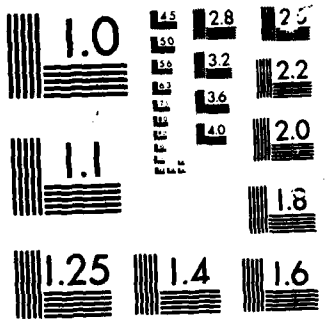
AD-A166 625 MULTICHANNEL RELATIVE-ENTROPY SPECTRUM ANALYSIS(U) 1/1
NAVAL RESEARCH LAB WASHINGTON DC R W JOHNSON ET AL.
25 MAR 86 NRL-MR-5723

UNCLASSIFIED F/G 9/4 NL



MICROCOPY CHART

(2)

NRL Memorandum Report 5723

Multichannel Relative-Entropy Spectrum Analysis

R. W. JOHNSON

Computer Science and Systems Branch
Information Technology Division

B. R. MUSICUS

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MS 02139

AD-A166 625

DTIC FILE COPY

DTIC
ELECTE
APR 1 6 1986
S D
D

March 25, 1986

NAVAL RESEARCH LABORATORY
Washington, D.C.

Approved for public release, distribution unlimited

62

AD-A166 625

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NRL Memorandum Report 5723	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Research Laboratory 6b. OFFICE SYMBOL (If applicable) Code 7593	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Washington, DC 20375-5000	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO: 61153N PROJECT NO: TASK NO: RR021-05-42 WORK UNIT ACCESSION NO: DN980-167

11. TITLE (Include Security Classification)
Multichannel Relative-Entropy Spectrum Analysis

12. PERSONAL AUTHOR(S)
Johnson, R.W. and Musicus, B.R.*

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM 10/83 TO 10/85	1986 March 25	26

16. SUPPLEMENTARY NOTATION
*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MS 02139

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Cross entropy, Information theory, Spectrum estimation, Relative entropy, Spectrum analysis, Spectral analysis

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A new relative-entropy method is presented for estimating the power spectral density matrix for multichannel data, given correlation values for linear combinations of the channels, and given an initial estimate of the spectral density matrix. A derivation of the method from the relative-entropy principle is given. The basic approach is similar in spirit to Multisignal Relative-Entropy Spectrum Analysis ~~of Johnson and Shore~~, but the results differ significantly because the present method does not arbitrarily require the final distributions of the various channels to be independent. For the special case of separately estimating the spectra of the signal and noise, given the correlations of their sum, Multichannel Relative-Entropy Spectrum Analysis turns into a two stage procedure. First a smooth power spectrum model is fitted to the correlations of the signal plus noise. Then final estimates of the spectra and cross spectra are obtained through linear filtering. For the special case where p uniformly spaced correlations are known, and where the initial estimate of the signal plus noise spectrum is all-pole with order p or less, this method fits a standard Maximum Entropy autoregressive spectrum to the noisy correlations, then linearly filters to calculate the signal and noise spectra and cross spectra. Consideration is given to the case where only an initial estimate of the noise power spectrum is available. An illustrative numerical example is given.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL R. W. Johnson	22b. TELEPHONE (Include Area Code) (202) 767-3490 22c. OFFICE SYMBOL Code 7591

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

CONTENTS

I. INTRODUCTION 1

II. RELATIVE ENTROPY 2

III. MULTICHANNEL RELATIVE-ENTROPY SPECTRUM ANALYSIS 3

IV. SPECTRUM ESTIMATION FROM CORRELATIONS OF SIGNAL PLUS NOISE 5

V. INTERPRETATION OF MULTICHANNEL RESA 7

VI. COMPUTATIONAL CONSIDERATIONS 10

VII. UNKNOWN INITIAL SIGNAL POWER SPECTRUM ESTIMATE 11

VIII. IMPROVING THE ESTIMATES WITH ADDITIONAL KNOWLEDGE 14

IX. EXAMPLE 15

X. DISCUSSION 20

REFERENCES 21

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Avail and/or
Dist Special
A-1
QUALITY INSPECTED 3

MULTICHANNEL RELATIVE-ENTROPY SPECTRUM ANALYSIS

I. INTRODUCTION

We examine the problem of estimating power spectra and cross spectra for multiple signals, given selected correlations of various linear combinations of the signals, and given an *initial* estimate of the spectral density matrix. We present a method that produces *final* estimates that are consistent with the given correlation information and otherwise as similar as possible to the initial estimates in a precise information-theoretic sense. The method is an extension of the Relative-Entropy Spectrum Analysis (RESA) of Shore [1] and of the Maximum-Entropy Spectral Analysis (MESA) of Burg [2, 3]. It reduces to RESA when there is a single signal and to MESA when the initial estimate is flat.

MESA starts with a set of p known data correlations. It then estimates a probability density for the signal that has as large an entropy as possible (is maximally "flat") but still satisfies the known correlations. Intuitively, the method seeks the most "conservative" density estimate that would explain the observed data. The resulting algorithm fits a smooth p^{th} order autoregressive power-spectrum model to the known correlations. This technique gives good, high-resolution spectrum estimates, particularly if the signal either is sinusoidal or has been generated by an autoregressive process of order p or less.

RESA [1] is based on an information-theoretic derivation that is quite similar to that of MESA, except that it incorporates an initial spectrum estimate. This prior knowledge can often improve the spectrum estimates when a reliable estimate of the shape of the overall signal spectrum is available. In the case where the initial spectral density estimate is flat, RESA reduces to MESA.

In this paper we derive a multichannel RESA method that estimates the joint probability density of a set of signals given correlations of various linear combinations of the signal and given an initial estimate of the signal probability densities. The estimator was briefly presented in [4]. Our basic approach is similar in spirit to the multisignal spectrum-estimation procedure in [5, 6], but the result differs significantly because that paper not only assumed that the initial probability-density estimates for the various signals were independent, but in effect imposed the same condition on the final estimates as well. We show that if this assumption is not made, the resulting final estimates are in fact *not* independent, but do take a form that is more intuitively satisfying. When applied to the case of estimating the power spectra and cross spectra of a signal and noise given selected correlations of their sum, our method first fits a smooth power spectrum model of the signal plus noise spectrum to the given correlations. It then uses a smoothing Wiener-Hopf filter to obtain the final estimates of the signal and the noise spectra. This Multichannel Relative-Entropy Spectrum Analysis method thus represents a bridge between the information theoretic methods and Bayesian methods for spectrum estimation from noisy data.

The last issue we consider in this paper is treated in a more tentative and exploratory manner. In certain filtering applications such as speech enhancement, relatively good estimates of a stationary noise background can be found during quiet periods when no signal is present.

The work of Bruce Musicus has been supported in part by the Advanced Research Projects Agency monitored by ONR under Contract N00014-81-K-0742 NR-049-506 and in part by the National Science Foundation under Grant ECS80-07102.

Manuscript approved October 31, 1985.

However, the signal spectrum may be changing relatively rapidly so that good initial estimates for this spectrum are not found as easily. Unfortunately, our technique, like the Bayesian methods, requires good initial estimates of both the signal and noise spectra. The simplest fix in the Bayesian estimation problem is to estimate the signal spectrum by spectral subtraction [7]. More sophisticated Bayesian methods estimate the signal model along with the signal and iterate between filtering steps and spectrum estimation steps [8, 9]. With these methods in mind, we consider several modifications to our Multichannel RESA method when a good initial signal estimate is not known. We try letting the initial signal spectrum estimate be infinite or flat, we try spectral subtraction, and we try estimating the initial signal density along with the final joint signal and noise density. Unfortunately, none of these approaches gives a truly convincing solution to the problem, and so the issue remains open.

II. RELATIVE ENTROPY

The Relative-Entropy Principle [10] can be characterized in the following way. Let $\mathbf{v}$ be a random variable with values drawn from a set $\mathbf{v} \in \mathbf{D}$ with probability density $q^{\dagger}(\mathbf{v})$. We will assume that this "true" density is unknown, and that all we have available is an *initial* estimate $p(\mathbf{v})$. Now suppose we obtain some information about the actual density that implies that $q^{\dagger}$, though unknown, must be an element of some convex set $\mathbf{Q}$ of densities. Suppose $p \notin \mathbf{Q}$. Since $\mathbf{Q}$ may contain many (possibly infinitely many) different probability densities, which of these should be chosen as the best estimate q of $q^{\dagger}$? And how should the initial estimate be incorporated into this decision?

The Relative Entropy Principle states that we should choose this *final* density $q(\mathbf{v})$ to be the one that minimizes the *relative entropy:*

$$H(q, p) = \int_{\mathbf{D}} q(\mathbf{v}) \log \frac{q(\mathbf{v})}{p(\mathbf{v})} \, d\mathbf{v} \tag{1}$$

subject to the condition $q \in \mathbf{Q}$. It has been shown [10] that minimizing any function other than $H(q, p)$ to estimate q must either give the same answer as minimizing relative entropy or else must contradict one of four axioms that any "reasonable" estimation technique must satisfy. These axioms require, for example, that the estimation method must give the same answer regardless of the coordinate system chosen. The function $H(q, p)$ has a number of useful properties: it is convex in q, it is convex in p, it is positive, and it is relatively convenient to work with computationally. If the convex set $\mathbf{Q}$ is closed and contains some q with $H(q, p) < \infty$, then there exists a $q \in \mathbf{Q}$ that minimizes (1) [11]. This solution is unique up to a set of measure zero.

Relative-entropy minimization was introduced as a general method of statistical inference by Kullback [12] and has been advocated by a variety of authors [13, 14, 15] under a variety of names, including cross-entropy [16], expected weight of evidence [17, p.72], directed divergence [12, p.7], discrimination information [12, p.37], and relative entropy [18, p.19]. The principle of Maximum Entropy [19, 20, 21] is a special case of the Relative-Entropy principle [10, 22] where the initial density is "flat" over the domain $\mathbf{D}$.

One application in which we can explicitly state the form of the relative-entropy solution q is where we observe the expected values $\overline{g_k}$ of a finite set of known functions $g_k(\mathbf{v})$ given the actual density $q^{\dagger}(\mathbf{v})$. Then the set $\mathbf{Q}$ of possible densities is defined by the *constraints:*

$$\int_{\mathbf{D}} g_k(\mathbf{v}) q(\mathbf{v}) \, d\mathbf{v} = \overline{g_k} \qquad \text{for } k = 1, \ldots, M \tag{2}$$

In addition, the density $q(\mathbf{v})$ must be properly normalized:

$$\int_{\mathbf{D}} q(\mathbf{v}) \, d\mathbf{v} = 1 \tag{3}$$

Because the constraints (2) and (3) are linear in q, the set $\mathbf{Q}$ of all probability densities satisfying these constraints must be convex. If the g_k are bounded functions, then $\mathbf{Q}$ is closed, and therefore there exists a density q that minimizes $H(q, p)$ subject to the constraints (provided these

are compatible with $H(q, p) < \infty$). In fact, even when the g_k are unbounded, the minimum-relative-entropy density q can be shown to exist under fairly general conditions; see [11] for a statement of such results.

