

A Separable Cross-Entropy Approach to Power Spectral Estimation

CHENG-YUAN LIOU, MEMBER, IEEE, AND BRUCE R. MUSICUS, MEMBER, IEEE

Abstract—We present a new approach for power spectrum estimation based on a separable cross-entropy modeling procedure. We start with a model of a multichannel, multidimensional stationary Gaussian random process which is sampled on a nonuniform grid. An approximate separable model is then fit to this, in which selected frequency samples of the process are modeled as independent random variables. Two cross-entropy-like criteria are used to select optimal separable approximations. One of our methods yields a spectral estimation algorithm which is a generalized version of Capon's MLM method, and the other is similar to classical windowing methods. We conclude with a discussion of different strategies for designing bandpass filters for use with this method.

I. INTRODUCTION

THERE are a large number of applications in which it is necessary to estimate the power spectrum of a stationary random process given samples of the covariance kernel. The simplest approach is to form the periodogram by computing the magnitude squared of the Fourier Transform of a set of windowed data samples. Unfortunately, the periodogram yields inconsistent and biased spectral estimates. Bias is introduced by the data windowing process, and the variance of each frequency sample remains constant regardless of the number of data samples [1]. Blackman-Tukey [2] suggested a simple method for improving the variance of the estimate by windowing the correlations of the data before Fourier Transforming. A similar effect was achieved by Welch [3] and others by averaging multiple periodograms formed from overlapping short frames of data. In both cases, the methods trade off resolution for decreased variance by using a window to average the power across a number of frequency components. The shape and width of the window controls the tradeoff between resolution and variance.

More modern methods use precise objective functions or models of the underlying random process to achieve higher resolution, smoother estimates, and lower variance. Burg's Maximum Entropy method (MEM) [4]

chooses a power spectrum estimate to maximize the entropy of the spectrum subject to the constraint that the spectrum must match the first few correlations of the data. Given one-dimensional, uniformly sampled correlations, the resulting spectrum is all-pole, and can be calculated using various linear prediction methods. An alternative approach is Capon's Maximum Likelihood Method (MLM) which designs an optimal Finite Impulse Response (FIR) bandpass filter at each frequency to pass that frequency unmodified but reject as much other signal energy as possible. The output power of the filter is used as the estimate of the power spectral density at that frequency. This spectral estimate can be shown to have the form of an all-pole model, although the coefficients are different than those of MEM. Numerous other interpretations of MLM exist in the literature. Marzetta [5], for example, shows the MLM gives the upper bound on the amplitude of each frequency component given the known correlations. MLM is easy to compute even for nonuniformly sampled or multidimensional data, while MEM is quite difficult for these cases. The resolution of MLM, however, tends to be lower than that of MEM [6].

In this paper, we consider a very different approach to the problem of power spectrum estimation. We start by noting that a stationary Gaussian random process can be loosely characterized as having a Fourier Transform which is a Gaussian white noise process. As the amount of data grows asymptotically large, samples of the transform approach independent zero-mean Gaussian random variables with variance equal to the power spectrum. Our new approach to spectral estimation starts by fitting an approximate model to the given data, in which we assume that a selected set of up to N frequency samples are independent random variables. The best such separable model is chosen by minimizing one of two discriminant or cross-entropy functions. The approach used is precisely that developed by Musicus [7], but applied to the spectral estimation problem. One of our methods yields a generalized form of the MLM method, and the other is similar to a classical windowing approach. Examples are shown demonstrating the performance of these techniques.

Our presentation is somewhat indirect. We will start by discussing the use of cross-entropy to build model approximations in which variables are forced to be independent. After discussing various special cases of the results, we will consider the power spectrum estimation problem.

Manuscript received May 6, 1985; revised April 4, 1988. The work of C.-Y. Liou was supported in part by the Technology Assessment Research Program of the Minerals Management Service. The work of B. R. Musicus was 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 ECS-8407285.

C.-Y. Liou was with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA. He is now with the National Taiwan University, Taiwan.

