}^{K-1} d_{n,k}^2(f)}$$

where, as earlier, the d_k 's are the solution weights.

By analogy with the stationary case we define the magnitude-squared coherence as

$$\tilde{C}(f, f') = \frac{|\tilde{C}_{m,n}(f, f')|^2 + |\tilde{C}_{n,m}(f, f')|^2}{\tilde{S}_n(f) \tilde{S}_m(f')}$$

which is the sum of the squared reverse coherence and the squared forward coherence and so may exceed 1.

Fig. 30 is a contour plot of the squared cross-coherence function showing the reverse term above the diagonal and the

forward term below the main diagonal. Observe that, in the square defining the band edges on the two frequencies, there is a well-defined and significant structure clearly showing the frequency reversal in the band.

XV. SUMMARY AND CONCLUSIONS

In this paper we have attempted to present a different approach to the problems of spectrum estimations from those currently fashionable. Thus instead of modeling the series in terms of rational polynomials, as is done in ARMA approaches, we have used the eigenfunctions of the finite Fourier transform to find approximate solutions of the integral equation defining the limits of our knowledge in the frequency domain. This approach has yielded considerable new information. First, the use of data windows is seen to be justifiable in that ones used previously approximate the first term in the series expansion used here. Second, stable estimates are obtained without the usual decrease in resolution. Third, a method for combining the basic high-resolution estimate in a free-parameter expansion was demonstrated. This estimate was then shown to produce eigenvalues reasonably close to the true values. In addition, using the sample eigenfunctions to expand the process gave high values of the log-likelihood function. Fourth, because the estimate can be used to separate moments of the effective spectral representation, it provides a distinction between spectrum and harmonic analysis and new techniques for dealing with mixed spectra. Some applications to multivariate processes were discussed.

Clearly, these techniques are not suited for all applications any more than is any other method. In problems where there is sufficient additional *a priori* information to permit specific techniques to be developed, a method optimized for a specific problem should always outperform a general technique using less *a priori* information about that problem. The extra computations required will probably exclude the techniques described here in those cases where the biggest problem is too much data and statistical efficiency is irrelevant. Similarly, the difference equations for the discrete prolate spheroidal sequences have not been exploited so that the technique cannot yet be applied in problems having severe real-time constraints. Also, the calculations are easier on a large computer. However, those with too little data, the range of the spectrum large, little information about the process, and statistical efficiency essential may find the additional computations worthwhile. In such cases, it is reasonable to assume that, as with most statistical techniques, it will be impossible to verify that the conditions assumed in this paper, i.e., stationarity, etc., are met. Indeed, since stationarity and line spectra are at best convenient fictions, these conditions will almost certainly *not* be satisfied by real data and one must always question the validity of the results; see [339]. My impression is that, if the data are not too far from Gaussian and stationary, the method works well. The reasons for this are as follows. First, normality has been assumed only for the variance expressions, so that these may be expected to break down as the data become less Gaussian in nature. Second, the orthogonal increment properties of $dZ(f)$ implied by strict stationarity are most important in the adaptive weighting of Section V. If one chooses simply to minimize the squared error for a particular sample, $dZ(f)$ may be replaced by an ordinary Fourier transform.

Again, we emphasize that the method discussed in this paper is *not* a complete spectrum estimation procedure but is in-

tended to fit into the larger framework of robust prewhitening, etc., mentioned in the Introduction.

Finally, the methods we have presented are still new, many of them appearing here for the first time. As such, I make no claim that they are optimal and many are, in fact, heuristic. They are presented in the hope of suggesting some new approaches to difficult problems.

APPENDIX COMPUTATION OF SPHEROIDAL WAVE FUNCTIONS

When N is large, it becomes inconvenient both from storage and precision viewpoints to directly solve the Toeplitz matrix equation. In such cases it is simpler to proceed as follows:

- 1) In [306, sec. 2.6], Slepian gives asymptotic expressions for both the discrete prolate spheroidal wave functions and sequences in terms of the continuous-time prolate spheroidal wave functions $\psi_k(c, x)$

$$v_n^{(k)}(N, W) \approx \pm \sqrt{\frac{2}{N\lambda_k(c)}} \psi_k \left(c, \frac{2n}{N} - 1 \right) \quad (A1)$$

where $c = N\pi W$ and the prolate spheroidal wave functions satisfy the integral equation

$$\lambda_k(c)\psi_k(c, x) = \int_{-1}^1 \frac{\sin c(x-u)}{\pi(x-u)} \psi_k(c, u) du.$$

- 2) Convert the integral equation to an algebraic eigenvalue equation by use of a Gauss-Legendre quadrature formula. (See [5], [318] for an introduction to these procedures.) To do this we assume a J point formula so that the integral equation becomes

$$\lambda_k(c)\psi_k(c, x) \approx \sum_{j=1}^J w_j \frac{\sin c(x-x_j)}{\pi(x-x_j)} \psi_k(c, x_j) \quad (A2)$$

where the x_j 's and w_j 's are the abscissas and weights of the quadrature rule. Note that there is a PORT routine [115], *dgqm11*, available to directly compute abscissas and quadrature points for a given J . If we define a modified eigenfunction

$$\Psi_k(j) = \sqrt{w_j} \psi_k(c, x_j)$$

and symmetric kernel

$$K(m, j) = \sqrt{w_m w_j} \frac{\sin c(x_m - x_j)}{\pi(x_m - x_j)}$$

and evaluate (A2) at the quadrature abscissas, the result is a symmetric algebraic eigenvalue problem

$$\lambda_k(c)\Psi_k(m) = \sum_{j=1}^J K(m, j)\Psi_k(j). \quad (A3)$$

As typically only a few eigenvalues are needed, the EISPACK procedures of [310, sec. 2.1.13] are very convenient.

- 3) The discrete prolate spheroidal sequences may now be interpolated at the N points required by using the kernel. Substituting in (A2) gives

$$v_n^{(k)}(N, W) \approx g_k \sum_{j=1}^J \sqrt{w_j} \frac{\sin c \left(\frac{2n}{N} - 1 - x_j \right)}{\pi \left(\frac{2n}{N} - 1 - x_j \right)} \Psi_k(j) \quad (A4)$$

where g_k is a normalization factor chosen so that

$$\sum_{n=0}^{N-1} |v_n^{(k)}(N, W)|^2 = 1.$$

This procedure works well because the error of a J -point Gaussian quadrature rule can be bounded in terms of the $2J$ th derivative of the integrand f

$$\epsilon_J \leq \frac{2^{2J+1}(J!)^4}{(2J+1)[(2J)!]^3} \max |f^{(2J)}(\xi)|$$

and in this problem the functions involved are all entire. Using Fourier integral representations for both the kernel and eigenfunctions, it may be shown that

$$\max |f^{(2J)}(\xi)| \leq \frac{16}{\pi} \frac{(\sqrt{2c})^{2J}}{(J+1)(2J+1)}.$$

Consequently, the error due to the Gaussian quadrature procedure decreases very rapidly with J .

In opposition to this rapidly decreasing quadrature error is the roundoff error in the algebraic eigenvalue procedure which increases with J but whose detailed characteristics depend both on the floating-point hardware used and the implementation of the routine. Taking s to be the number of bits in the floating point mantissa [44, ch. 6], gives

$$108 \times 2^{-s} J^{3/2} (1 + 9 \times 2^{-s})^{24J-42}$$

for a relative *rms* error bound on computing *all* the eigenvalues of a symmetric matrix using Jacobi's method. For small J the dominant error term is proportional to $J^{3/2}$ but, when J becomes large, the error increases exponentially. Note carefully, however, that this formula is given only as an example and that the accuracy of numerical routines depends on many factors besides basic machine precision so that this formula is not valid for other eigenvalue procedures and not even for other implementations of Jacobi's method. Also, when the matrices involved are stored in single precision, it is common to accumulate sums in double precision so that the effective precision is increased somewhat. Details are available in [352].

The total error, therefore, consists of two terms: the quadrature error, which decreases with J ; and the numerical error, which increases with J . Thus for a given value of c , there is an "optimum" J . Directly finding this minimum gives for $c = 4\pi$ and $\epsilon = 2^{-27}$, $J \approx 18$, and for $\epsilon = 2^{-111}$, $J \approx 35$. However, since we are usually interested in only a few of the largest eigenvalues, and the matrix is very structured, when using the EISPACK routines mentioned above, use of a somewhat larger J is advisable.