Given the constraints (2) and (3), we wish to choose the final estimate $q(\boldsymbol{v})$ of $q^{\dagger}(\boldsymbol{v})$ by minimizing the relative entropy (1) subject to (2) and (3). To do this, we introduce Lagrange multipliers λ_k, construct the Lagrangian:

$$H(q, p) + (\lambda_0 - 1)\left[\int_D q(\boldsymbol{v})d\boldsymbol{v} - 1\right] + \sum_{k=1}^{M} \lambda_k \left[\int_D g_k(\boldsymbol{v})\, q(\boldsymbol{v})\, d\boldsymbol{v} - \overline{g}_k\right] \tag{4}$$

and set the variation with respect to q to zero. We obtain:

$$q(\boldsymbol{v}) = p(\boldsymbol{v}) \exp\left[-\lambda_0 - \sum_{k=1}^{M} \lambda_k\, g_k(\boldsymbol{v})\right] \tag{5}$$

It can be shown that if there is a solution $q(\boldsymbol{v})$ to the constrained minimization problem, then it must have the form (5) with the possible exception of a set of points on which the constraints imply that q vanishes [12, p.38; 11]. Conversely, if there are multipliers λ_k such that $q(\boldsymbol{v})$ in (5) satisfies the constraints (2) and (3), then $q(\boldsymbol{v})$ must be the unique element of $\mathbf{Q}$ that minimizes the relative entropy subject to the constraints [11]. When the g_k are complex functions, (2) is equivalent to two real constraints for each k. We then write (5) with complex Lagrange multipliers, define complex conjugate quantities $g_{-r} = g_r^*$, $\lambda_{-k} = \lambda_k^*$, and let k in the sum range over negative as well as positive values. In general, it is difficult to find closed-form solutions for the λ_k in terms of the constraints $\overline{g}_k$. Computational methods using gradient search have been developed, however [23].

III. MULTICHANNEL RELATIVE-ENTROPY SPECTRUM ANALYSIS

Let us apply this theory to estimating the spectra and cross-spectra of a set of L signals, or "channels", $x_0(t), \ldots, x_{L-1}(t)$, which we collect into a single vector-valued "multichannel" signal $\boldsymbol{x}(t) = (\, x_0(t) \;\cdots\; x_{L-1}(t)\,)^{\mathrm{T}}$. (In what follows, we will use italic type, such as x, P, for scalars; bold italic, such as $\boldsymbol{x}$, for column vectors; and Roman, such as P, for matrices. Superscripts T and H denote the transpose and the Hermitian adjoint, and a star denotes the complex conjugate.)

We assume that $\boldsymbol{x}(t)$ is a bandlimited stationary random complex process. To simplify the mathematics, we will assume that $\boldsymbol{x}(t)$ is a finite sum of complex exponentials at frequencies ω_n with random vector amplitudes $\boldsymbol{e}_n$:

$$\boldsymbol{x}(t) = \sum_{n=0}^{N-1} \boldsymbol{e}_n\, e^{i\omega_n t} \tag{6}$$

This involves no essential loss of generality, since an arbitrary stationary complex random process may be approximated by the form (6) with arbitrarily small mean square error on arbitrarily large finite time intervals by choosing the number of frequencies large enough and their spacing close enough [24, p.36].

Let $q^{\dagger}(\boldsymbol{e}_0, \ldots, \boldsymbol{e}_{N-1})$ be a joint probability density for the vector amplitudes $\boldsymbol{e}_n$. We can express the correlation matrix of the signal as

$$\begin{aligned} \mathrm{R}(\tau) &= \mathrm{E}\left[\lim_{T\to\infty} \frac{1}{T}\int_0^T \boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{H}}(t-\tau)\, dt\right] \\ &= \mathrm{E}\left[\sum_n \boldsymbol{e}_n \boldsymbol{e}_n^{\mathrm{H}}\, e^{i\omega_n \tau}\right] \end{aligned}$$

(expectation with respect to $q^{\dagger}$). Fourier transformation gives the power spectral matrix

$$
\begin{aligned}
S(\omega_n) &= E\left[c_n c_n^H \right] \\
&= \int c_n c_n^H q^\dagger(c_0, \ldots, c_{N-1})\, dc_0 \cdots dc_{N-1}
\end{aligned}
\tag{7}
$$

Let us choose an initial probability density estimate p of $q^\dagger$ such that the c_n are independent Gaussian random variables with zero mean and covariance $P(\omega_n)$:

$$
p(c_0, \ldots, c_{N-1}) = \prod_{n=0}^{N-1} p(c_n) \tag{8}
$$

$$
p(c_n) = N(\mathbf{0}, P(\omega_n)) = \frac{1}{\pi^L \det P(\omega_n)} \exp\left(- c_n^H P(\omega_n)^{-1} c_n \right)
$$

This choice of p corresponds to choosing the initial power spectrum estimate of $S(\omega_n)$ to be $P(\omega_n)$:

$$
P(\omega_n) = \int c_n c_n^H p(c_0, \ldots, c_{N-1})\, dc_0 \cdots dc_{N-1}
$$

This Gaussian assumption is usually considered reasonable and is often implicit in spectrum-analysis approaches such as Blackman-Tukey periodograms [25] or estimation procedures such as Wiener-Hopf filters [26]. For further discussion of the assumed form, see [27].

Now suppose we learn correlations R_k at various lags τ_k of various pairs of linear combinations $\alpha_k^H x(t)$, $\beta_k^H x(t)$ of the vector signal components:

$$
\begin{aligned}
R_k &= E\left[\lim_{T\to\infty} \frac{1}{T} \int_0^T \left(\alpha_k^H x(t) \right) \left(\beta_k^H x(t-\tau_k) \right)^* dt \right] \\
&= \alpha_k^H R(\tau_k) \beta_k = \alpha_k^H \sum_n S(\omega_n) e^{i\omega_n \tau_k} \beta_k
\end{aligned}
\tag{9}
$$

This rather general form includes measurements of correlations of pairs of single signal components, *i.e.* individual matrix elements of $R(\tau_k)$. As another special case, treated in the next section, it includes measurements of autocorrelations of the sum of the signal components. With the help of (7), this gives constraints in the standard form of (2) as follows:

$$
R_k = \int \left[\alpha_k^H \left(\sum_n c_n c_n^H e^{i\omega_n \tau_k} \right) \beta_k \right] q(c_0, \ldots, c_{N-1})\, dc_0 \cdots dc_{N-1} \tag{10}
$$

The Relative-Entropy final estimate of the probability density of the c_n coefficient given the initial estimate (8) and constraints (10) is then:

$$
q(c_0, \ldots, c_{N-1}) = p(c_0, \ldots, c_{N-1}) \exp\left(-\lambda_0 - \sum_k \lambda_k \sum_n c_n^H \beta_k \alpha_k^H c_n e^{i\omega_n \tau_k} \right) \tag{11}
$$

for some set of Lagrange multipliers λ_k, which are chosen so that $q(\mathbf{v})$ satisfies the constraints and is normalized (2). (Again we use the device of setting $\lambda_{-k} = \lambda_k^*$ and letting k in the sum run over negative values as well as positive. With the definitions $\tau_{-k} = -\tau_k$, $\alpha_{-k} = \beta_k$, $\beta_{-k} = \alpha_k$, this ensures a real result.) Substituting the formula (8) for $p(c_0, \ldots, c_{N-1})$ into (11) and simplifying puts the probability density estimate into the following elegant form:

$$
q(c_0, \ldots, c_{N-1}) = \prod_{n=0}^{N-1} q(c_n) \tag{12}
$$

where

$$
q(c_n) = N(\mathbf{0}, Q(\omega_n))
$$

$$Q(\omega_n) = \left[P(\omega_n)^{-1} + \sum_k \lambda_k \boldsymbol{\beta}_k \boldsymbol{\alpha}_k^H e^{i\omega_n \tau_k} \right]^{-1}$$

and where the unknowns λ_k must be determined from the constraints. Substituting this probability density into (10) and simplifying reduces the constraints to the form:

$$R_k = \boldsymbol{\alpha}_k^H \left[\sum_n Q(\omega_n) e^{i\omega_n \tau_k} \right] \boldsymbol{\beta}_k \tag{13}$$

Adjusting the λ_k until the latter equations are satisfied with $Q(\omega_n) \geq 0$ is a non-linear problem that must be solved, in general, by a non-linear gradient search technique.

The amplitudes $\boldsymbol{e}_n$ are *a posteriori* independent Gaussian random variables (*i.e.* have independent Gaussian final densities). Even if the channels of $\boldsymbol{x}(t)$ are *a priori* independent (*i.e.* have independent initial densities), so that the $P(\omega_n)$ matrices are all diagonal, the observation information concerns linear combinations of the channels, and as a result the covariance of the final density, $Q(\omega_n)$, will generally not be diagonal. Thus the final estimates of the various channels, unlike those in [5], will generally be correlated with each other.

IV. SPECTRUM ESTIMATION FROM CORRELATIONS OF SIGNAL PLUS NOISE

A special case of great practical interest is that in which we observe autocorrelations only for the sum of the signal components

$$y(t) = \sum_{i=0}^{L-1} x_i(t) = \boldsymbol{e}^T \boldsymbol{x}(t) \tag{14}$$

where $\boldsymbol{e} = (1 \;\; 1 \cdots 1)^T$. We then have:

$$R_k = \boldsymbol{e}^T \sum_n Q(\omega_n) \boldsymbol{e} \; e^{i\omega_n \tau_k} \tag{15}$$

These constraints are identical in form to those in (13) with $\boldsymbol{\alpha}_k = \boldsymbol{\beta}_k = \boldsymbol{e}$ for all k. We may often take the signal components $x_i(t)$ to be *a priori* uncorrelated, so that the power spectral density matrix $P(\omega_n)$ is diagonal for all ω_n. This restriction, however, is not necessary.

The Multichannel Relative-Entropy Spectrum Analysis estimate for $\boldsymbol{x}(t)$ from (12) is given by:

$$Q(\omega_n) = \left[P(\omega_n)^{-1} + \boldsymbol{e} \left(\sum_k \lambda_k e^{i\omega_n \tau_k} \right) \boldsymbol{e}^T \right]^{-1} \tag{16}$$

where the Lagrange multipliers λ_k are chosen to satisfy (15). The structure of this estimate is quite similar to the single-channel Relative-Entropy Spectrum Analysis (RESA) estimate given by [1] except that the quantities involved are matrices. Namely, the second term inside the brackets is the product of a scalar Σ, the summation, with $\boldsymbol{e}\,\boldsymbol{e}^T$, a square matrix of all 1's. In the single-signal case, $P(\omega_n)$ and $Q(\omega_n)$ become scalars, and we can replace $\boldsymbol{e}\,\Sigma\,\boldsymbol{e}^T$ with Σ; the result is just the RESA estimate. On the other hand, there is also a close formal connection with the Multisignal RESA estimate given in [5]. That is equivalent to the result of replacing $\boldsymbol{e}\,\Sigma\,\boldsymbol{e}^T$ in (16) by ΣI, where I is an identity matrix.