B. R. Musicus is with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139.
IEEE Log Number 8931691.

0096-3518/90/0100-0105\$01.00 © 1990 IEEE

II. GAUSSIAN ESTIMATION PROBLEM

Suppose we are given a finite set of L jointly Gaussian real or complex valued random variables, $\mathbf{x}_1, \dots, \mathbf{x}_L$, each a vector of length N_i

$$E[\mathbf{x}_i] = \mathbf{m}_{x_i}$$

$$\text{Cov}[\mathbf{x}_i, \mathbf{x}_j] = E[(\mathbf{x}_i - \mathbf{m}_{x_i})(\mathbf{x}_j - \mathbf{m}_{x_j})^H] = R_{i,j} \quad (2.1)$$

where $E[\cdot]$ is the expectation operator, $\text{Cov}[\cdot]$ is the covariance, and $\mathbf{x}^H = \mathbf{x}^{T*}$ is the complex conjugate transpose. $R_{i,j}$ is an $N_i \times N_j$ matrix. It will be convenient to define variables

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_L \end{pmatrix} \quad R_x = \begin{pmatrix} R_{1,1} & \cdots & R_{1,L} \\ \vdots & \cdots & \vdots \\ R_{L,1} & \cdots & R_{L,L} \end{pmatrix}$$

$$\mathbf{m}_x = \begin{pmatrix} \mathbf{m}_{x_1} \\ \vdots \\ \mathbf{m}_{x_L} \end{pmatrix} \quad (2.2)$$

where \mathbf{x} and \mathbf{m}_x have length $N = \sum N_i$, and R_x is an $N \times N$ conjugate symmetric positive-definite covariance matrix. Then the probability density of \mathbf{x} is Gaussian with mean \mathbf{m}_x and covariance R_x

$$p(\mathbf{x}_1, \dots, \mathbf{x}_L) \equiv p(\mathbf{x}) = N(\mathbf{m}_x, R_x). \quad (2.3)$$

Let $\mathbf{y}_1, \dots, \mathbf{y}_P$ be a set of P vectors, each of length M_i , which are formed from linear combinations of the \mathbf{x}_j

$$\mathbf{y}_i = \sum_{j=1}^L T_{i,j} \mathbf{x}_j \quad \text{for } i = 1, \dots, P. \quad (2.4)$$

The \mathbf{y}_i are jointly Gaussian. Let

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_P \end{pmatrix} \quad T = \begin{pmatrix} T_{1,1} & \cdots & T_{1,L} \\ \vdots & \cdots & \vdots \\ T_{P,1} & \cdots & T_{P,L} \end{pmatrix} \quad (2.5)$$

where \mathbf{y} has length $M = \sum M_i$, and T has dimensions $M \times N$. Assume that $M \leq N$, and that T has rank M . Then \mathbf{y} is Gaussian with mean \mathbf{m}_y and positive definite covariance R_y

$$p(\mathbf{y}_1, \dots, \mathbf{y}_P) \equiv p(\mathbf{y}) = N(\mathbf{m}_y, R_y)$$

where

$$\mathbf{m}_y = T\mathbf{m}_x$$

$$R_y = TR_x T^H. \quad (2.6)$$

It will be convenient to partition R_y into blocks $[R_y]_{i,j}$ of size $M_i \times M_j$, corresponding to the covariance of \mathbf{y}_i and \mathbf{y}_j . Similarly, we will partition the inverse covariance matrix R_y^{-1} into blocks $[R_y^{-1}]_{i,j}$.