ACKNOWLEDGMENT

It is my pleasure to thank J. F. Kaiser, C. L. Mallows, A. J. Rainal, D. W. Tufts, and J. W. Tukey for their many helpful suggestions and comments.

REFERENCES

- [1] *Bell Syst. Tech. J.* (Special Issue on the WT4 Millimeter Wave-guide System), vol. 56, no. 9, 1977.
- [2] *Proc. RADC Spectrum Estimation Workshop*, 1978.
- [3] *Proc. 2nd RADC Spectrum Estimation Workshop*, 1979.
- [4] *Directions in Time Series*, D. R. Brillinger & G. C. Tiao, Eds., Institute of Mathematical Statistics, 1980.
- [5] M. Abramowitz, and I. A. Stegun, Eds. *Handbook of Mathematical Functions* (Applied Math Series 55). Washington, DC: NBS, 1965.
- [6] D. Achilles, "Deconvolution algorithms based on spline interpolation," in *Proc. ICASSP 80*, pp. 950-953, 1980.
- [7] A. C. Aitken, "On least squares and linear combination of observations," *Proc. Roy. Soc. Edinburgh*, vol. 55, pp. 42-48, 1935.
- [8] H. Akaike, "Some problems in the application of the cross-spectral method," in *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967, pp. 81-108.
- [9] —, "On the use of an index of bias in the estimation of power spectra," *Ann. Inst. Statist. Math.*, vol. 20, pp. 55-69, 1968.
- [10] V. G. Alekseev, "Calculation of spectra of stationary random processes on the basis of large samples," *Prob. Infor. Trans.*, vol. 16, pp. 30-35, 1980.
- [11] —, "On estimation of some functionals of spectral density of Gaussian random processes," *Theory Prob. & Appl.*, vol. XXV, pp. 267-273, 1980.
- [12] V. G. Alkseev and A. M. Yaglom, *Nonparametric and Parametric Spectrum Estimation Methods for Stationary Time Series* (Time Series), O. D. Anderson, Ed. Amsterdam, The Netherlands: North-Holland, 1980, pp. 401-422.
- [13] J. B. Allen and L. R. Rabiner, "Unbiased spectrum estimation and system identification using short-time spectral analysis methods," *Bell System Tech. J.*, vol. 58, pp. 1743-1763, 1979.
- [14] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1958.
- [15] —, "An asymptotic expansion of the distribution of the limited information maximum likelihood estimate of a coefficient in a simultaneous equation system," *J. Amer. Statist. Assoc.*, vol. 69, pp. 565-573, 1974.
- [16] T. W. Anderson, and H. Rubin, "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Ann. Math. Statist.*, vol. 20, pp. 46-63, 1949.
- [17] —, "The asymptotic properties of estimates of the parameters of the parameters of a single equation in a complete system of stochastic equations," *Ann. Math. Statist.*, vol. 21, pp. 570-582, 1950.
- [18] W. W. Anderson, "Optimum estimation of the mean of a Gaussian random process," *Proc. IEEE*, vol. 53, pp. 1640-1641, 1965.
- [19] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, pp. 337-404, 1950.
- [20] T. Aulin, "A modified model for the fading signal at a mobile radio channel," *IEEE Trans. Vehicular Technology*, vol. VT-28, pp. 182-203, 1979.
- [21] H. Babić, G. C. Temes, "Optimum low-order windows for discrete Fourier transform systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 512-517, 1976.
- [22] C.T.H. Baker, et al. "Numerical solution of Fredholm integral equations of the first kind," *Comput. J.*, vol. 7, pp. 141-148, 1964.
- [23] C.T.H. Baker, *The Numerical Treatment of Integral Equations*. New York: Oxford Univ. Press, 1977.
- [24] E. J. Baghdady, R. N. Lincoln, and B. D. Nelin, "Short-term frequency stability: Characterization, theory, and measurements," *Proc. IEEE*, vol. 53, pp. 704-722, 1965.
- [25] T. E. Barnard, "Legendre polynomial expressions for the probability density function of magnitude-squared coherence estimates," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 107-108, 1986.
- [26] C. W. Barnes, "Object restoration in a diffraction-limited imaging system," *J. Opt. Soc. Amer.*, vol. 56, pp. 575-578, 1966.
- [27] Y. Bar-Shalom, "On the asymptotic properties of the maximum likelihood estimate obtained from dependent observations," *J. Roy. Statist. Soc.*, vol. B-33, pp. 72-77, 1971.
- [28] M. S. Bartlett, "On the theoretical specification and sampling properties of autocorrelated time series," *J. Roy. Statist. Soc. Suppl.*, vol. 8, pt. 1, pp. 27-41, 1946.
- [29] —, "Periodogram analysis and continuous spectra," *Biometrika*, vol. 37, pp. 1-16, 1950.
- [30] —, "Some remarks on the analysis of time series," *Biometrika*, vol. 54, pp. 25-38, 1967.
- [31] M. Båth, *Spectral Analysis in Geophysics*. Amsterdam, The Netherlands: Elsevier, 1974.
- [32] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [33] K. N. Berk, "Consistent autoregressive spectral estimates," *Ann. Statist.*, vol. 2, pp. 489-502, 1974.
- [34] M. Bertero, C. De Mol, and G. A. Viano, "On the problems of object restoration and image extrapolating in optics," *J. Math. Phys.*, vol. 20, pp. 509-521, 1979.
- [35] —, "Resolution beyond the diffraction limit for regularized object restoration," *Optica Acta*, vol. 27, pp. 307-320, 1980.
- [36] M. Bertran-Salvans and V. Casares-Giner, "On band-pass energy concentration for finite duration sequences," in *Signal Processing: Theories and Applications*, M. Kunt and F. de Coulon, Eds. Amsterdam, The Netherlands: North-Holland, 1980, pp. 181-186.