The expression (16) can be put into another interesting form by using the Woodbury-Sherman formula $(A + BDC)^{-1} = A^{-1} - A^{-1}B(CA^{-1}B + D^{-1})^{-1}CA^{-1}$:

$$Q(\omega_n) = P(\omega_n) - \frac{P(\omega_n)\boldsymbol{e}\;\boldsymbol{e}^T P(\omega_n)}{\boldsymbol{e}^T P(\omega_n)\boldsymbol{e} + \dfrac{1}{\sum_k \lambda_k e^{i\omega_n \tau_k}}} \tag{17}$$

Defining initial and final power-spectrum estimates for the summed signal $y(t)$ by

$$P_{yy}(\omega_n) = e^{\mathrm{T}}\mathrm{P}(\omega_n)e$$
$$Q_{yy}(\omega_n) = e^{\mathrm{T}}\mathrm{Q}(\omega_n)e$$

we obtain from (17):

$$Q_{yy}(\omega_n) = P_{yy}(\omega_n) - \frac{P_{yy}(\omega_n)^2}{P_{yy}(\omega_n) + \dfrac{1}{\sum_k \lambda_k e^{i\omega_n \tau_k}}} \tag{18}$$

and thus

$$Q_{yy}(\omega_n) = \frac{1}{\dfrac{1}{P_{yy}(\omega_n)} + \sum_k \lambda_k e^{i\omega_n \tau_k}} \tag{19}$$

This is precisely the form of the single-signal RESA final estimate with initial estimate $P_{yy}(\omega_n)$. We can write (15) as

$$R_k = \sum_n Q_{yy}(\omega_n)\, e^{i\omega_n \tau_k} \tag{20}$$

The Lagrange multipliers in (19) must be chosen to make $Q_{yy}(\omega_n)$ satisfy the correlation constraints (20). We can thus determine λ_k in (16) by solving a single-channel problem. That provides everything necessary to determine the solution $\mathrm{Q}(\omega_n)$ of the multichannel problem. We can in fact express $\mathrm{Q}(\omega_n)$ directly in terms of $Q_{yy}(\omega_n)$ and $\mathrm{P}(\omega_n)$; from (17) and (18) we obtain

$$\mathrm{Q}(\omega_n) = \mathrm{P}(\omega_n) + \frac{Q_{yy}(\omega_n) - P_{yy}(\omega_n)}{P_{yy}(\omega_n)^2}\, \mathrm{P}(\omega_n)e\ \ e^{\mathrm{T}}\mathrm{P}(\omega_n) \tag{21}$$

These equations summarize the Multichannel Relative-Entropy Spectrum Analysis method for correlations of a sum of signals. The calculation of the final spectral matrix proceeds in two steps. First we must find Lagrange multipliers such that the final estimate of the power spectrum of the sum of signals, $Q_{yy}(\omega_n)$ in (19), has the observed correlation values (20). Computationally, this generally requires a nonlinear gradient search algorithm to locate the correct λ_k [23]. Next the final spectral density matrix, $\mathrm{Q}(\omega_n)$ in (21), containing the cross-spectra as well as the power spectra of the individual signals, is formed by combining a linear multiple of the fitted power spectrum $Q_{yy}(\omega_n)$ with a constant term that depends only on the initial densities.

Frequently the multichannel signal $x(t)$ will comprise just two components, a signal $s(t)$ and an additive disturbance $d(t)$:

$$x(t) = \begin{pmatrix} s(t) \\ d(t) \end{pmatrix} = \sum_{n=0}^{N-1} \begin{pmatrix} \sigma_n \\ \delta_n \end{pmatrix} e^{i\omega_n t} \tag{22}$$

The initial estimate takes the form

$$\mathrm{P}(\omega_n) = \begin{pmatrix} P_{ss}(\omega_n) & P_{sd}(\omega_n) \\ P_{ds}(\omega_n) & P_{dd}(\omega_n) \end{pmatrix} = \mathrm{E}\left[\begin{pmatrix} \sigma_n \\ \delta_n \end{pmatrix} \begin{pmatrix} \sigma_n^* & \delta_n^* \end{pmatrix} \right]$$

The expression for $P_{yy}(\omega_n)$ specializes to:

$$P_{yy}(\omega_n) = \begin{pmatrix} 1 & 1 \end{pmatrix} \mathrm{P}(\omega_n) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \mathrm{E}\left[\left| \sigma_n + \delta_n \right|^2 \right]$$

We also define the initial cross-power spectra of $s(t)$ and $d(t)$ with respect to $y(t)$ as follows:

$$\begin{pmatrix} P_{sy}(\omega_n) \\ P_{dy}(\omega_n) \end{pmatrix} = \mathrm{P}(\omega_n) \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\left\{ P_{ys}(\omega_n) \;\; P_{yd}(\omega_n) \right\} = \left\{ 1 \;\; 1 \right\} \mathrm{P}(\omega_n)$$

We define the components $Q_{ss}(\omega_n)$, $Q_{sd}(\omega_n)$, $Q_{dd}(\omega_n)$, $Q_{yy}(\omega_n)$, $Q_{sy}(\omega_n)$ and $Q_{dy}(\omega_n)$ similarly. Then (21) becomes:

$$\mathrm{Q}(\omega_n) = \mathrm{P}(\omega_n) + \frac{Q_{yy}(\omega_n) - P_{yy}(\omega_n)}{P_{yy}(\omega_n)^2} \begin{bmatrix} P_{sy}(\omega_n) \\ P_{dy}(\omega_n) \end{bmatrix} \left\{ P_{ys}(\omega_n) \;\; P_{yd}(\omega_n) \right\} \tag{23}$$

An alternative formula for $\mathrm{Q}(\omega_n)$ in terms of $Q_{yy}(\omega_n)$ is:

$$\mathrm{Q}(\omega_n) = Q_{yy}(\omega_n) \begin{bmatrix} \dfrac{P_{sy}(\omega_n)}{P_{yy}(\omega_n)} \\ \dfrac{P_{dy}(\omega_n)}{P_{yy}(\omega_n)} \end{bmatrix} \left\{ \frac{P_{ys}(\omega_n)}{P_{yy}(\omega_n)} \;\; \frac{P_{yd}(\omega_n)}{P_{yy}(\omega_n)} \right\} + \frac{\det \mathrm{P}(\omega_n)}{P_{yy}(\omega_n)} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \left\{ 1 \;\; -1 \right\} \tag{24}$$

V. INTERPRETATION OF MULTICHANNEL RESA

The formulas defining Multichannel RESA have an interesting and profound structure that may not be obvious at first glance. First of all, formula (21) makes it easy to state conditions under which the matrix estimate $\mathrm{Q}(\omega_n)$ is positive definite. Next, the appearance of $Q_{yy}(\omega_n)$ in the constraint equation for the λ_k is actually something we should have expected by the property of subset aggregation that Relative-Entropy estimators satisfy [28]. Furthermore, formula (24), which builds the spectral density matrix estimate $\mathrm{Q}(\omega_n)$ by linearly filtering the fitted model spectrum $Q_{yy}(\omega_n)$, is identical in form to the standard Bayesian formula for the final expected power and cross-power in two signals given the value of their sum. In particular, the first term in (24) applies the well-known Wiener-Hopf smoothing filter [26] to $Q_{yy}(\omega_n)$, while the second term can be interpreted as the expected final variance of σ_n and δ_n. Finally, we will show that the relative entropy $H(q, p)$ has the same form as a generalized Itakura-Saito distortion measure [29]. Thus minimizing relative entropy in this problem is equivalent to finding the spectral matrix $\mathrm{Q}(\omega_n)$ with minimum Itakura-Saito distortion.

A. Positive Definiteness

Assume that the initial spectral density matrices are positive definite, $\mathrm{P}(\omega_n)>0$; then $P_{yy}(\omega_n)>0$ also for all ω_n. This implies that $\mathrm{Q}(\omega_n)$ in (21) is at least well-defined, provided we can find some $Q_{yy}(\omega_n)$ that satisfies the correlation constraints. Assume moreover that $Q_{yy}(\omega_n)$ is strictly positive, $Q_{yy}(\omega_n)>0$. Let $\mathbf{u}$ be any nonzero vector. Since $\mathrm{P}(\omega_n)$ is positive definite, we can write $\mathbf{u} = a\mathbf{e} + \mathbf{v}$ for some scalar a and some vector $\mathbf{v}$ such that $\mathbf{v}^{\mathrm{H}}\mathrm{P}(\omega_n)\mathbf{e} = 0$. Then (21) implies $\mathbf{u}^{\mathrm{H}}\mathrm{Q}(\omega_n)\mathbf{u} = |a|^2 Q_{yy}(\omega_n) + \mathbf{v}^{\mathrm{H}}\mathrm{P}(\omega_n)\mathbf{v}$, and at least one of the two terms on the right-hand side must be positive. Thus $\mathbf{u}^{\mathrm{H}}\mathrm{Q}(\omega_n)\mathbf{u} > 0$ for every nonzero vector $\mathbf{u}$; that is, $\mathrm{Q}(\omega_n)$ is strictly positive definite, $\mathrm{Q}(\omega_n)>0$.

B. Subset Aggregation Property

Consider the following two approaches to estimating the final density of the frequency components $\eta_n = \sigma_n + \delta_n$ of $y(t) = s(t) + d(t)$, or equivalently, estimating the final power spectrum $Q_{yy}(\omega_n)$:

Method #1:

a) Apply Relative Entropy to estimate the joint final density of σ_n, δ_n:

$$q\left(\begin{bmatrix} \sigma_n \\ \delta_n \end{bmatrix} \right) = N(\mathbf{0}, \mathrm{Q}(\omega_n))$$

b) Form the final density $q(\eta_n)$ for $\eta_n = \sigma_n + \delta_n$ from $q\left(\begin{pmatrix} \sigma_n \\ \delta_n \end{pmatrix} \right)$:

$$q(\eta_n) = N\left(0 , \begin{pmatrix} 1 & 1 \end{pmatrix} Q(\omega_n) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right)$$

Method #2:

a) Form an initial probability density estimate $p(\eta_n)$ from $p\left(\begin{pmatrix} \sigma_n \\ \delta_n \end{pmatrix} \right)$:

$$p(\eta_n) = N\left(0 , \begin{pmatrix} 1 & 1 \end{pmatrix} P(\omega_n) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) = N(0 , P_{yy}(\omega_n))$$

b) Solve the Relative-Entropy problem to find the final density of $q(\eta_n)$ given the constraints on the correlations of $y(t)$ and the initial density $p(\eta_n)$.

Method #1 is exactly the approach we have taken in this paper—the resulting spectrum estimates $Q_{yy}(\omega_n) = \begin{pmatrix} 1 & 1 \end{pmatrix} Q(\omega_n) \begin{pmatrix} 1 & 1 \end{pmatrix}^T$ are given by equations (20) and (19). According to the theory of Relative Entropy, both methods for estimating $q(\eta_n)$ ought to give identical results. In fact, if we follow method #2 and apply the Relative-Entropy result in (16) to the "single-channel" signal $y(t)$, we will find the following final density estimate $\overline{q}$ for $y(t)$:

$$\overline{q}(\eta_0, \ldots, \eta_{N-1}) = \prod_{n=0}^{N-1} \overline{q}(\eta_n)$$

$$\overline{q}(\eta_n) = N(0 , \overline{Q}_{yy}(\omega_n))$$

where

$$\overline{Q}_{yy}(\omega_n) = \left(\frac{1}{P_{yy}(\omega_n)} + \sum_{k=1}^{M} \lambda_k e^{i\omega_n \tau_k} \right)^{-1}$$

and the Lagrange multipliers λ_k are chosen so that the correlations of $\overline{Q}_{yy}(\omega_n)$ have the correct values:

$$R_k = \sum_n \overline{Q}_{yy}(\omega_n) e^{i\omega_n \tau_k}$$

As expected, these formulas for $\overline{Q}_{yy}(\omega_n)$ are identical to those for $Q_{yy}(\omega_n)$ in (19) and (20), and thus the two methods do indeed give identical results.