III. SEPARABLE CROSS-ENTROPY MODELING

In general, the components $\{\mathbf{y}_i\}$ will be correlated Gaussian random variables. Suppose, however, that the components $\{\mathbf{y}_i\}$ are "close" to being independent in the sense that the off-diagonal blocks of the R_y covariance matrix are "small" compared to the diagonal blocks. It might be computationally and analytically convenient in this case to find an approximation to the Gaussian model for \mathbf{y} in which the components $\{\mathbf{y}_i\}$ are treated as independent random variables, but in which the shape of the probability density is otherwise about the same. More formally, we will try to find a separable probability density model $q(\mathbf{y})$ in which the components $\{\mathbf{y}_i\}$ are independent random variables

$$q(\mathbf{y}) = q_1(\mathbf{y}_1)q_2(\mathbf{y}_2) \cdots q_P(\mathbf{y}_P). \quad (3.1)$$

Our goal is to find a separable $q(\mathbf{y})$ which optimally approximates the original $p(\mathbf{y})$ in some manner.

Kullback's two discriminant functions [8] are good criteria for measuring the goodness-of-fit between p and q . The first discriminant function is equivalent to the cross-entropy criterion suggested by Shore and Johnson [9] and developed for this class of problems by Musicus [7]. The second function is just the reverse of the first. We will develop these two approaches in the next two sections, and then compare the results.

A. Cross-Entropy (CE) Method

Shore and Johnson considered the following problem. Suppose that we are given an *a priori* estimate $p(\cdot)$ of a probability density for some unknown \mathbf{a} , but that we are not very certain of its validity. Suppose that new observation data arrive which suggest that this estimated density is incorrect, and that the actual density must belong to some set \mathcal{Q} . Shore and Johnson started with four invariance and consistency axioms which, they argued, any reasonable estimation procedure must obey. These insist, for example, that the same answer must result regardless of the coordinate system chosen for the unknown. Shore and Johnson were then able to prove that *any* estimation procedure which satisfies these axioms must yield the same estimate as that found by minimizing the cross-entropy $H(q, p)$ over the set $q \in \mathcal{Q}$

$$\hat{q}_{CE}(\cdot) \leftarrow \min_{q \in \mathcal{Q}} H(q, p) \quad (3.2)$$

where

$$H(q, p) = \int q(\mathbf{a}) \log \frac{q(\mathbf{a})}{p(\mathbf{a})} d\mathbf{a}.$$

We will call this the "Cross-Entropy" Method (CE), and will denote the solution by \hat{q}_{CE} . For our problem, \mathcal{Q} is the set of separable densities, and so

$$\hat{q}_{CE} \leftarrow \min_{q_1, \dots, q_P} \int q_1(\mathbf{y}_1) \cdots q_P(\mathbf{y}_P) \cdot \log \frac{q_1(\mathbf{y}_1) \cdots q_P(\mathbf{y}_P)}{p(\mathbf{y}_1, \dots, \mathbf{y}_P)} d\mathbf{y}_1 \cdots d\mathbf{y}_P. \quad (3.3)$$

We show in Appendix A that this minimization problem has a unique solution which has the form

$$\text{CE: } \hat{q}_{CE}(\mathbf{y}) = \hat{q}_1(\mathbf{y}_1) \cdots \hat{q}_P(\mathbf{y}_P)$$

where

$$\hat{q}_i(\mathbf{y}_i) = N(\mathbf{m}_{y_i}, [\mathbf{R}_y^{-1}]_{i,i}^{-1}). \quad (3.4)$$

This is exactly the conditional density of \mathbf{y}_i given that all the other components \mathbf{y}_j are equal to their *a priori* mean \mathbf{m}_{y_j}

$$\hat{q}_i(\mathbf{y}_i) = p(\mathbf{y}_i | \{ \mathbf{y}_j = \mathbf{m}_{y_j} \text{ for } j \neq i \}). \quad (3.5)$$

The value of the cross-entropy at this solution is

$$H(\hat{q}_{CE}, p) = \xi \log \frac{\prod_{i=1}^P |[R_y^{-1}]_{i,i}|}{|R_y^{-1}|} \quad (3.6)$$

where $|R_y^{-1}|$ is the determinant of R_y^{-1} . Note that $H(q, p)$ is always positive, and attains its absolute minimum value of 0 if and only if $q = p$. Thus, the minimum cross-entropy value will be zero if and only if the original probability density is already separable. If the minimum $H(\hat{q}_{CE}, p)$ is close to 0, then \hat{q}_{CE} is a good separable approximation to p . If the minimum $H(\hat{q}_{CE}, p)$ is large, however, then the \mathbf{y}_i are strongly correlated and the best separable approximation is still quite poor.