- [37] R. J. Bhansali, "A Monte Carlo comparison of the regression method and the spectral methods of prediction," *J. Amer. Statist. Assoc.*, vol. 68, pp. 621-625, 1973.
- [38] R. J. Bhansali, "Asymptotic properties of the Wiener-Kolmogorov predictor," *J. Roy. Statist. Soc.*, pp. 61-73, 1974.
- [39] F. W. Biegler-König, "Sufficient conditions for the solubility of inverse eigenvalue problems," *Linear Algebra Appl.*, vol. 40, pp. 89-100, 1981.
- [40] C. Bingham, M. D. Godfrey, and J. W. Tukey, "Modern techniques of power spectrum estimation," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 56-67, 1967.
- [41] R. B. Blackman, and J. W. Tukey, "The measurement of power spectra," *Bell Syst. Tech. J.*, vol. 37 (reprinted by Dover).
- [42] A. Blomqvist, "Figures of merit of windows for power spectrum estimation with the DFT," *Proc. IEEE*, vol. 67, pp. 438-439, 1979.
- [43] P. Bloomfield, *Fourier Analysis of Time Series: An Introduction*. New York: Wiley, 1976.
- [44] E. K. Blum, *Numerical Analysis and Computation: Theory and Practice*. Reading, MA: Addison-Wesley, 1972.
- [45] L. E. Blumenson, and K. S. Miller, "Properties of generalized Rayleigh distributions," *Ann. Math. Statist.*, vol. 34, pp. 903-910, 1963.
- [46] R. P. Bogert, M. J. Healy, and J. W. Tukey, "The frequency analysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking," in *Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, 1962, pp. 209-243.
- [47] E. Boileau and B. Picinbono, "Statistical study of phase fluctuations and oscillator stability," *IEEE Trans. Instrum. Meas.*, vol. IM-25, pp. 66-75, 1976.
- [48] B. A. Bolt and D. R. Brillinger, "Estimation of uncertainties in eigenspectral estimates from decaying geophysical time series," *Geophys. J. Roy. Astr. Soc.*, vol. 59, pp. 593-603, 1979.
- [49] C. de Boor, "Package for calculating with b-splines," *SIAM J. Numer. Analysis*, vol. 14, pp. 441-472, 1977.
- [50] G.E.P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, pp. 317-346, 1949.
- [51] G.E.P. Box and G. M. Jenkins, *Time Series Analysis Forecasting and Control*. San Francisco: Holden-Day, 1970.
- [52] D. R. Brillinger, "An introduction to polyspectra," *Ann. Math. Statist.*, vol. 36, pp. 1351-1374, 1965.
- [53] —, "Computation and interpretation of k-th order spectra," *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967, pp. 81-108.
- [54] —, "Asymptotic properties of spectral estimates of second order," *Biometrika*, vol. 56, pp. 375-390, 1969.
- [55] —, "The frequency analysis of relations between stationary spatial series," *Proc. 12th Biennial Sem. Canadian Math. Congress*, R. Pyke, Ed. (Canadian Math. Congress, Montreal) pp. 39-81, 1970.
- [56] —, "Fourier analysis of stationary processes," *Proc. IEEE*, vol. 62, pp. 1628-1643, 1974.
- [57] —, *Time Series, Data Analysis and Theory*. New York: Holt, Rinehart and Winston, 1975.
- [58] —, "The key role of tapering in spectrum estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 1075-1076, 1981.
- [59] D. R. Brillinger and M. Rosenblatt, "Asymptotic theory of estimates of k-th order spectra," in *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967, pp. 81-108.
- [60] W. S. Brown, *ALTRAN User's Manual*. Murray Hill, NJ: Bell Laboratories (Available through AT&T patent licensing, Greensboro, NC), 1977.
- [61] G. J. Buck, "Radar mapping: Prolate spheroidal wave functions versus truncated inverse Fourier transform," *IEEE Trans. Antennas Propagat.*, vol. AP-20, pp. 188-193, 1972.
- [62] J. P. Burg, "The relationship between maximum entropy spectra and maximum likelihood spectra," *Geophysics*, vol. 37, pp. 375-376, 1972.
- [63] J. P. Butler, J. A. Reeds, and S. V. Dawson, "Estimating solutions of first kind integral equations with nonnegative constraints and optimal smoothing," *SIAM J. Numer. Anal.*, vol. 18, pp. 381-397, 1981.
- [64] J. A. Cadzow, "An extrapolation procedure for band-limited signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 4-12, 1979.
- [65] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408-1418, 1969.
- [66] —, "Maximum-likelihood spectral estimation," in *Nonlinear Methods of Spectral Analysis*, S. Haykin, Ed. New York: Springer-Verlag, 1979, pp. 155-180.
- [67] G. C. Carter, "Statistics of the estimate of Coherence," *Proc. IEEE*, vol. 60, pp. 465-466, 1972.
- [68] —, "Bias in magnitude-squared coherence estimation due to misalignment," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 97-99, 1980.
- [69] G. C. Carter, C. H. Knapp, and A. H. Nuttall, "Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 337-344, 1973.
- [70] —, "Statistics of the estimate of the magnitude-coherence function," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 388-389, 1973.
- [71] G. C. Carter and A. H. Nuttall, "Evaluation of the statistics of the estimate of magnitude-squared coherence," NUSC Tech. Memo. TC-193-71, 1971.
- [72] —, "On the weighted overlapped segment averaging method for power spectrum estimation," *Proc. IEEE*, vol. 68, pp. 1352-1354, 1980.
- [73] W. Y. Chen and G. R. Stegen, "Experiments with maximum entropy power spectra of sinusoids," *J. Geophys. Res.*, vol. 79, pp. 3019-3022, 1974.
- [74] D. G. Childers, *Modern Spectrum Analysis*. New York: IEEE Press, 1978.
- [75] D. G. Childers, J. I. Aunon, and C. D. McGillem, "Spectral analysis: Prediction and extrapolation," in *Critical Reviews in Bioengineering*, vol. 6, issue 2. Boca Raton, FL: CRC Press, 1981, pp. 133-175.
- [76] D. G. Childers and M.-T. Pao, "Complex demodulation for transient wavelet detection and extraction," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 295-308, 1972.
- [77] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell Syst. Tech. J.*, vol. 47, pp. 957-1000, 1968.
- [78] W. S. Cleveland and E. Parzen, "The estimation of coherence, frequency response, and envelope delay," *Technometrics*, vol. 17, pp. 167-172, 1975.
- [79] J. W. Cooley, and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comp.*, vol. 19, pp. 297-301, 1965.
- [80] J. W. Cooley, P.A.W. Lewis, and P. D. Welch, *The Application of the Fast Fourier Transform Algorithm to the Estimation of Spectra and Cross-Spectra*, Computer Processing in Communications, Polytechnic Institute of Brooklyn Microwave Research Institute Symp. Ser., vol. 19, pp. 5-20, 1969.
- [81] D. R. Cox and D. V. Hinkley, *Theoretical Statistics*. London, England: Chapman & Hall, 1979.
- [82] H. Cramer, "On the theory of stationary random processes," *Ann. Math.*, vol. 41, pp. 215-230, 1940.
- [83] —, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946.
- [84] R. E. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 99-102, 1980.
- [85] H. B. Curry and I. J. Schoenberg, "On Pólya frequency functions IV: The fundamental spline functions and their limits," *J. Analyse Math.*, vol. 17, pp. 71-107, 1966.
- [86] G. Dahlquist, A. Björck, and N. Anderson, *Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
- [87] H. E. Daniels, "The estimation of spectral densities," *J. Roy. Statist. Soc.*, vol. B-24, pp. 185-198, 1962.
- [88] J. N. Darroch, "An optimal property of principal components," *Ann. Math. Statist.*, vol. 36, pp. 1579-1582, 1965.
- [89] A. R. Davies, T. Cochrane, and O. M. Al-Faour, "The numerical inversion of truncated autocorrelation functions," *Opt. Acta*, vol. 27, pp. 107-118, 1980.
- [90] H. T. Davis and R. H. Jones, "Estimation of the innovations variance of a stationary time series," *J. Amer. Statist. Assoc.*, vol. 63, pp. 141-149, 1968.
- [91] Yu. M. Davydov, "Moments of the bivariate envelope distribution," *Radio Eng.*, vol. 25, no. 12, pp. 77-81, 1970.
- [92] L. M. Delves and J. Walsh, Eds., *Numerical Solution of Integral Equations*. New York: Oxford Univ. Press, 1974.
- [93] A. M. Despain, and J. W. Bell, "Increased spectral resolution from fixed length interferograms," *Aspen Int. Conf. on Fourier Spectroscopy*, 1970, G. A. Vanasse, A. T. Stair, and D. J. Baker, Eds., Tech. Rep. AFCRL-71-0019, pp. 397-406, 1970.
- [94] E. L. Dodd, "The length of the cycles which result from the graduation of chance elements," *Ann. Math. Statist.*, vol. 10, pp. 254-264, 1939.
- [95] J. L. Doob, *Stochastic Processes*. New York: Wiley, 1953.
- [96] V. Dose, and H. Scheidt, "Deconvolution of appearance potential spectra," *Appl. Phys.*, vol. 19, pp. 19-23, 1979.
- [97] V. Dose, and Th. Fauster, "Deconvolution of appearance potential spectra II," *Appl. Phys.*, vol. 20, pp. 299-303, 1979.
- [98] S. C. Dutta Roy and A. Agrawal, "Digital low-pass filtering using the discrete Hilbert transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 465-467, 1978.
- [99] H. Dym and H. P. McKean, *Gaussian Processes, Function Theory, and the Inverse Spectral Problem*. New York: Academic Press, 1976.
- [100] K. O. Dzhaparidze, "Estimation of parameters of a spectral density with fixed zeroes," *Theory Prob. and Appl.*, vol. XXII, pp. 708-729, 1977.
- [101] K. O. Dzhaparidze and A. M. Yaglom, "Asymtotically efficient