C. Bayesian Filtering Interpretation

Formula (24) appears complicated, but it is in fact precisely the same form as the result found by a purely Bayesian approach to the problem. Consider the following situation. Suppose we are given, not the correlations R_k, but the actual exact frequency components η_n of $y(t)$. What is the final density of σ_n and δ_n given these values $\eta_n = \sigma_n + \delta_n$? We will take a purely Bayesian approach. Because σ_n and δ_n are *a priori* jointly Gaussian, so are σ_n and η_n, their joint density is

$$p\left(\begin{pmatrix} \sigma_n \\ \eta_n \end{pmatrix} \right) = N\left(\mathbf{0} , \begin{pmatrix} P_{ss}(\omega_n) & P_{sy}(\omega_n) \\ P_{ys}(\omega_n) & P_{yy}(\omega_n) \end{pmatrix} \right) \tag{25}$$

Now given η_n, we can find the conditional density of σ_n given η_n by the standard Bayesian formula:

$$p(\sigma_n \mid \eta_n) = N\left(\frac{P_{sy}(\omega_n)}{P_{yy}(\omega_n)} \eta_n \ , \ P_{ss}(\omega_n) - \frac{P_{sy}(\omega_n)P_{ys}(\omega_n)}{P_{yy}(\omega_n)} \right) \tag{26}$$

Let us define the Bayesian estimate $Q_B(\omega_n)$ of the power spectral density matrix of $(\ \sigma_n \quad \delta_n\ )^T$ given η_n as follows:

$$Q_B(\omega_n) \equiv E\left[\begin{pmatrix} \sigma_n \\ \delta_n \end{pmatrix} \begin{pmatrix} \sigma_n^* & \delta_n^* \end{pmatrix} \middle| \eta_n \right]$$

But then:

$$\begin{aligned} Q_B(\omega_n) &= E\left[\begin{pmatrix} \sigma_n \\ \eta_n - \sigma_n \end{pmatrix} \begin{pmatrix} \sigma_n & \eta_n - \sigma_n \end{pmatrix}^* \middle| \eta_n \right] \\ &= E\left[\begin{pmatrix} \sigma_n \\ \eta_n - \sigma_n \end{pmatrix} \middle| \eta_n \right] E\left[\begin{pmatrix} \sigma_n & \eta_n - \sigma_n \end{pmatrix} \middle| \eta_n \right] + \mathrm{Var}\left[\begin{pmatrix} \sigma_n \\ \eta_n - \sigma_n \end{pmatrix} \middle| \eta_n \right] \end{aligned} \tag{27}$$

Substituting the final mean and variance of σ_n given η_n in (26) into (27) and recognizing that:

$$\begin{aligned} \frac{P_{ss}(\omega_n)P_{yy}(\omega_n) - P_{sy}(\omega_n)P_{ys}(\omega_n)}{P_{yy}(\omega_n)} &= \frac{1}{P_{yy}(\omega_n)} \det \begin{pmatrix} P_{ss}(\omega_n) & P_{sy}(\omega_n) \\ P_{ys}(\omega_n) & P_{yy}(\omega_n) \end{pmatrix} \\ &= \frac{1}{P_{yy}(\omega_n)} \det \mathrm{P}(\omega_n) \end{aligned}$$

gives:

$$Q_B(\omega_n) = |\eta_n|^2 \begin{pmatrix} \dfrac{P_{sy}(\omega_n)}{P_{yy}(\omega_n)} \\ \dfrac{P_{dy}(\omega_n)}{P_{yy}(\omega_n)} \end{pmatrix} \begin{pmatrix} \dfrac{P_{ys}(\omega_n)}{P_{yy}(\omega_n)} & \dfrac{P_{yd}(\omega_n)}{P_{yy}(\omega_n)} \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \frac{\det \mathrm{P}(\omega_n)}{P_{yy}(\omega_n)} \tag{28}$$

This Bayesian spectral density estimate, however, is identical in form to the Multichannel Relative-Entropy Spectrum Analysis estimate $Q(\omega_n)$ in (24), except that the Bayesian method uses the known signal plus noise spectral power $|\eta_n|^2$ in (28) while the Multichannel RESA estimate must use the smooth fitted signal plus noise power spectrum estimate $Q_{yy}(\omega_n)$. This similarity between the two methods provides an interesting interpretation of the Multichannel RESA formula. The first term in (24) is the product of the expectation of $(\ \sigma_n \quad \delta_n\ )^T$ and the expectation of $(\ \sigma_n^* \quad \delta_n^*\ )$. The vector elements $P_{sy}(\omega_n)/P_{yy}(\omega_n)$ and $P_{dy}(\omega_n)/P_{yy}(\omega_n)$ are exactly the smoothing Wiener-Hopf filter expressions that arise when estimating a signal $(\ \sigma_n \quad \delta_n\ )^T$ from an observation η_n. The second term in the Multichannel RESA estimate depends only on the prior information and corresponds to the *a posteriori* covariance of the spectrum estimates. The term:

$$\frac{\det \mathrm{P}(\omega_n)}{P_{yy}(\omega_n)}$$

thus can be used as a crude estimate of the variance of our model at ω_n.

D. Asymptotic Behavior

It is interesting to examine how the Multichannel Relative-Entropy algorithm performs in the asymptotic limit as the true correlations of the signal plus noise are known at all possible lags. In this asymptotic limit, the only solution $Q_{yy}(\omega_n)$ that can possibly satisfy all the correlation constraints is the true power spectrum of the signal plus noise. Unfortunately, the allocation of this power between the signal and noise spectral estimates and cross-spectrum estimates is determined entirely by the initial estimates of these spectra via the formula (24). Therefore, beyond a certain limit, gathering more and more correlation values will only improve the estimate of

$Q_{yy}(\omega_n)$ and will not improve the estimates of the signal or noise power spectra or cross-power spectra.

E. Generalized Itakura-Saito Distortion Measure

If we substitute any zero-mean, Gaussian densities

$$q(e_n) = N(\mathbf{0}, \mathrm{Q}(\omega_n))$$

and

$$p(e_n) = N(\mathbf{0}, \mathrm{P}(\omega_n))$$

into the relative-entropy formula we get:

$$H(q, p) = \sum_n \mathrm{tr}\left\{ \mathrm{Q}(\omega_n)\mathrm{P}(\omega_n)^{-1} - \mathrm{I} \right\} - \log\det\left(\mathrm{Q}(\omega_n)\mathrm{P}(\omega_n)^{-1} \right)$$

This is just a generalized version of the Itakura-Saito distortion measure [29]. We therefore could have derived the same spectrum estimate by minimizing the Itakura-Saito distortion measure over all possible spectral matrices $\mathrm{Q}(\omega_n)$ subject to the constraints (13).

VI. COMPUTATIONAL CONSIDERATIONS

The difficult step in the Multichannel RESA procedure is to solve for the Lagrange multipliers that will give $Q_{yy}(\omega_n)$ the appropriate correlations in (20). Gradient search algorithms for computing this in general are given by Johnson [23]. Once $Q_{yy}(\omega_n)$ is known, the components of $\mathrm{Q}(\omega_n)$ may be easily found by filtering $Q_{yy}(\omega_n)$ and adding in the final covariance estimate, an amount of computation that is linear in the number of frequency samples.

One special case is particularly easy to solve. This is when correlations of $y(t)$ are given for uniformly spaced lags $\tau_k = -p, -p+1, \ldots, p$ and when the initial spectral density of the signal plus noise, $P_{yy}(\omega_n) = (1\ 1)\,\mathrm{P}(\omega_n)\,(1\ 1)^{\mathrm{T}}$ is autoregressive of order at most p. Let us take the limiting form of our equations for equispaced frequencies as the spacing becomes extremely small, so that we can treat the spectral densities as continuous functions of ω. Then because $P_{yy}(\omega_n)$ is autoregressive (all-pole), the term $1/P_{yy}(\omega_n)$ in the denominator of

$$Q_{yy}(\omega) = \frac{1}{\dfrac{1}{P_{yy}(\omega)} + \sum_{k=-p}^{p} \lambda_k e^{i\omega k}}$$

has the same form as the sum over k. We can therefore combine coefficients in the two sums and write

$$Q_{yy}(\omega) = \frac{1}{\sum_{k=-p}^{p} \beta_k e^{i\omega k}}$$

where $\beta_{-p}, \ldots, \beta_p$ are to be determined so that:

$$R_k = \int_{-\pi}^{\pi} Q_{yy}(\omega) e^{i\omega k} \frac{d\omega}{2\pi}$$

This, however, is the standard Maximum Entropy Spectrum Analysis problem (MESA), and can be solved with $O(p^2)$ calculation by Levinson recursion [30, 31, 32]. Thus in this special case, the power spectrum of the signal plus noise $Q_{yy}(\omega)$ will be set to the MESA estimate. The initial estimate $P_{yy}(\omega)$ will be completely ignored in this step. (This is an example of "prior washout", as discussed by Shore and Johnson [28].) The spectral density matrix estimate $\mathrm{Q}(\omega)$ for $s(t)$ and $d(t)$ will then be formed as the appropriately filtered function of this MESA spectrum. In this case, therefore, the prior information is used solely to control how the estimates for the signal and

the disturbance are to be obtained from the MESA estimate. Note that although $P_{yy}(\omega)$ and $Q_{yy}(\omega)$ will both be autoregressive spectra, the individual components of the spectral density matrices $P(\omega)$ and $Q(\omega)$ will generally *not* be autoregressive.

VII. UNKNOWN INITIAL SIGNAL POWER SPECTRUM ESTIMATE

One potential difficulty in applying our Multichannel RESA algorithm is that, like the Bayesian algorithm, it requires good prior knowledge of both the signal and the noise spectra. Unfortunately, in some applications such as speech enhancement, good prior knowledge may be available only about the noise spectrum. In such cases it will be necessary to choose an initial estimate of $P_{ss}(\omega_n)$ by some rule.

In [6], the method of [5] is extended to allow for the possibility that not all initial estimates are equally reliable. It is shown how to associate weights with initial estimates. When there is a good initial noise estimate but a poor initial speech estimate, one gives a large weight to the initial noise estimate and a small weight to the initial speech estimate. The result is that the final noise estimate tends to stay close to the initial estimate, while the final speech estimate does most of the varying to conform to the constraints. An analogous method for incorporating weights in our present procedure would be desirable, but we do not yet have such a method. Straightforward imitation of the development in [6], but without the assumption of independent final estimates, yields final estimates that do not depend on the weights introduced. Accordingly we explore several other approaches.