B. Reversed Cross-Entropy Approach

A related approach which gives quite a different result is to reverse the order of the arguments in the cross-entropy function, choosing \hat{q} to minimize $H(p, q)$

$$\begin{aligned} \hat{q}_{RCE} \leftarrow \min_{q_1, \dots, q_P} \int p(\mathbf{y}_1 \cdots \mathbf{y}_P) \\ \cdot \log \frac{p(\mathbf{y}_1 \cdots \mathbf{y}_P)}{q_1(\mathbf{y}_1) \cdots q_P(\mathbf{y}_P)} d\mathbf{y}_1 \cdots d\mathbf{y}_P. \end{aligned} \quad (3.7)$$

We will call this the ‘‘Reverse Cross-Entropy’’ (RCE) Method, and will denote the solution as \hat{q}_{RCE} . This minimization is considerably easier to solve than that in (3.3). We can write (3.7) in the form

$$\begin{aligned} \hat{q}_{RCE} \leftarrow \min_{q_1, \dots, q_P} \int p(\mathbf{y}_1 \cdots \mathbf{y}_P) \\ \cdot \log \frac{p(\mathbf{y}_1 \cdots \mathbf{y}_P)}{p(\mathbf{y}_1) \cdots p(\mathbf{y}_P)} d\mathbf{y}_1 \cdots d\mathbf{y}_P \\ + \int p(\mathbf{y}_1, \cdots, \mathbf{y}_P) \log \frac{p(\mathbf{y}_1) \cdots p(\mathbf{y}_P)}{q_1(\mathbf{y}_1) \cdots q_P(\mathbf{y}_P)} \\ \cdot d\mathbf{y}_1 \cdots d\mathbf{y}_P \end{aligned} \quad (3.8)$$

where $p(\mathbf{y}_i)$ is the marginal density

$$\begin{aligned} p(\mathbf{y}_i) &= \int p(\mathbf{y}_1, \cdots, \mathbf{y}_P) d\mathbf{y}_1 \cdots d\mathbf{y}_{i-1} d\mathbf{y}_{i+1} \cdots d\mathbf{y}_P \\ &= N(\mathbf{m}_{y_i}, [R_y]_{i,i}). \end{aligned} \quad (3.9)$$

Only the second term in (3.8) depends on q . This term can be further decomposed, reducing the minimization problem to the form:

$$\hat{q}_{RCE} \leftarrow \min_{q_1, \dots, q_P} \sum_{i=1}^P \int p(\mathbf{y}_i) \log \frac{p(\mathbf{y}_i)}{q_i(\mathbf{y}_i)} d\mathbf{y}_i \quad (3.10)$$

where the minimization is subject to the constraint $\int q_i(\mathbf{y}_i) d\mathbf{y}_i = 1$. Each term in (3.10) is strictly convex in q_i , and thus the unique minimum occurs at

$$\text{RCE: } \hat{q}_{RCE}(\mathbf{y}) = \hat{q}_1(\mathbf{y}_1) \cdots \hat{q}_P(\mathbf{y}_P)$$

where

$$\begin{aligned} \hat{q}_i(\mathbf{y}_i) &= p(\mathbf{y}_i) = N(\mathbf{m}_{y_i}, [R_y]_{i,i}) \\ &\text{for } i = 1, \cdots, P. \end{aligned} \quad (3.11)$$

With this objective function, therefore, the separable function simply models each component as having the same mean and covariance as the original marginal density.