- estimation of the spectrum parameters of stationary stochastic processes," in *Proc. Prague Symp. on Asymptotic Statistics*, vol. 1. Prague, Czechoslovakia: Charles Univ. Press, 1974.
- [102] —, "Applications of a modified "scoring method" of Fisher to the estimation of spectral parameters of random processes," *Sov. Math.-Dokl.*, vol. 15, pp. 1077-1082, 1974.
- [103] A. Eberhard, "An optimal discrete window for the calculation of power spectra," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 37-43, 1973.
- [104] B. Efron, "Computers and the theory of statistics: Thinking the unthinkable," *SIAM Rev.*, vol. 21, pp. 460-480, 1979.
- [105] M. F. Egerton and P. J. Laycock, "Some criticisms of stochastic shrinkage and ridge regression, with counterexamples," *Technometrics*, vol. 23, pp. 155-159, 1981.
- [106] M. P. Ekstrom, "A spectral characterization of ill-conditioning in numerical deconvolution," *IEEE Trans. Audio Electroacoust.*, vol. Au-21, pp. 344-348, 1973.
- [107] R. H. Farrell, "Asymptotic lower bounds for the risk of estimators of the value of a spectral density function," *Z. Wahrscheinlichkeitstheorie Verw.*, vol. 49, pp. 221-234, 1979.
- [108] J. L. Fields, "Rational approximations to generalized hypergeometric functions," *Math. Comp.*, vol. 19, pp. 606-624, 1965.
- [109] R. A. Fisher, "Tests of significance in harmonic analysis," *Proc. Roy. Soc. London*, vol. 125 A, pp. 54-59, 1929.
- [110] —, "On the similarity of the distributions found for the test of significance in harmonic analysis and in Steven's problem in geometrical probability," *Ann. Eugen.*, vol. 10, pp. 14-17, 1940.
- [111] —, *Statistical Methods and Scientific Inference*, 3rd. ed. New York: Hafner, 1973.
- [112] —, *Statistical Methods for Research Workers*, 14th ed. New York: Hafner, 1973.
- [113] R. Fitzgerald, "Extrapolation of band-limited signals: A tutorial," *Signal Processing: Theories and Applications*, M. Kunt and F. de Coulon, Eds. Amsterdam, The Netherlands: North-Holland, 1980, pp. 175-180.
- [114] C. Flammer, *Spheroidal Wave Functions*. Stanford, CA: Stanford Univ. Press, 1967.
- [115] P. Fox, Ed., *PORT Mathematical Subroutine Library*. Murray Hill, NJ: Bell Telephone Laboratories, 1977 (available through AT&T patent licensing, Greensboro, NC).
- [116] J. N. Franklin, "On Tikhonov's method for ill-posed problems," *Math. Comp.*, vol. 28, pp. 889-907, 1974.
- [117] B. R. Frieden, "Band-unlimited reconstruction of optical objects and spectra," *J. Opt. Soc. Amer.*, vol. 57, pp. 1013-1019, 1967.
- [118] M. J. Gans, "A power-spectral theory of propagation in the mobile radio environment," *IEEE Trans. Vehicular Technology*, vol. VT-20, p. 27, 1972.
- [119] B. S. Garbow, J. M. Boyle, J. J. Dongarra, and C. B. Moler, "Matrix eigensystem routines—EISPACK guide extension," Lecture Notes in Computer Science no. 51. New York: Springer, 1977.
- [120] N. C. Geçkinli, "Power spectra estimation using discrete Fourier transform," in *Signal Processing: Theories and Applications*, M. Kunt and F. de Coulon, Eds. Amsterdam, The Netherlands: North-Holland, 1980, 567-575.
- [121] N. C. Geçkinli and D. Yavuz, "Some novel windows and a concise tutorial comparison of window families," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 501-507, 1978.
- [122] R. W. Gerchberg, "Superresolution through error energy reduction," *Opt. Acta*, vol. 21, pp. 709-720, 1974.
- [123] J. Geweke, "A comparison of tests of the independence of two covariance-stationary time series," *J. Amer. Statist. Assoc.*, vol. 76, pp. 363-373, 1981.
- [124] I. C. Gohberg and I. Fel'dman, *Convolution Equations and Projection Methods for their Solution*. New York: Amer. Math. Soc., 1974.
- [125] N. R. Goodman, "On the joint estimation of the spectra, cospectrum of a two-dimensional stationary Gaussian process," Scientific Paper No. 10, College of Eng., New York Univ., ASTIA Doc. AD-134919.
- [126] —, "Measuring amplitude and phase," *J. Franklin Inst.*, vol. 270, pp. 437-450, 1960.
- [127] —, "Spectral analysis of multiple time series," *Proc. Symp. on Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, 1962, pp. 260-266.
- [128] —, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, pp. 152-177, 1963.
- [129] F. Gori, and C. Palma, "On the eigenvalues of sinc² Kernel," *J. Phys. A: Math. Gen.*, vol. 8, pp. 1709-1719, 1975.
- [130] A. H. Gray, Jr. and J. D. Markel, "Distance measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-24, pp. 380-391, 1976.
- [131] R. M. Gray, "Sliding-block source coding," *IEEE Trans. Informat. Theory*, vol. IT-21, pp. 357-368, 1975.
- [132] R. M. Gray, D. L. Neuhoff, and P. C. Shields, "A generalization of Ornstein's \bar{d} distance with applications to information theory," *Ann. Prob.*, vol. 3, pp. 315-328, 1975.
- [133] R. M. Gray, A. Buzo, A. H. Gray, Jr., and Y. Matsuyama, "Distortion measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 367-376, 1980.
- [134] R. M. Gray, A. H. Gray, Jr., G. Rebolledo, and J. E. Shore, "Rate-distortion speech coding with a minimum discrimination information distortion measure," *IEEE Trans. Informat. Theory*, vol. IT-27, pp. 708-721, 1981.
- [135] U. Grenander, H. O. Pollak, and D. Slepian, "The distribution of quadratic forms in normal variates: A small sample theory with applications to spectral analysis," *J. SIAM* vol. 7, pp. 374-401, 1959.
- [136] U. Grenander and M. Rosenblatt, "An extension of a theorem of G. Szegö and its application to the study of stochastic processes," *Trans. Amer. Math. Soc.*, vol. 76, pp. 112-126, 1954.
- [137] —, *Statistical Analysis of Stationary Time Series*. New York: Wiley, 1957.
- [138] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. Los Angeles, CA: Univ. Cal. Press, 1958.
- [139] C. Gumacos, "On the optimum estimation of the spectra of certain discrete stochastic processes," *IEEE Trans. Informat. Theory*, vol. IT-13, pp. 298-304, 1967.
- [140] P. R. Gutowski, E. A. Robinson, and S. Trietel, "Spectral estimation: Fact or fiction," *IEEE Trans. Geosci. Electron.*, vol. GE-16, pp. 80-84, 1978.
- [141] J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. New Haven, CT: Yale Univ. Press, 1923.
- [142] P. H. Halpern, "Optimum finite duration Nyquist signals," *IEEE Trans. Commun.*, vol. COM-27, pp. 884-888, 1979.
- [143] E. J. Hannan, *Multiple Time Series*. New York: Wiley, 1970.
- [144] —, "Non-linear time series regression," *J. Appl. Prob.*, vol. 8, pp. 767-780, 1971.
- [145] —, "The estimation of frequency," *J. Appl. Prob.*, vol. 10, pp. 510-519, 1973.
- [146] E. J. Hannan and P. J. Thomson, "Spectral inference over narrow bands," *J. Appl. Prob.*, vol. 8, pp. 157-169, 1971.
- [147] —, "The estimation of coherence and group delay," *Biometrika*, vol. 58, pp. 469-481, 1971.
- [148] —, "Estimating group delay," *Biometrika*, vol. 60, pp. 241-253, 1973.
- [149] R. J. Hanson, "A numerical method for solving Fredholm integral equations of the first kind using singular values," *SIAM J. Numer. Anal.*, vol. 8, pp. 616-622, 1971.
- [150] R. F. Harrington, *Field Computation by Moment Methods*. New York: MacMillan, 1968.
- [151] B. Harris, Ed., *Spectral Analysis of Time Series*. New York: Wiley, 1967.
- [152] F. J. Harris, "High-resolution spectral analysis with arbitrary spectral centers and arbitrary spectral resolutions," *Comput. Electr. Eng.*, vol. 3, pp. 171-191, 1976.
- [153] —, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, pp. 51-83, 1978.
- [154] J. L. Harris, "Diffraction and resolving power," *J. Opt. Soc. Amer.*, vol. 54, pp. 931-936, 1964.
- [155] S. Haykin, *Nonlinear Methods of Spectral Analysis*. New York: Springer, 1979.
- [156] J. R. Higgins, *Completeness and Basis Properties of Sets of Special Functions*. New York: Cambridge Univ. Press, 1977.
- [157] I. I. Hirschman, Jr., "Recent developments in the theory of finite Toeplitz operators," *Adv. Prob. and Related Topics*, vol. 1, pp. 103-167, 1971.
- [158] —, "On the eigenvalues of certain integral operators," *SIAM J. Math. Anal.*, vol. 6, pp. 1024-1050, 1975.
- [159] I. I. Hirschman and D. V. Widder, *The Convolution Transform*. Princeton, NJ: Princeton Univ. Press, 1955.
- [160] W. S. Hodgkiss and J. A. Presley, "Adaptive tracking of multiple sinusoids whose power levels are widely separated," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 550-561, 1981.
- [161] H. Hotelling, "Relations between two sets of variables," *Biometrika*, vol. 28, pp. 321-377, 1936.
- [162] P. J. Huber, "Robust statistics: a review," *Ann. Math. Statist.*, vol. 43, pp. 1041-1067, 1972.
- [163] I. A. Ibragimov, "On estimation of the spectral function of a stationary Gaussian process," *Theory Prob. Appl.*, vol. VIII, pp. 366-401, 1963.
- [164] I. S. Ibrahimilov, "An estimator of the spectral density of a stationary Gaussian process," *Theory Prob. Math. Statist.*, vol. 17, pp. 87-94, 1979.
- [165] E. Isaacson and H. B. Keller, *Analysis of Numerical Methods*. New York: Wiley, 1966.
- [166] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 67-72, 1975.
- [167] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electr. Commun. Japan*, vol. 53-A, no. 1, pp. 36-43, 1970.
- [168] G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Application*.