The one that probably best fits the general philosophy of Multichannel RESA is to assume that the signal and noise are *a priori* independent, and that the initial signal density $p(\sigma_0, \ldots, \sigma_{N-1})$ is "flat" over the domain of interest, or equivalently, that the signal variance is very large. Substituting this flat initial signal density times a zero-mean Gaussian initial noise density into the relative entropy formula (1), and then solving for the resulting spectral estimate $Q(\omega_n)$ yields:

$$Q(\omega_n) = \left[\begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{P_{dd}}(\omega_n) \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \left(\sum_k \lambda_k e^{i\omega_n \tau_k} \right) \begin{pmatrix} 1 & 1 \end{pmatrix} \right]^{-1}$$

This is the same solution as (16) but with $P_{sd}(\omega_n)=0$ and with $P_{ss}(\omega_n)$ set to infinity. Inverting the spectral matrix and using the same algebraic manipulations as before gives the following equivalent formula for $Q(\omega_n)$:

$$R_k = \sum_n Q_{yy}(\omega_n) e^{i\omega_n \tau_k}$$

$$Q_{yy}(\omega_n) = \frac{1}{\sum_k \lambda_k e^{i\omega_n \tau_k}}$$

$$Q(\omega_n) = \begin{pmatrix} Q_{yy}(\omega_n) + P_{dd}(\omega_n) & -P_{dd}(\omega_n) \\ -P_{dd}(\omega_n) & P_{dd}(\omega_n) \end{pmatrix}$$

This, unfortunately, is not quite the answer one might have expected. We first fit a smooth autoregressive spectrum $Q_{yy}(\omega_n)$ to the correlations of the signal plus noise; this is exactly the MESA spectrum estimate, and it ignores the prior information about the noise. This spectrum is then allocated between the signal and noise in a rather peculiar way: the signal spectrum is estimated by *adding* the noise spectrum to the smooth estimate of the signal plus noise. Also the final noise power spectrum is estimated as exactly equal to the initial noise spectrum $P_{dd}(\omega_n)$; none of the observed data affects this estimate at all.

In retrospect, this structure is not terribly surprising. We started by assuming infinite signal variance at all frequencies. The observed correlation data is then taken into account by striking a compromise between the fitted noisy power spectrum $Q_{yy}(\omega_n)$ and the prior belief that $P_{ss}(\omega_n) = \infty$. This compromise, inevitably, is *larger* than the fitted noisy signal spectrum. Because virtually nothing is known about the signal, any deviation of $Q_{yy}(\omega_n)$ from the noisy spectrum $P_{dd}(\omega_n)$ is considered to be due to the signal, and thus the initial noise spectrum estimate is accepted as truth.

Thus, although setting $P_{ss}(\omega_n) = \infty$ may be the philosophically correct approach when no information about the signal is given, it leads to results that are counterintuitive and not very useful. The problem is that we generally know something about the signal, if only that it has finite energy. This might lead us to consider choosing as the initial signal density estimate a density that has maximum entropy subject to a constraint on the total signal energy. This corresponds to an initial signal power spectrum estimate that is flat across all frequencies, with constant value P_0. The problem is picking a good choice for P_0. Choose P_0 too small and most of the fitted spectrum $Q_{yy}(\omega_n)$ will be allocated to the noise; choose it too large, and most of $Q_{yy}(\omega_n)$ will be allocated to the signal. P_0 thus becomes an experimental parameter that must be varied to achieve a desirable effect.

Another, more *ad hoc* idea is to try estimating $P_{ss}(\omega_n)$ by spectral subtraction. Suppose we are given the observed data waveform $y(t)$ with complex exponential amplitudes η_n. To obtain the initial signal power spectrum estimate, subtract the estimated noise power from the observed periodogram and clip the result to positive values:

$$P_{ss}(\omega_n) = \max\left(\ |\eta_n|^2 - P_{dd}(\omega_n) \, , 0 \ \right) \tag{29}$$

Then calculate a few low order correlations from $y(t)$, and combine them with this initial estimate to estimate $Q(\omega_n)$. Unfortunately, unless $|\eta_n|^2$ dips below $P_{dd}(\omega_n)$ over significantly large numbers of frequency samples, it is easy to see that our estimation procedure will approximately set:

$$Q_{yy}(\omega_n) \approx P_{yy}(\omega_n) = \max\left(\ |\eta_n|^2 \, , P_{dd}(\omega_n) \ \right) \tag{30}$$

$$Q(\omega_n) \approx P(\omega_n) = \begin{bmatrix} P_{ss}(\omega_n) & 0 \\ 0 & P_{dd}(\omega_n) \end{bmatrix}$$

This spectrum estimate unfortunately differs little from the spectral subtraction estimate, and it has the large variance of the periodogram, a situation that Multichannel RESA was supposed to cure.

We consider one final approach, which seems too reasonable to resist. Since the initial signal density is unknown, one might argue, perhaps it should be estimated along with the final density q. This turns out to be a rather bad idea, but is interesting enough to pursue for a while. Let us assume that the signal and noise are *a priori* independent, so that:

$$p(\sigma,\delta) = p_s(\sigma) p_d(\delta) \tag{31}$$

We will assume that the noise density $p_d(\delta)$ has the form in

$$p(\delta) = \prod_{n=0}^{N-1} p(\delta_n) \tag{32}$$

$$p(\delta_n) = N(0, P_{dd}(\omega_n))$$

but that the initial signal density $p_s(\sigma)$ is to be estimated along with the final joint density $q(\sigma,\delta)$ by minimizing the relative entropy over both q and p_s subject to the correlation constraints on q. We write this as:

$$q\,,\,p_s \leftarrow \min_{q\in\mathbf{Q},\,p_s} H(q\,,\,p_s\,p_d) \tag{33}$$

where $\mathbf{Q}$ is the set of densities satisfying the correlation constraints (15). Let us solve this minimization iteratively, minimizing first over all $q \in \mathbf{Q}$, then over all p, iterating back and forth until the estimates converge:

For $k = 0,1,\cdots$

$$q_{k+1} \leftarrow \min_{q\in\mathbf{Q}} H(q\,,\,p_{s_k}p_d) \tag{34}$$

$$p_{s_{k+1}} \leftarrow \min_{p_s} H(q_{k+1},\,p_s\,p_d)$$

The first problem is precisely our usual Multichannel RESA algorithm for estimating the final density $q(\sigma,\delta)$ from the initial density $p_{s_k}(\sigma)p_d(\delta)$ and the correlations R_k. To solve the second step, factor $q(\sigma,\delta) = q(\sigma)q(\delta \mid \sigma)$ and rewrite $H(q_k\,,p_s\,p_d)$ in the form:

$$H(q_{k+1},p_s\,p_d) = \int q_{k+1}(\sigma)\log\frac{q_{k+1}(\sigma)}{p_s(\sigma)}\,d\sigma \tag{35}$$

$$+\iint q_{k+1}(\sigma,\delta)\log\frac{q_{k+1}(\delta \mid \sigma)}{p_d(\delta)}\,d\delta\,d\sigma$$

Since $p_s(\sigma)$ only appears in the first term, and since this first term is bounded below by 0, the minimum of $H(q_{k+1},\,p_s\,p_d)$ over p_s is achieved by:

$$p_{s_{k+1}}(\sigma) = q_{k+1}(\sigma) = \int q_{k+1}(\sigma,\delta)\,d\sigma \tag{36}$$

Putting all this together implies that $q(\sigma,\delta)$ will generally be Gaussian with the same form as before, and that $p_{s_{k+1}}(\sigma)$ will also be Gaussian with covariance $P_{ss}(\omega_n) = Q_{ss}(\omega_n)$. The complete iterative algorithm then takes the following form:

Guess $P_{ss}(\omega_n)$

For $k = 0,1,\cdots$

a) Use the Multichannel RESA algorithm to compute $Q(\omega_n)$

b) Set $P_{ss}(\omega_n) = Q_{ss}(\omega_n)$

Thus we run our usual Multichannel RESA algorithm to get a good estimate of our signal and noise spectra. We then set our initial signal spectrum estimate to this improved final signal spectrum and iterate. The improved initial density should lead to an even better final density. Each iteration drives the relative entropy to ever smaller values, and thus each set of spectrum estimates ought to be better than the last.

Alas, appearances can be deceiving. This idea of feeding the Multichannel RESA process back on itself, using the final estimates as a new initial estimate for the next iteration, has been informally suggested by numerous researchers, but in this case, where the procedure converges depends entirely on where it starts. The problem is that the original double minimization problem (33) may have infinitely many solutions. For example, suppose there is at least one spectrum $P_{yy}(\omega_n)$ that is larger than the initial noise spectrum estimate, $P_{yy}(\omega_n) > P_{dd}(\omega_n)$, and has the correct correlations R_k. (There may be infinitely many such densities.) Let $P_{ss}(\omega_n) \equiv P_{yy}(\omega_n) - P_{dd}(\omega_n)$. Then the corresponding initial density $p_s(\sigma)p_d(\delta)$ will satisfy the constraints (15), and thus the Relative-Entropy final estimate will just be $q(\sigma,\delta) = p_s(\sigma)p_d(\delta)$ and the Relative Entropy at this solution will be $H(q\,,\,p_s\,p_d) = 0$, which is as low as it will ever get. Because there could be an infinite number of choices for $P_{yy}(\omega_n)$, this relative entropy problem could have an infinite number of minimizing solutions. Which solution our iterative algorithm converges to will thus depend on where we start. It could be argued that this iteration is so appealing that perhaps we should use it anyway, but stop after only a couple of iterations. Unfortunately, this approach is difficult to justify on anything but empirical grounds.

VIII. IMPROVING THE ESTIMATES WITH ADDITIONAL KNOWLEDGE

The spectral estimates gained from the Multichannel RESA algorithm can be significantly improved if additional knowledge about the process can be incorporated. We will discuss two particular types of knowledge which improve the spectral estimates substantially for the case where we observe correlations of signal plus noise.

Suppose that we observe the correlations R_k of a signal plus noise, $y(t) = s(t) + d(t)$, as in section IV. Suppose, however, that we add the constraint that the signal and noise are known to be independent, so that the densities must have the form:

$$p(\sigma_0, \ldots, \sigma_{N-1}, \delta_0, \ldots, \delta_{N-1}) = p(\sigma_0, \ldots, \sigma_{N-1})\, p(\delta_0, \ldots, \delta_{N-1}) \tag{37}$$

and:

$$q(\sigma_0, \ldots, \sigma_{N-1}, \delta_0, \ldots, \delta_{N-1}) = q(\sigma_0, \ldots, \sigma_{N-1})\, q(\delta_0, \ldots, \delta_{N-1}) \tag{38}$$

If we solve the Minimum Relative Entropy problem with this additional restriction on the form of q, then we will find that the spectral estimate $Q(\omega_n)$ is now a diagonal matrix with elements $Q_{sd}(\omega_n) = Q_{ds}(\omega_n) = 0$, and:

$$\begin{aligned} Q_{ss}(\omega_n) &= \frac{1}{\dfrac{1}{P_{ss}(\omega_n)} + \sum_k \lambda_k e^{i\omega_n \tau_k}} \\ Q_{dd}(\omega_n) &= \frac{1}{\dfrac{1}{P_{dd}(\omega_n)} + \sum_k \lambda_k e^{i\omega_n \tau_k}} \end{aligned} \tag{39}$$

The Lagrange multipliers λ_k are chosen so that the sum of these spectra, $Q_{yy}(\omega_n) = Q_{ss}(\omega_n) + Q_{dd}(\omega_n)$ has the correct correlations. These formulas are precisely the results given by Johnson and Shore [5].