The value of the reverse cross-entropy function (3.7) at this solution is

$$H(p, \hat{q}_{RCE}) = \xi \log \frac{\prod_{i=1}^P |[R_y]_{i,i}|}{|R_y|}. \quad (3.12)$$

Once again, this will be zero if and only if p is already separable. If the value is close to 0, then \hat{q}_{RCE} is a good separable approximation to p . Otherwise, the \mathbf{y}_i are not modeled well as independent random variables.

C. Comparison of the Techniques

It is interesting to compare the behavior of the separable approximations for these two methods. If the original density p were separable,

$$p(\mathbf{y}) = p(\mathbf{y}_1) \cdots p(\mathbf{y}_P), \quad (3.13)$$

then both methods would simply set $\hat{q}_{CE} = \hat{q}_{RCE} = p$, and the value of the cross-entropies would both be zero, $H(\hat{q}_{CE}, p) = H(p, \hat{q}_{RCE}) = 0$. In general, however, when p is not separable, then the two methods will give somewhat different answers, and the cross-entropy values will be somewhat different. The CE method (3.5) sets each component of the density \hat{q}_i to the conditional probability density of \mathbf{y}_i given that the remaining components are set equal to their *a priori* means. The RCE method (3.11) sets each component density to the marginal density of \mathbf{y}_i . In general, with correlated \mathbf{y}_i , the CE density estimates will be ‘‘narrower’’ than the RCE estimates. In fact, it is straightforward to show through matrix manipulation that the CE covariance estimate is always smaller than the RCE covariance estimate

$$\text{Cov}(\mathbf{y}_i | \hat{q}_{CE}) = [R_y^{-1}]_{i,i}^{-1} \leq \text{Cov}(\mathbf{y}_i | \hat{q}_{RCE}) = [R_y]_{i,i}. \quad (3.14)$$

Both methods have a similar structure; the primary difference is that CE bases its estimates on diagonal blocks

of the inverse covariance matrix, $[R_y^{-1}]_{i,i}^{-1}$, while RCE uses diagonal blocks of the covariance matrix, $[R_y]_{i,i}$. This suggests a simple duality between these methods: the \hat{q}_{CE} and \hat{q}_{RCE} approximations to $p(\mathbf{y}) = N(\mathbf{m}_y, R_y)$ are equal respectively to the \hat{q}_{RCE} and \hat{q}_{CE} approximations to $\tilde{p}(\mathbf{y}) = N(\mathbf{m}_y, R_y^{-1})$. Also, $H(\hat{q}_{CE}, p) = H(\tilde{p}, \hat{q}_{CE})$ and $H(p, \hat{q}_{RCE}) = H(\hat{q}_{RCE}, \tilde{p})$.

This duality also implies that neither method can be better than the other in all cases. For two unknowns, example 1 below shows that both methods always achieve equal cross-entropy, $H(\hat{q}_{CE}, p) = H(p, \hat{q}_{RCE})$. For 3 or more unknowns, however, examples can be constructed in which the difference between $H(\hat{q}_{CE}, p)$ and $H(p, \hat{q}_{RCE})$ is arbitrarily large (see Appendix B).

D. Example 1—Two Variables

Let y_1, y_2 be two jointly Gaussian real-valued random variables with means $E[y_i] = m_{y_i}$ and covariances $\text{Cov}(y_i, y_j) = R_{ij}$. We will try approximating this model with a separable model $q_1(y_1)q_2(y_2)$. The CE method gives

$$\hat{q}_{CE}(y_1, y_2) = \hat{q}_1(y_1)\hat{q}_2(y_2) \quad (3.15)$$

where

$$\begin{aligned} \hat{q}_1(y_1) &= N(m_{y_1}, R_{11} - R_{12}R_{22}^{-1}R_{21}) \\ \hat{q}_2(y_2) &= N(m_{y_2}, R_{22} - R_{21}R_{11}^{-1}R_{12}) \end{aligned} \quad (3.16)$$

and the value of the cross-entropy function is

$$H(\hat{q}_{CE}, p) = \frac{1}{2} \log \frac{|R_{11}|}{|R_{11} - R_{12}R_{22}^{-1}R_{21}|}. \quad (3.17)$$