- cations. San Francisco, CA: Holden-Day, 1968.
- [169] F. John, "Continuous dependence on data for solutions of partial differential equations with a prescribed bound," *Comm. Pure Appl. Math.*, vol. XIII, pp. 551-585, 1960.
- [170] R. Johnson and J. Shore, "Minimum-cross-entropy spectral analysis of multiple signals," Naval Research Lab. Memo. 4492, Apr. 7, 1981.
- [171] A. F. Jones, and D. L. Misell, "The problem of error in deconvolution," *J. Phys. A: Gen. Phys.*, vol. 3, pp. 462-472, 1970.
- [172] R. H. Jones, "Spectral estimates and their distributions," *Skandinavisk Aktuarietidskrift*, vol. 45, pp. 39-69, 135-153, 1962.
- [173] —, "A reappraisal of the periodogram in spectral analysis," *Technometrics*, vol. 7, pp. 531-542, 1965.
- [174] —, "Phase free estimation of coherence, *Ann. Math. Statist.*, vol. 40, pp. 510-518, 1969.
- [175] —, "Estimation of the innovation generalized variance of a multivariate stationary time series," *J. Amer. Statist. Assoc.*, vol. 71, pp. 386-388, 1976.
- [176] T. T. Kadota, "Simultaneous orthogonal expansion of two stationary Gaussian processes-examples," *Bell Syst. Tech. J.*, vol. 46, pp. 1071-1096, 1966.
- [177] T. T. Kadota and L. A. Shepp, "On the best finite set of linear observables for discriminating two Gaussian signals," *IEEE Trans. Informat. Theory*, vol. IT-13, pp. 278-284, 1967.
- [178] T. Kailath, "An innovations approach to least-squares estimation—Part I: Linear filtering in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 645-655, 1968.
- [179] —, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun.*, vol. COM-15, pp. 52-60, 1967.
- [180] —, "A view of three decades of linear filtering theory," *IEEE Trans. Informat. Theory*, vol. IT-20, pp. 146-180, 1974..
- [181] T. Kailath and P. Frost, "An innovations approach to least-squares estimation—Part II: Linear smoothing in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 655-660, 1968.
- [182] J. F. Kaiser, "Nonrecursive digital filter design using the $I_0 - \sinh$ window function," in *IEEE Inter. Symp. Circuits Syst. Proc.*, pp. 20-23.
- [183] T. Kaneko and P. Liu, "Accumulation of round-off error in fast Fourier transforms," *J. Assoc. Comput. Mach.*, vol. 17, pp. 637-654, 1970.
- [184] J. K. Kauppinen, D. J. Moffatt, D. G. Cameron, and H. H. Mantsch, "Noise in Fourier self-deconvolution," *Appl. Opt.*, vol. 20, pp. 1866-1879, 1981.
- [185] S. M. Kay, and S. L. Marple, Jr., "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, pp. 1380-1419, 1981.
- [186] R. B. Kelman and R. P. Feinerman, "Dual orthogonal series," *SIAM J. Math. Anal.*, vol. 5, pp. 489-502, 1974.
- [187] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics, Vol. I-III*. New York: Hafner, 1963.
- [188] C. G. Khatri, "Classical statistical analysis based on a certain multivariate complex Gaussian distribution," *Ann. Math. Statist.*, vol. 36, pp. 98-114, 1965.
- [189] J. L. Kinsey and R. D. Levine, "A performance criterion for information theoretic data analysis," *Chem. Phys. Lett.*, vol. 65, pp. 413-416, 1979.
- [190] B. Kleiner, R. D. Martin, and D. J. Thomson, "Three approaches towards making power spectra less vulnerable to outliers," *Buss. Econ. Statist. Sect. Proc., Amer. Statist. Assoc.*, pp. 386-391, 1976.
- [191] —, "Robust estimates of spectra (with discussion)," *J. Roy. Statist. Soc.*, vol. B-41, pp. 313-351, 1979.
- [192] G. Koleyni and F. Ingels, "Windowed sinusoidal waveforms analysis using the short time averaged cepstrum," in *Proc. Southeast-con.*, pp. 379-383, 1981.
- [193] A. N. Kolmogorov, "Stationary sequences in Hilbert space," *Bull. Math. Moscow Univ.*, vol. 2, no. 6, pp. 1-40 (translation), 1941.
- [194] L. H. Koopmans, "On the coefficient of coherence for weakly stationary stochastic processes" *Ann. Math. Statist.*, vol. 35, pp. 532-549, 1964.
- [195] L. H. Koopmans, "On the multivariate analysis of weakly stationary stochastic processes," *Ann. Math. Statist.*, vol. 35, pp. 1765-1780, 1964.
- [196] —, *The Spectral Analysis of Time Series*. New York: Academic Press, 1974.
- [197] S. Kullbach, *Information Theory and Statistics*. New York: Wiley, 1959.
- [198] R. W. Kulp, "An optimal sampling procedure for use with the Prony method," *IEEE Trans. Electromag. Comput.*, vol. EMC-23, pp. 67-71, 1981.
- [199] R. Kumaresan and D. W. Tufts, "Improved spectral resolution III: Efficient realization," *Proc. IEEE*, vol. 68, pp. 1354-1355, 1980.
- [200] —, "Singular value decomposition and spectral analysis," *Proc. IEEE Workshop on Spectral Estimation* (Hamilton, Ont.), pp. 6.4.1-6.4.12, 1981.
- [201] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—I," *Bell Syst. Tech. J.*, vol. 40, pp. 65-84, 1961.
- [202] —, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—III," *Bell Syst. Tech. J.*, vol. 40, pp. 1295-1336, 1962.