In addition to the independence constraint, in some cases we may believe that our noise model is quite accurate, and may wish to constrain the noise density to exactly match our *a priori* noise model:

$$q(\delta_0, \ldots, \delta_{N-1}) = p(\delta_0, \ldots, \delta_{N-1}) \tag{40}$$

With this additional constraint, the Minimum Relative Entropy solution now gives $Q_{dd}(\omega_n) = P_{dd}(\omega_n)$, $Q_{sd}(\omega_n) = Q_{ds}(\omega_n) = 0$, and:

$$Q_{ss}(\omega_n) = \frac{1}{\dfrac{1}{P_{ss}(\omega_n)} + \sum_k \lambda_k e^{i\omega_n \tau_k}} \tag{41}$$

where the Lagrange multipliers λ_k are again chosen so that $Q_{yy}(\omega_n) = Q_{ss}(\omega_n) + Q_{dd}(\omega_n)$ has the correct correlations. Since $Q_{dd}(\omega_n)$ is known to equal $P_{dd}(\omega_n)$, however, this is equivalent to choosing the λ_k so that:

$$\sum_n Q_{ss}(\omega_n)\, e^{i\omega_n \tau_k} = R_k - \sum_n P_{dd}(\omega_n)\, e^{i\omega_n \tau_k} \tag{42}$$

The right hand side is just the observed correlations minus the contribution due to the noise, and is therefore an estimate of the signal correlations. With the added assumption of independence and known noise spectrum, therefore, our Multichannel RESA method acts like a correlator-subtractor technique [7], subtracting the noise correlations from the observed correlations, and then fitting a standard RESA model to these estimated signal correlations.

IX. EXAMPLE

We define a pair of spectra, S_{dd} and S_{ss}, which we think of as a known "background" and an unknown "signal" component of a total spectrum. Both are symmetric and defined in the frequency band from $-\pi$ to π, though we plot only their positive-frequency parts. (The abscissas in the figures are the frequency in Hz, $\omega/2\pi$, ranging from 0 to 0.5.) S_{dd} is the sum of white noise with total power 5 and a peak at frequency $0.215\times 2\pi$ corresponding to a single sinusoid with total power 2. S_{ss} consists of a peak at frequency $0.165\times 2\pi$ corresponding to a sinusoid of total power 2. Figure 1 shows a discrete-frequency approximation to the sum $S_{ss} + S_{dd}$, using 100 equispaced frequencies. From the sum, six autocorrelations were computed exactly. S_{dd} itself was used as the initial estimate P_{dd} of S_{dd}. For P_{ss} we used a uniform (flat) spectrum with the same total power as P_{dd}. The two initial estimates are shown in Figure 2. Figures 3 and 4 show multisignal RESA final estimates Q_{dd} and Q_{ss} by the method of [5] —independence was assumed for the final joint probability densities of the two signals. Figures 5 and 6 show final spectrum estimates obtained by the present method from the same autocorrelation data and initial spectrum estimates. The initial cross-spectrum estimates were taken to be zero (P was diagonal). No such assumption was made for the final estimate, of course, and indeed the final cross-spectrum estimates (not shown) are non-zero. Figures 7 and 8 show final estimates for the total spectrum of signal plus noise by the methods of [5] and this paper, respectively. Figure 7 is just the sum of Figures 3 and 4. Figure 8 coincides with the single-signal RESA final estimate obtained when $P_{dd} + P_{ss}$ is used as the initial estimate; it is not the sum of Figures 5 and 6, since it includes contributions from the cross-spectrum estimates.

In the results for both methods, the signal peak shows up primarily in Q_{ss}, but some evidence of it is in Q_{dd} as well. Comparison of Figures 5 and 6 with Figures 3 and 4 shows that both final spectrum estimates by the present method are closer to the respective initial estimates than are the final estimates by the method of [5]

In view of the fact that the present method has the logically more satisfying derivation and is computationally cheaper, the comparison of Figure 6 with Figure 4 is somewhat disappointing; the signal peak shows up less strongly in Figure 6. It must be pointed out, however, that in this example the signal and noise are truly uncorrelated. Our technique does not use this information, and in fact estimates a non-zero cross-correlation between the signal and noise. The method in [5], however, uses this additional knowledge and therefore, in this case, is able to produce better estimates than our technique.

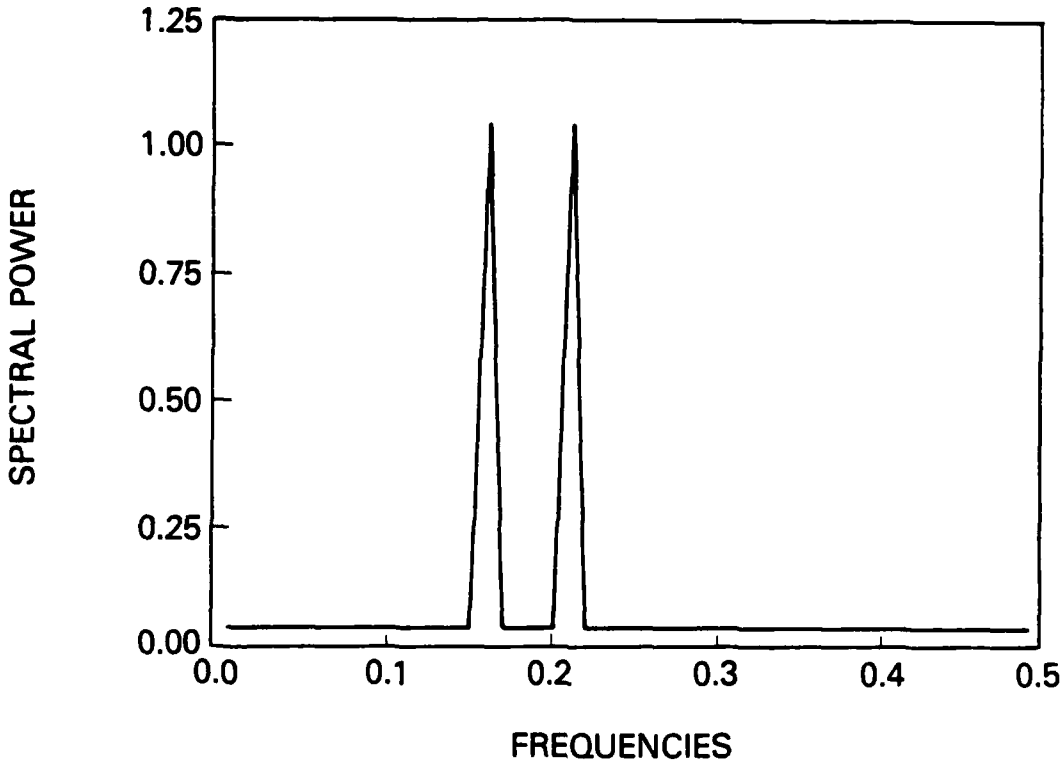


Fig. 1. Assumed total original spectrum for example.

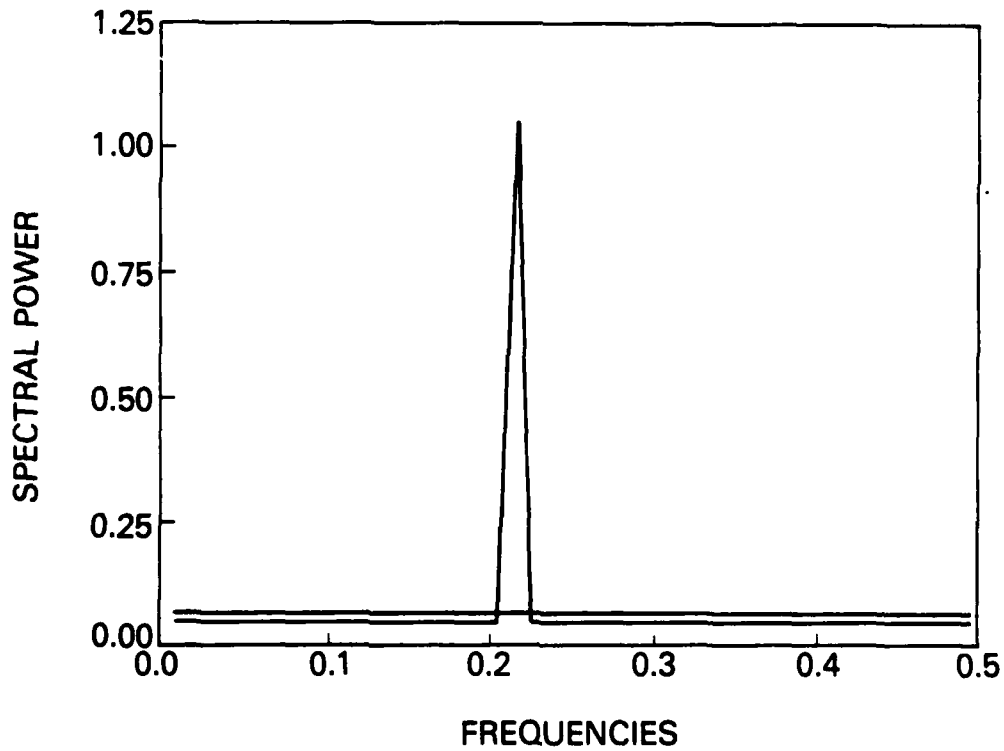


Fig. 2. Initial spectrum estimates P_{dd} (with peak) and P_{ee} (flat) for example.

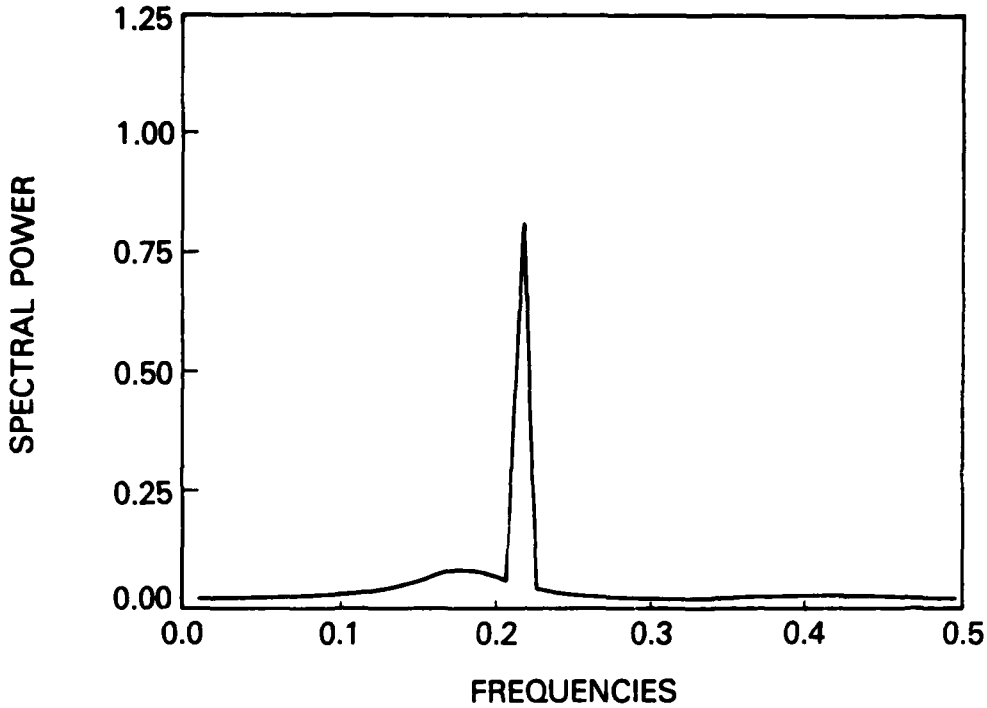


Fig. 3. Final spectrum estimate Q_{dd} by the method of [5].

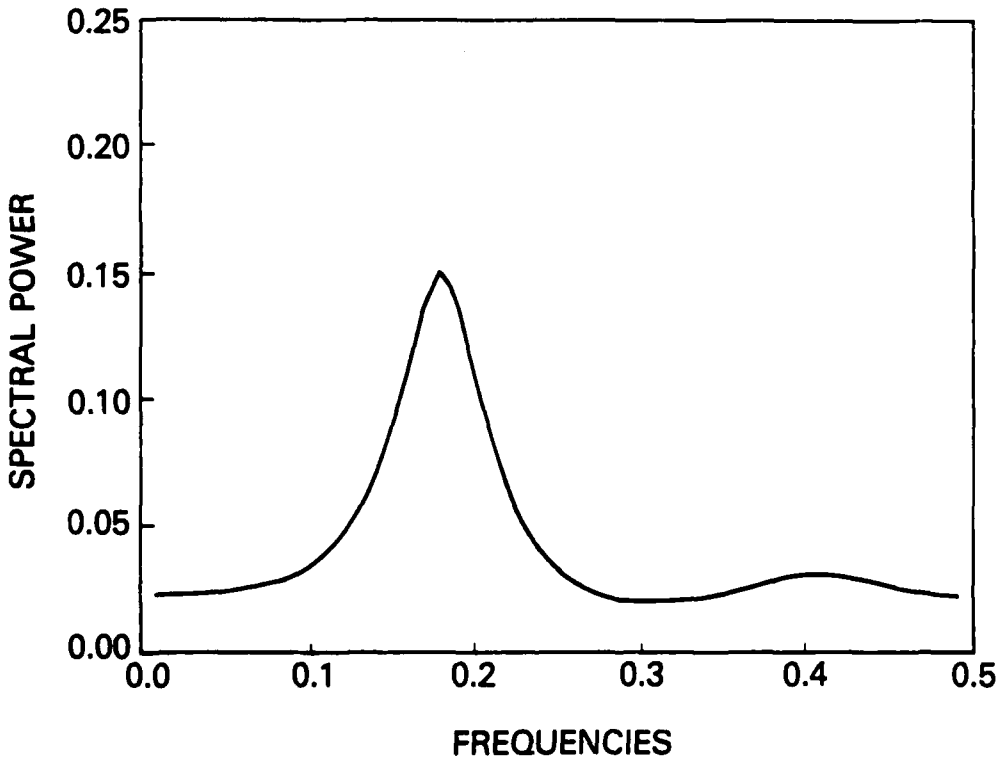


Fig. 4. Final spectrum estimate Q_{ee} by the method of [5].

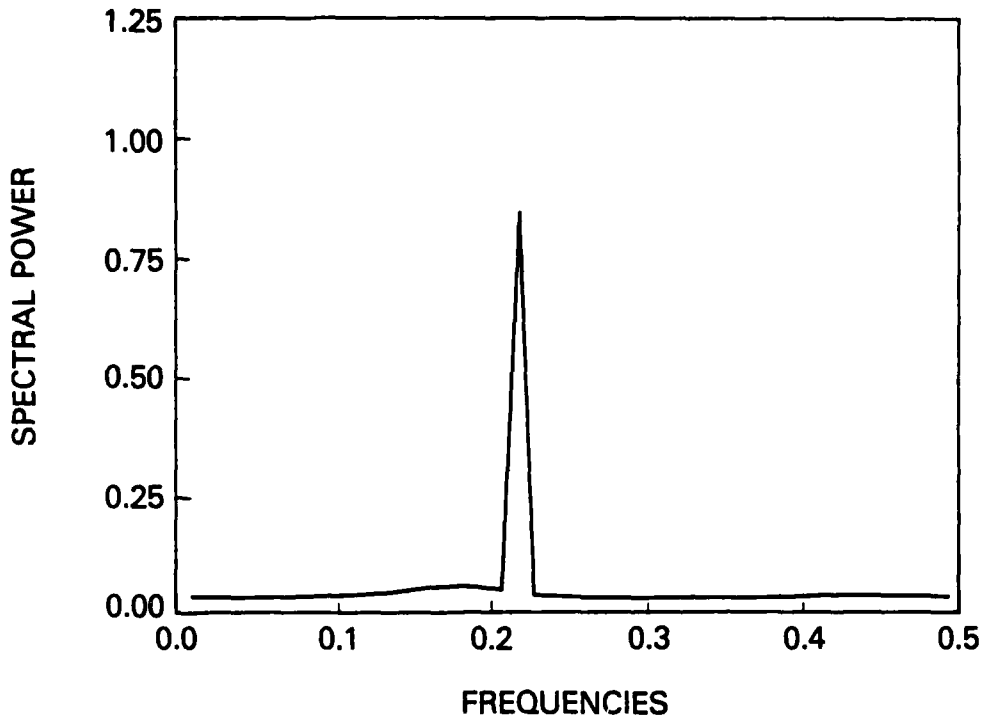


Fig. 5. Final spectrum estimate Q_{dd} by the method of this paper.

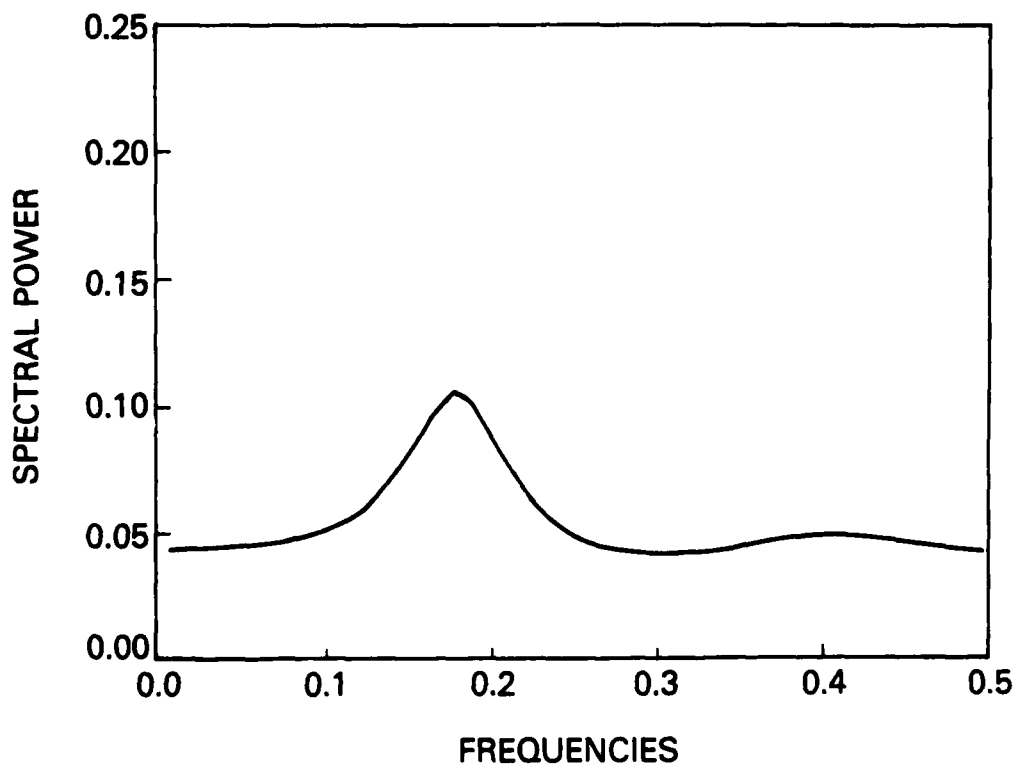


Fig. 6. Final spectrum estimate Q_{ee} by the method of this paper.

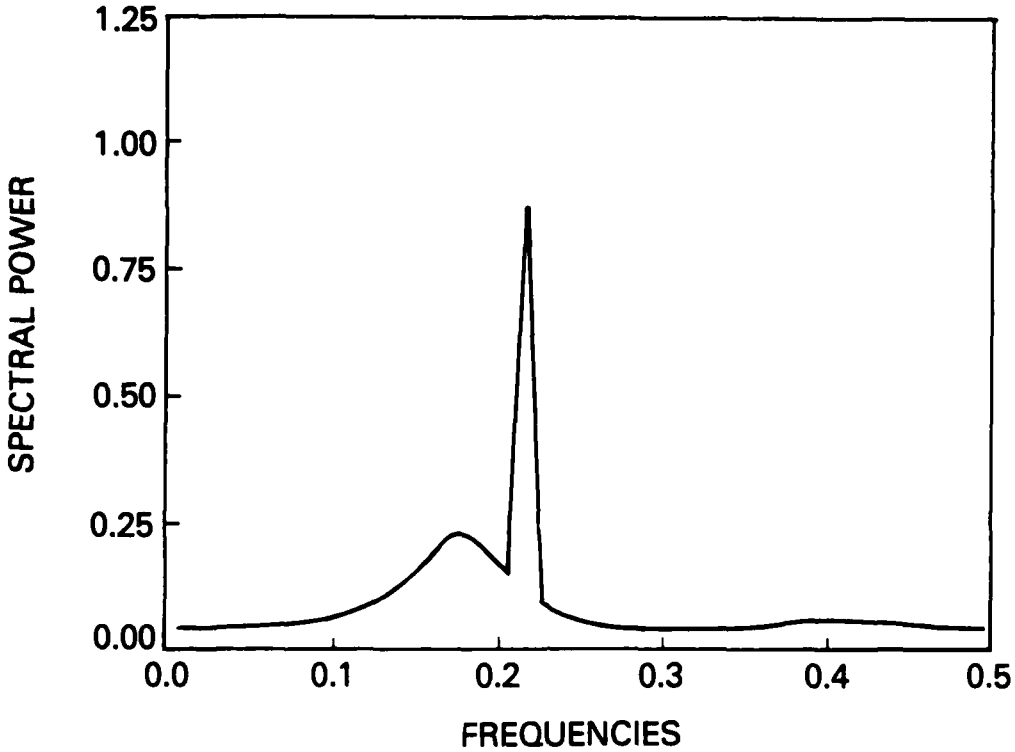


Fig. 7. Final estimate $Q_{dd} + Q_{ss}$ of total spectrum by method of [5].