- [203] R. D. Larsen, "Time series analysis in Fourier spectroscopy," in *Proc. 1970 Aspen Inter. Conf. on Fourier Spectroscopy*, G. A. Vanasse, A. T. Stair, and D. J. Baker, Eds., pp. 385-394, 1970.
- [204] P. F. Lee, "An algorithm for computing the cumulative distribution function for magnitude squared coherence estimates," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 117-119, 1981.
- [205] V. P. Leonov and A. N. Shiryaev, "On a method of calculation of semiinvariants," *Theory Prob. Appl.*, vol. IV, pp. 319-329, 1959.
- [206] —, "Some problems in the spectral theory of higher-order moments, II," *Theory Prob. Appl.*, vol. V, pp. 417-421, 1960.
- [207] R. V. Lenth, "On finding the source of a signal," *Technometrics*, vol. 23, pp. 149-154, 1981.
- [208] R. V. Lenth, "Robust measures of location for directional data," *Technometrics*, vol. 23, pp. 77-81, 1981.
- [209] G. J. Leppink, "Efficient estimators in spectral analysis," *Proc. 12th Biennial Seminar, Can. Math. Congress*, R. Pyke, Ed., pp. 83-87 (Can. Math. Congress, Montreal), 1970.
- [210] P. Lesage and C. Audoin, "Characterization of frequency stability: Uncertainty due to the finite number of measurements," *IEEE Trans. Instrum. Meas.*, vol. IM-22, pp. 157-161, 1973.
- [211] B. Ya. Levit, and A. M. Samarov, "Estimation of spectral functions," *Prob. Infor. Trans.*, vol. 14, pp. 120-124, 1978.
- [212] W. C. Lindsey and C. M. Chie, "Theory of oscillator instability based upon structure functions," *Proc. IEEE*, vol. 64, pp. 1652-1666, 1976.
- [213] M. Loève, *Probability Theory*. New York: Van Nostrand, 1963.
- [214] Z. A. Lomnicki and S. K. Zaremba, "On estimating the spectral density function of a stochastic process," *J. Roy. Statist. Soc.*, vol. B-19, pp. 13-37, 1957.
- [215] —, "Bandwidth and resolvability in statistical spectral analysis," *J. Roy. Statist. Soc.*, vol. B-21, pp. 169-171, 1959.
- [216] R. Lugannani, "Asymptotic distribution of the sample coherence using overlap processing," in *Proc. 16th Allerton Conf. on Control and Computing* (Univ. of Illinois), pp. 403-412, 1978.
- [217] E. Lukacs, "A characterization of a bivariate gamma distribution," in *Multivariate Analysis-IV*, P. R. Krishnaiah, Ed. Amsterdam, The Netherlands: North-Holland, pp. 119-128.
- [218] I. B. MacNeill, "Tests for periodic components in multiple time series," *Biometrika*, vol. 61, pp. 57-70, 1974.
- [219] —, "A test of whether several time series share common periodicities," *Biometrika*, vol. 64, pp. 495-508, 1977.
- [220] T. A. Magness, "Spectral response of a quadratic device to non-Gaussian noise," *J. Appl. Phys.*, vol. 25, pp. 1357-1365, 1954.
- [221] H. Maitre, "Iterative superresolution. Some fast new methods," *Opt. Acta*, vol. 28, pp. 973-980, 1981.
- [222] C. L. Mallows, "Linear processes are nearly Gaussian," *J. Appl. Prob.*, vol. 4, pp. 313-329, 1967.
- [223] J. D. Markel and A. H. Gray, Jr., "On autocorrelation equations as applied to speech analysis," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 69-79, 1973.
- [224] J. E. Mazo, "On the angle between two Fourier subspaces," *Bell Syst. Tech. J.*, vol. 56, pp. 411-426, 1977.
- [225] R. N. McDonough, "A canonical form of the likelihood detector for Gaussian random vectors," *J. Acoust. Soc. Amer.*, vol. 49, pp. 402-406, 1970.
- [226] J.-M. Mendel, "White-noise estimators for seismic data processing in oil exploration," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 694-706, 1977.
- [227] K. S. Miller, "Complex linear least squares," *SIAM Rev.*, vol. 15, pp. 706-726, 1973.
- [228] —, *Complex Stochastic Processes*. Reading, MA: Addison-Wesley, 1974.
- [229] K. S. Miller, R. I. Bernstein and L. E. Blumenson, "Generalized Rayleigh processes," *Quart. J. Math.*, vol. 16, pp. 137-145, 1958.
- [230] K. Miller and G. A. Viano, "On the necessity of nearly-best-possible methods for analytic continuation of scattering data," *J. Math. Phys.*, vol. 14, pp. 1037-1048, 1973.
- [231] M. Miyata, "Complex generalization of canonical correlation and its application at a sea-level study," *J. Marine Res.*, vol. 28, pp. 202-214, 1970.
- [232] J. A. Mullen and D. Middleton, "The rectification of non-Gaussian noise," *Quart. Appl. Math.*, vol. 15, pp. 395-419, 1958.
- [233] M. Z. Nashed and G. Wanba, "Generalized inverses in reproducing kernel spaces: An approach to regularization of linear operator equations," *SIAM J. Math. Anal.*, vol. 5, pp. 974-987, 1974.
- [234] P. G. Neval, "Eigenvalue distribution of Toeplitz matrices," *Proc. Amer. Math. Soc.*, vol. 80, pp. 247-253, 1980.
- [235] A. H. Nuttall, and C. G. Carter, "A generalized framework for power spectral estimation," *IEEE Trans. Acoust., Speech, Sig-*