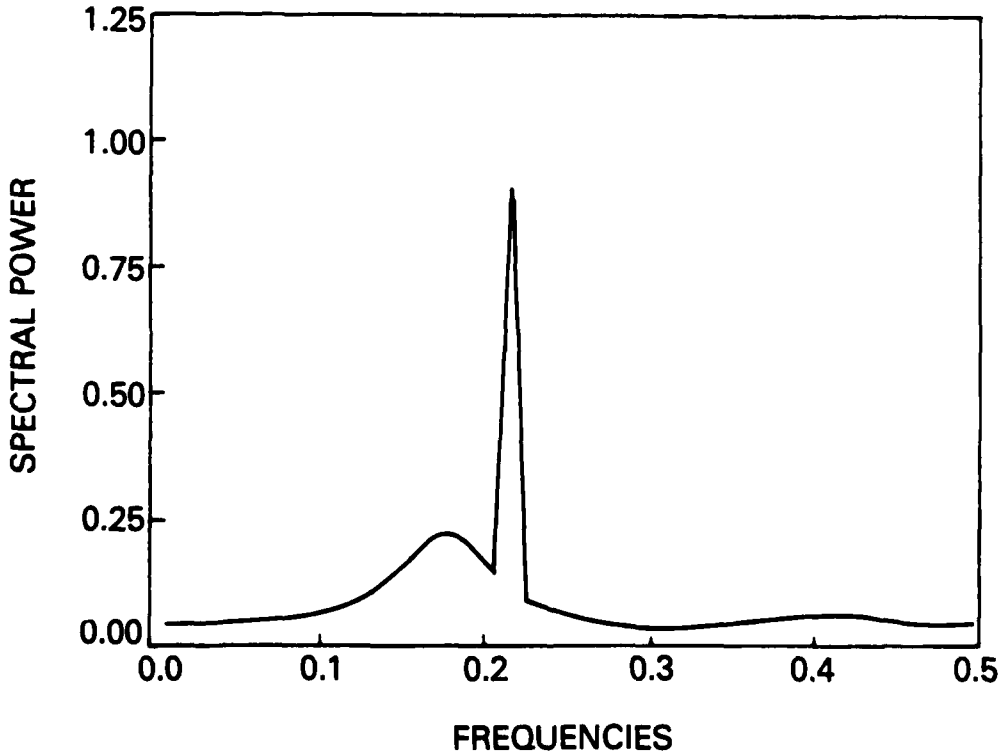


Fig. 8. Final estimate Q_{yy} of total spectrum by method of this paper.

X. DISCUSSION

In this paper we have derived a Multichannel Relative-Entropy Spectrum Analysis method that estimates the power spectra and cross-spectra of several signals, given an initial estimate of the spectral density matrix and given new information in the form of correlation values for linear combinations of the channels. Both this method and the multisignal method of [5] will estimate the power spectra of a signal and noise when prior information is available in the form of an initial estimate of each spectrum and given selected correlations of the signal plus noise. The present method can accept more general forms of correlation data and also produces cross-spectrum estimates, which are implicitly assumed to be zero in [5]. Even when the only correlation data are for the signal plus noise, and cross-spectrum estimates are not desired, there is a persuasive argument for preferring the present method to that of [5] —if the discrepancy between the given correlation values and those computed from the initial estimates can be accounted for in part by correlations between the signal and noise, then the correlation data should be regarded as evidence for such correlations, and correlated final estimates should be produced.

Estimates by the present method are considerably more economical to compute than estimates by the method of [5]. The algorithm first fits a smooth model power spectrum to the noisy signal using the given correlations. The available prior information is then used to linearly filter this spectrum estimate in order to obtain separate estimates for the signal and the noise. This allocation formula is virtually identical to that used by the usual Bayesian formula in which the signal and noise power spectra are estimated from the observed signal plus noise spectrum. The difference between the Multichannel RESA and Bayesian methods is that the relative-entropy technique starts by fitting a smooth power spectrum model to the observed correlations, while the Bayesian approach starts with the directly observed power spectrum. This Multichannel Relative-Entropy technique thus provides a smooth model fitting spectrum analysis procedure that is closely analogous to the Bayesian approach. If p uniformly spaced correlations are given, and if the prior information suggests that the power spectrum of the signal plus noise is autoregressive of order at most p, then the step of fitting a smooth model spectrum to the noisy signal is identical to using a standard MESA algorithm to fit a smooth autoregressive model to the given correlations. We concluded by searching for a way to treat the case where no prior information about the signal is known. Unfortunately, we were not able to find a theoretically sound approach with desirable characteristics.

In general, the method presented in this paper yields final spectral estimates that are closer to the initial estimates than those of [5]. This is not surprising. Our method starts with an initial estimate of the signal and noise spectra, and uses correlations of the signal plus noise to get better power spectra estimates. The method in [5] uses the same information, but also assumes that the signal and noise are uncorrelated. This additional knowledge further restricts the constraint space $\mathbf{Q}$ in which the probability density is known to lie, effectively leaving less unknown aspects of the density to estimate, and thus improving the final spectra. In general, the resulting spectral estimate will have higher relative entropy than the solution from our method, and will thus be "farther" from the initial density p than the solution from our method.

Our estimate of $Q_{yy}(\omega_k)$ can be improved by observing more and more correlations of the signal plus noise. Regardless of how much data is gathered, however, our method relies exclusively on the initial estimate of the signal and noise spectra and cross spectra to allocate $Q_{yy}(\omega_k)$ between the signal, noise, and cross terms. The fundamental difficulty is that observing correlations of the signal plus noise gives no insight into how this observation energy should be partitioned between the signal and the noise. Achieving accurate estimation of the signal and noise spectra separately requires a different type of observation data. Learning that the signal and noise are uncorrelated as in [5], for example, will improve our spectral estimates, as would learning the exact noise power spectrum. The best solution, of course, would be to use an accurate model of the signal and noise processes, or to directly observe the signal and/or noise correlations.

REFERENCES

1. J. E. Shore, "Minimum cross-entropy spectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing* **ASSP-29**, pp. 230-237 (Apr. 1981).
2. J. P. Burg, "Maximum entropy spectral analysis," presented at the 37th Annual Meeting Soc. of Exploration Geophysicists, Oklahoma City, Okla. (1967).
3. J. P. Burg, "Maximum Entropy Spectral Analysis," Ph.D. Dissertation, Stanford University, Stanford, California (1975). (University Microfilms No. AAD75-25,499)
4. R. W. Johnson and B. R. Musicus, "Multichannel relative-entropy spectrum analysis," pp. 13.6.1–4 in *Proc. ICASSP 84*, San Diego, Ca. (March 1984). 1984 IEEE International Conference on Acoustics, Speech, and Signal Processing. IEEE 84CH1945-5.
5. R. W. Johnson and J. E. Shore, "Minimum-cross-entropy spectral analysis of multiple signals," *IEEE Trans. Acoust., Speech, Signal Processing* **ASSP-31**, pp. 574-582 (June 1983). Also see NRL MR 4492 (AD-A097531)
6. R. W. Johnson, J. E. Shore, and J. P. Burg, "Multisignal minimum-cross-entropy spectrum analysis with weighted initial estimates," *IEEE Trans. Acoustics, Speech, Signal Processing* **ASSP-32**, pp. 531-539 (June, 1984).
7. Jae S. Lim, "Evaluation of a Correlation-Subtraction Method for Enhancing Speech Degraded by Additive White Noise," *IEEE Trans. Acoust., Speech, and Signal Processing* **ASSP-26**, pp. 471-472 (Oct 1978).
8. Bruce R. Musicus, *An Iterative Technique for Maximum Likelihood Parameter Estimation on Noisy Data*, S.M. Thesis, M.I.T., Cambridge, Mass. (Feb 1979).
9. Bruce R. Musicus and Jae S. Lim, "Maximum Likelihood Parameter Estimation of Noisy Data," in *Proc. ICASSP 79*, Washington, DC (April 2–4, 1979). 1979 IEEE International Conference on Acoustics, Speech, and Signal Processing (AD-A069771)
10. J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory* **IT-26**, pp. 26-37 (Jan. 1980). (See also comments and corrections in *IEEE Trans Inform. Theory* **IT-29**, Nov. 1983, p. 942)
11. I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *Ann. Prob.* **3**, pp. 146-158 (1975).
12. S. Kullback, *Information Theory and Statistics*, Dover, New York (1968). (Wiley, New York, 1959)
13. R. W. Johnson, "Axiomatic characterization of the directed divergences and their linear combinations," *IEEE Trans. Inform. Theory* **IT-25**, pp. 709-716 (Nov. 1979).
14. E. T. Jaynes, "Prior probabilities," *IEEE Trans. Systems Science and Cybernetics* **SSC-4**, pp. 227-241 (1968).
15. A. Hobson and B.-K. Cheng, "A comparison of the Shannon and Kullback information measures," *J. Stat. Phys.* **7**, pp. 301-310 (1973).
16. I. J. Good, "Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables," *Ann. Math. Stat.* **34**, pp. 911-934 (1963).
17. I. J. Good, *Probability and the Weighing of Evidence*, Griffin, London (1950).
18. M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco (1964).
19. W. M. Elsasser, "On quantum measurements and the role of the uncertainty relations in statistical mechanics," *Phys. Rev.* **52**, pp. 987-999 (Nov. 1937).
20. E. T. Jaynes, "Information theory and statistical mechanics I," *Phys. Rev.* **106**, pp. 620-630 (1957).

21. E. T. Jaynes, "Information theory and statistical mechanics II," *Phys. Rev.* **108**, pp. 171-190 (1957).

22. 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. Inform. Theory* **IT-27**, pp. 708-721 (Nov. 1981).

23. R. W. Johnson, "Algorithms for single-signal and multisignal minimum-cross-entropy spectral analysis," NRL Report 8667, Naval Research Laboratory, Washington, D.C. (August 1983). (AD-A132 400)

24. A. M. Yaglom, *An Introduction to the Theory of Stationary Random Functions*, Prentice-Hall, Englewood Cliffs, NJ (1962).

25. R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra from the Point of View of Communications Engineering*, Dover, New York (1959).

26. H. L. Van Trees, *Detection, Estimation and Modulation Theory—Part 1*, Wiley, New York, N.Y. (1968).

27. J. E. Shore, "Minimum Cross-Entropy Spectral Analysis," NRL Memorandum Report 3921, Naval Research Laboratory, Washington, DC 20375 (Jan. 1979). (AD-A064183)

28. J. E. Shore and R. W. Johnson, "Properties of cross-entropy minimization," *IEEE Trans. Inform. Theory* **IT-27**, pp. 472-482 (July 1981).

29. F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," *Proc. 6th International Congress on Acoustics*, pp. C-17-20 (August 1968).

30. J. Makhoul, "Linear Prediction: A Tutorial Review," *Proc. IEEE* **63**(4), pp. 561-580 (April 1975). reprinted in Modern Spectrum Analysis, ed. Donald G. Childers, IEEE Press, New York 1978

31. N. Levinson, "The Wiener RMS (root mean square) error criterion in filter design and prediction," *J. Math Phys.* **25**, pp. 261-278 (1947).

32. J. Durbin, "The fitting of time series models," *Rev. Inst. Int. de Stat.* **28**, pp. 233-244 (1960).

END
FILMED

5-86

DTIC