- nal Processing*, vol. ASSP-28, pp. 334–335, 1980.
- [236] R. L. Obenchain, “Good and optimal ridge estimates,” *Ann. Statist.*, vol. 6, pp. 1111–1121, 1978.
- [237] A. Oppenheim and J. S. Lim, “The importance of phase in signals,” *Proc. IEEE*, vol. 69, pp. 529–541, 1981.
- [238] R. K. Ottes and L. Enochson, *Digital Time Series Analysis*. New York: Wiley-Interscience, 1972.
- [239] K. Ozeki, “A coordinate-free theory of eigenvalue analysis related to the method of principal components and the Karhunen-Loeve expansion,” *Inform. Contr.*, vol. 42, pp. 38–59, 1979.
- [240] B. V. Pal’cev, “The asymptotics of the spectrum and eigenfunctions of convolution operators on a finite interval with a kernel with a homogeneous Fourier transform,” *Sov. Math.—Dokl.*, vol. 15, pp. 1243–1247, 1974.
- [241] H. A. Panofsky, “Meteorological applications of cross-spectrum analysis,” in *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967, pp. 109–132.
- [242] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1965.
- [243] —“Minimum-bias windows for high-resolution spectral estimates,” *IEEE Trans. Informat. Theory*, vol. IT-19, pp. 9–12, 1973.
- [244] —, “A new algorithm in spectral analysis and band-limited extrapolation,” *IEEE Trans. Circuits Syst.*, vol. CAS-22, pp. 735–742, 1975.
- [245] A. Papoulis and M. S. Bertran, “Digital filtering and prolate functions,” *IEEE Trans. Circuit Theory*, vol. CT-19, pp. 674–681, 1972.
- [246] E. Parzen, “On consistent estimates of the spectrum of a stationary time series,” *Ann. Math. Statist.*, vol. 28, pp. 329–348, 1957.
- [247] —, “Mathematical considerations in the estimation of spectra,” *Technometrics*, vol. 3, pp. 167–190, 1961.
- [248] —, *Time Series Analysis Papers*. San Francisco, CA: Holden-Day, 1967.
- [249] —, “Time series analysis for models of signal plus white noise,” in *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967, pp. 233–258.
- [250] —, “Statistical inference on time series by RKHS methods,” *Proc. 12th Biennial Seminar Can. Math. Congress*, R. Pyke, Ed. (Can. Math Congress, Montreal), pp. 1–37, 1970.
- [251] —, “Modern empirical statistical spectral analysis,” Tech. Rep. N-12, Texas A & M Univ., 1980.
- [252] —, “Autoregressive spectral estimation, log spectral smoothing, and entropy,” Tech. Rep. N-26, Texas A & M Univ., 1981.
- [253] K. L. Peacock, and S. Triel, “Predictive deconvolution: Theory and practice,” *Geophys.*, vol. 34, pp. 155–169, 1969.
- [254] W. L. Perry, “Approximate solutions of inverse problems with piecewise continuous solutions,” *Radio Sci.*, vol. 12, pp. 637–642, 1977.
- [255] D. L. Phillips, “A technique for the numerical solution of certain integral equations of the first kind,” *J. Assoc. Comput. Mach.*, vol. 9, pp. 84–97, 1962.
- [256] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes* (translated by A. Feinstein). San Francisco: Holden-Day, 1964.
- [257] V. F. Pisarenko, “On the estimation of spectra by means of nonlinear functions of the covariance matrix,” *Geophys. J., Roy. Astronomical Soc.*, vol. 28, pp. 511–531, 1972.
- [258] —, “The retrieval of harmonics from a covariance function,” *Geophys. J., Roy. Astronomical Soc.*, vol. 33, pp. 347–366, 1973.
- [259] J. Pliva, A. S. Pine, and P. D. Wilson, “Deconvolution of infrared spectra beyond the Doppler limit,” *Appl. Opt.*, vol. 19, pp. 1833–1837, 1980.
- [260] K. M. M. Prabhu and V. U. Reddy, “Data windows in digital signal processing—A review,” *J. Inst. Electr. Telecom. Eng.*, vol. 26, no. 1, pp. 69–76, 1980.
- [261] P. Prescott, “Variances and covariances of order statistics from the gamma distribution,” *Biometrika*, vol. 61, pp. 607–613, 1974.
- [262] M. B. Priestley, “The analysis of stationary processes with mixed spectra,” *J. Royal Statist. Soc.*, vol. B-24; p. I, pp. 215–233; p. II, pp. 511–529, 1962.
- [263] —, “Estimation of the spectral density function in the presence of harmonic components,” *J. Roy. Statist. Soc.*, vol. B-26, pp. 123–132, 1964.
- [264] —, “The role of bandwidth in spectral analysis,” *Appl. Statist.*, vol. 14, pp. 33–47, 1965.
- [265] —, *Spectral Analysis and Time Series I & II*. New York: Academic Press, 1981.
- [266] R. Prost and R. Goutte, “Deconvolution when the convolution kernel has no inverse,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 542–549, 1977.
- [267] A. J. Rainal, “Phase principle for measuring location or spectral shape of a discrete radio source,” *Bell Syst. Tech. J.*, vol. 47, pp. 415–428, 1968.
- [268] —, “Detecting a weak sine wave and measuring its parameters,” *IEEE Trans. Instrum. Meas.*, vol. IM-17, pp. 127–133, 1968.
- [269] C. R. Rao, *Linear Statistical Inference and Its Applications*. New York: Wiley, 1965.
- [270] Lord Rayleigh (J. W. Strutt), [1879–1880] “Investigations in optics, with special reference to the spectroscope,” in *Scientific Papers by Lord Rayleigh*. New York: Dover, 1964, article 62.
- [271] Lord Rayleigh (J. W. Strutt), [1912], “Remarks concerning Fourier’s theorem as applied to physical problems,” in *Scientific Papers by Lord Rayleigh*. New York: Dover, 1964, article 369.
- [272] D. R. Rhodes, “On the spheroidal functions,” *J. Res. Nat. Bur. Stand.*, vol. 74B, pp. 187–209, 1970.
- [273] —, “On a third kind of characteristic numbers of the spheroidal functions,” *Proc. Nat. Acad. Sci. (USA)*, vol. 67, pp. 351–355, 1970.
- [274] —, “On the Taylor distribution,” *IEEE Trans. Antennas Propagat.*, vol. AP-20, pp. 143–145, 1972.
- [275] G. G. Ricker and J. R. Williams, “Redundant processing sensitivity,” *IEEE Trans. Automat. Contr.*, vol. AU-21, pp. 348–354, 1973.
- [276] D. C. Rife, and R. R. Boorstyn, “Multiple tone parameter estimation from discrete-time observations,” *Bell Syst. Tech. J.*, vol. 55, pp. 1389–1410, 1976.
- [277] D. C. Rife, and G. A. Vincent, “Use of the discrete Fourier transform in the measurement of frequencies, and levels of tones,” *Bell Syst. Tech. J.*, vol. 49, pp. 197–228, 1970.
- [278] H. B. Riley and W. E. Alexander, “Deconvolution using the complex cepstrum,” *Proc. Southeastcon.*, pp. 854–858, 1981.
- [279] A. L. Roark and G. M. Wing, “A method for computing the eigenvalues of certain integral equations,” *Numerische Math.*, vol. 7, pp. 159–170, 1965.
- [280] L. Robin, *Fonctions Sphériques de Legendre et Fonctions Sphéroïdales, I-II*. Paris: Gauthier-Villars, 1957–1959.
- [281] Y. Ronen and M. Zakai, “The maximum likelihood estimator for a phase comparison angle measuring system,” *Proc. IEEE*, vol. 52, pp. 1669–1670, 1963.
- [282] M. Rosenblatt, “Some comments on narrow band-pass filters,” *Quart. Appl. Math.*, vol. 18, pp. 387–393, 1961.
- [283] M. Rosenblatt and J. W. Van Ness, “Estimation of the bispectrum,” *Ann. Math. Statist.*, vol. 36, pp. 1120–1136, 1965.
- [284] Y. A. Rozanov, “Spectral analysis of abstract functions,” *Theory Prob. Appl.*, vol. 4, pp. 271–287, 1959.
- [285] —, “On the applicability of the central limit theorem to stationary processes which have passed through a linear filter,” *Theory Prob. Appl.*, vol. 6, pp. 321–322, 1961.
- [286] —, *Stationary Stochastic Processes*. San Francisco, CA: Holden-Day, 1967.
- [287] C. K. Rushforth and R. W. Harris, “Restoration, resolution, and noise,” *J. Opt. Soc. Amer.*, vol. 58, pp. 539–545, 1968.
- [288] J. Rutman, “Characterization of phase and frequency instabilities in precision frequency sources: Fifteen years of progress,” *Proc. IEEE*, vol. 66, pp. 1048–1075, 1978.
- [289] M. S. Sabri and W. Steenaart, “An approach to band-limited extrapolation: The extrapolation matrix,” *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 74–78, 1978.
- [290] A. M. Samarov, “Lower bound on the risk for spectral density estimates,” *Prob. Infor. Trans.*, vol. 13, pp. 48–52, 1977.
- [291] T. K. Sarkar, D. D. Weiner, and V. K. Jain, “Some mathematical considerations in dealing with the inverse problem,” *IEEE Trans. Antennas Propagat.*, vol. AP-29, pp. 373–379, 1981.
- [292] H. Scheffé, *The Analysis of Variance*. New York: Wiley, 1959.
- [293] A. Schuster, “On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena,” *Terr. Magn.*, vol. 3, pp. 13–41, 1898.
- [294] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, 1948.
- [295] A. N. Shiryaev, “Some problems in the spectral theory of higher-order moments I,” *Theory Prob. Appl.*, vol. V, pp. 265–284, 1960.
- [296] Y. G. Shiryaev, “On conditions for ergodicity of stationary processes in terms of higher order moments,” *Theory Prob. Appl.*, vol. 436–439, 1963.
- [297] R. A. Silverman, “The fluctuation rate of the chi process,” *IRE Trans. Informat. Theory*, vol. IT-4, pp. 30–34, 1958.
- [298] Y. G. Sinai, “On higher order spectral measures of ergodic stationary processes,” *Theory Prob. Appl.*, vol. VIII, pp. 429–436, 1963.
- [299] M. T. Silvia, “The role of statistical filtering in the geophysical inverse scattering problem,” in *Proc. 24th Midwest Symp. on Circuits and Syst.*, S. Karni, Ed. (Univ. of New Mexico), pp. 76–81, 1981.
- [300] E. Sjøtoft, “A straightforward deconvolution method for use in small computers,” *Nuclear Inst. Methods*, vol. 163, pp. 519–522, 1979.
- [301] D. Slepian, “Estimation of signal parameters in the presence of noise,” *IRE Trans. Informat. Theory*, vol. IT-3, pp. 68–89, 1953.
- [302] —, “Fluctuations of random noise power,” *Bell Syst. Tech. J.*, vol. 37, pp. 163–184, 1958.
- [303] —, “Prolate spheroidal wave functions, Fourier analysis and

- uncertainty—IV," *Bell Syst. Tech. J.*, vol. 43, pp. 3009–3057, 1964.
- [304] —, "Some asymptotic expansions for prolate spheroidal wave functions," *J. Math. Phys.*, vol. 44, pp. 99–140, 1965.
- [305] —, "On bandwidth," *Proc. IEEE*, vol. 64, pp. 292–300, 1976.
- [306] —, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case," *Bell Syst. Tech. J.*, vol. 57, pp. 1371–1429, 1978.
- [307] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—I," *Bell Syst. Tech. J.*, vol. 40, pp. 43–64, 1961.
- [308] D. Slepian and E. Sonnenblick, "Eigenvalues associated with prolate spheroidal wave functions of zero order," *Bell Syst. Tech. J.*, vol. 44, pp. 1745–1759, 1965.
- [309] I. H. Sloan, "Iterated Galerkin method for eigenvalue problems," *SIAM J. Numer. Anal.*, vol. 13, pp. 753–760, 1976.
- [310] B. T. Smith *et al.*, "Matrix eigensystem routines—EISPACK guide," Lecture Notes in Computer Science no. 6. Berlin: Springer, 1976.
- [311] F. Smithies, *Integral Equations*. New York: Cambridge Univ. Press, 1962.
- [312] S. D. Stearns, "Tests of coherence unbiasing methods," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 321–323, 1981.
- [313] B. O. Steenson and N. C. Stirling, "The amplitude distribution and false-alarm rate of filtered noise," *Proc. IEEE*, vol. 53, pp. 42–55, 1965.
- [314] R. L. Stens, "Error estimates for sampling sums based on convolution integrals," *Infomat. Contr.*, vol. 45, pp. 37–47, 1980.
- [315] F. Stenger, "Numerical methods based on Whittaker cardinal, or SINC functions," *SIAM Rev.*, vol. 23, pp. 165–224, 1981.
- [316] M. Stone, "Application of a measure of information to the design and comparison of regression experiments," *Ann. Math. Statist.*, vol. 30, pp. 55–70, 1959.
- [317] O. N. Strand and E. R. Westwater, "Minimum-rms estimation of the numerical solution of a Fredholm integral equation of the first kind," *SIAM J. Numer. Anal.*, vol. 5, pp. 287–295, 1968.
- [318] A. H. Stroud and D. Secrest, *Gaussian Quadrature Formulas*. Englewood Cliffs, NJ: Prentice-Hall, 1966.
- [319] G. Szegő, "Beiträge zur theorie det Toeplitzchen formen," *Math. Zeitschr.*, vol. 6, pp. 167–202, 1920.
- [320] —, "Beiträge zur theorie det Toeplitzchen formen (fortsetzung)," *Math. Zeitschr.*, vol. 9, pp. 167–190, 1921.
- [321] —, *Orthogonal polynomials*, 3rd ed. New York: American Math Soc., 1967.
- [322] L. S. Taylor, "The phase retrieval problem," *IEEE Trans. Antennas Propagat.*, vol. AP-29, pp. 386–391, 1981.
- [323] D. J. Thomson, "Spectral analysis of short series," Ph.D. dissertation, Polytechnic Institute of Brooklyn, Brooklyn NY, 1971.
- [324] —, "Spectrum estimation techniques for characterization and development of WT4 waveguide," *Bell Syst. Tech. J.*, vol. 56, pt. I, pp. 1769–1815, pt II, pp. 1983–2005, 1977.
- [325] —, "Some recent developments in spectrum and harmonic analysis," in *Computer Science and Statistics: Proc. of the 13th Symposium on the Interface*, W. F. Eddy, Ed. New York: Springer, 1981, pp. 167–171.
- [326] —, "Spectrum estimation and harmonic analysis," in *Proc. 1st ASSP Workshop on Spectral Estimation*, S. Haykin, Ed. pp. 6.2.1–6.2.8, 1981.
- [327] D. J. Thomson, M. F. Robbins, C. G. MacLennan, and L. J. Lanzerotti, "Spectral and windowing techniques in power spectral analysis of geomagnetic data," *Phys. Earth Planetary Interiors*, vol. 12, pp. 217–231, 1976.
- [328] L. J. Tick, "Conditional spectra, linear systems, and coherency," in *Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, 1962, pp. 197–203.
- [329] —, "Estimation of coherency," *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967, pp. 133–152.
- [330] E. C. Titchmarsh, *Eigenfunction Expansions, I & II*. New York: Oxford Univ. Press, 1962.
- [331] P. H. Todd, "Direct minimal-order Markov model for sliding-window detection probabilities," *Proc. Inst. Elec. Eng.*, vol. 128, pt. F, pp. 152–154, 1981.
- [332] E. A. Trachtenberg, "Construction of fast unitary transforms which are equivalent to Karhunen-Loève spectral representations," in *Proc. IEEE Int. Symp. Electro.*, pp. 376–379, 1980.
- [333] J. R. Treichler, "Transient and convergent behaviour of the adaptive line enhancer," *IEEE Tran. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 53–62, 1979.
- [334] F. G. Tricomi, *Integral Equations*. New York: Wiley-Interscience, 1957.
- [335] D. W. Tufts and J. T. Francis, "Designing digital low-pass filters—Comparison of some methods and criteria," *IEEE Trans. Audio Electroacoust.*, vol. AU-18, pp. 487–494, 1970.
- [336] —, "Estimation and tracking of parameters of narrow-band signals by iterative processing," *IEEE Trans. Informat. Theory*, vol. IT-23, pp. 742–751, 1977.
- [337] J. W. Tukey, "Discussion, emphasizing the connection between analysis of variance and spectrum analysis," *Technometrics*, vol. 3, pp. 191–219, 1961.
- [338] —, "An introduction to the calculations of numerical spectrum analysis," in *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967.
- [339] —, "Can we predict where 'time series' should go next?," in *Directions in Time Series*, D. R. Brillinger and G. C. Tiao, Eds. New York: Inst. Math. Statist., 1980.
- [340] J. W. Van Ness, "Asymptotic normality of bispectral estimates," *Ann. Math. Statist.*, vol. 37, pp. 1257–1272, 1966.
- [341] H. L. Van Trees, *Detection, Estimation, and Modulation Theory I*. New York: Wiley, 1968.
- [342] G. Wahba, "On the distribution of some statistics useful in the analysis of jointly stationary time series," *Ann. Math. Statist.*, vol. 39, pp. 1849–1862, 1968.
- [343] —, "Automatic smoothing of the log periodogram," *J. Amer. Statist. Assoc.*, vol. 75, pp. 122–132, 1980.
- [344] A. M. Walker, "On the estimation of a harmonic component in a time series with stationary independent residuals," *Biometrika*, vol. 58, pp. 21–36.
- [345] G. G. Walter, "A class of spectral density estimators," *Ann Inst. Statist. Math.*, vol. 32A, pp. 65–80, 1980.
- [346] P. D. Welch, "A direct digital method of power spectrum estimation," *IBM J. Res. Devel.*, vol. 5, pp. 141–156, 1961.
- [347] —, "The use of the fast Fourier transform for estimation of spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, AU-15, pp. 70–74, 1967.
- [348] P. Whittle, "The simultaneous estimation of a time series harmonic components and covariance structure," *Trabajos de Estadística*, vol. 13, pp. 43–57, 1952.
- [349] —, "Estimation and information in stationary time series," *Ark. Für Matematik*, vol. 2, pp. 423–434, 1953.
- [350] —, *Prediction and Regulation by Linear Least-Squares Methods*. Princeton, NJ: Van Nostrand, 1963.
- [351] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. Cambridge, MA: M.I.T. Press, 1949.
- [352] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. Oxford, England: Clarendon, 1965.
- [353] S. S. Wilks, *Mathematical Statistics*. New York: Wiley, 1962.
- [354] E. J. Williams, "The analysis of association among many variates, with discussion," *J. Roy. Statist. Soc.*, vol. B-29, pp. 199–242, 1967.
- [355] J. R. Williams and G. G. Ricker, "Spectrum analyzer overlap requirements and detectability using discrete Fourier transform and composite digital filters," *J. Acoust. Soc. Amer.*, vol. 64, pp. 815–822, 1978.
- [356] R. A. Williams and W.S.C. Chang, "Resolution and noise in Fourier transform spectroscopy," *J. Opt. Soc. Amer.*, vol. 56, pp. 167–170, 1966.
- [357] M. B. Woodroffe and J. W. Van Ness, "The maximum deviation of sample spectral densities," *Ann Math. Statist.*, vol. 38, pp. 1558–1569, 1967.
- [358] A. D. Wyner, "Applications of discrete prolate spheroidal sequences to digital filtering," unpublished work.
- [359] K. Yoshimura, "Characterization of frequency stability: Uncertainty due to the autocorrelation of the frequency fluctuations," *IEEE Trans. Instrum. Meas.*, vol. IM-27, pp. 1–7, 1978.
- [360] C. K. Yuen, "Comments on modern methods of spectrum estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 298–299, 1979.
- [361] I. G. Zhurbenko, "Local properties of estimate of spectral function," *Prob. Infor. Trans.*, vol. 14, pp. 218–222, 1976.
- [362] D. S. Zrnic, "Spectral statistics for complex coloured discrete-time sequence," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 596–599, 1980.