The RCE method gives

$$\hat{q}_{RCE}(y_1, y_2) = \hat{q}_1(y_1)\hat{q}_2(y_2) \quad (3.18)$$

where

$$\begin{aligned} \hat{q}_1(y_1) &= N(m_{y_1}, R_{11}) \\ \hat{q}_2(y_2) &= N(m_{y_2}, R_{22}) \end{aligned} \quad (3.19)$$

and its cross-entropy is exactly equal to that of CE

$$H(\hat{q}_{CE}, p) = H(p, \hat{q}_{RCE}). \quad (3.20)$$

Note that the covariance of the CE estimates is smaller than the covariance of the RCE estimates. Also, note that if the unknowns are initially almost uncorrelated, then $R_{12} = R_{21}^H \approx 0$. In this case, $H(\hat{q}_{CE}, p) = H(p, \hat{q}_{RCE}) \approx 0$, and both separable models will be good fits to the original density. Notice, in fact, that both methods also give virtually the same answer, since $R_{11} - R_{12}R_{22}^{-1}R_{21} \approx R_{11}$. On the other hand, if the variables are highly correlated, then $R_{11} - R_{12}R_{22}^{-1}R_{21} \approx 0$, the cross-entropy values H are large, and both separable approximations are equally poor. Note that in this case, CE assigns nearly zero variance to the variables, while RCE assigns variance R_{ij} .

E. Example 2—Invertible Transformations

Suppose $M = N$ and T is a square and invertible transform. Then

$$R_y = TR_x T^H \quad (3.21)$$

$$R_y^{-1} = T^{-H} R_x^{-1} T^{-1}. \quad (3.22)$$

Let $[T]_{i,*}$ be the i th block of rows of T

$$T = \begin{pmatrix} [T]_{1,*} \\ \dots \dots \dots \\ [T]_{p,*} \end{pmatrix} \quad (3.23)$$

so that $[T]_{i,*}$ is an $N_i \times N$ matrix, and

$$y_i = [T]_{i,*} \mathbf{x}. \quad (3.24)$$

Similarly, let $[T^{-1}]_{*,i}$ be the i th block of columns of T^{-1}

$$T^{-1} = ([T^{-1}]_{*,1} \dots [T^{-1}]_{*,p}) \quad (3.25)$$

so that

$$\mathbf{x} = T^{-1} \mathbf{y} = \sum_{i=1}^p [T^{-1}]_{*,i} y_i. \quad (3.26)$$

$TT^{-1} = I$ implies that

$$[T]_{i,*} [T^{-1}]_{*,j} = \delta_{i,j} I \quad (3.27)$$

where $\delta_{i,j}$ is the Kronecker delta function, $\delta_{i,j} = 1$ if $i = j$, $= 0$ else. Equation (3.27) implies that $[T]_{i,*}$ is orthogonal to all $[T^{-1}]_{*,j}$ for $i \neq j$, and has a unit projection onto $[T^{-1}]_{*,i}$. Then our two approximation methods give answers

$$\begin{aligned} \text{CE: } \hat{q}_i(y_i) &= N(m_{y_i}, ([T^{-1}]_{*,i}^H R_x^{-1} [T^{-1}]_{*,i})^{-1}) \\ \text{RCE: } \hat{q}_i(y_i) &= N(m_{y_i}, [T]_{i,*} R_x [T]_{i,*}^H). \end{aligned} \quad (3.28)$$

Note that the Cross-Entropy solution for the i th variance depends on $[T^{-1}]_{*,i}$, which in turn depends on the choice of all the linear transformations $[T]_{1,*}, \dots, [T]_{p,*}$. The Reversed Cross-Entropy solution for the i th variance, on the other hand, depends only on $[T]_{i,*}$.

IV. POWER SPECTRUM ESTIMATION

We can easily apply these ideas to power spectrum estimation. Let $\mathbf{x}(t)$ be a complex-valued, multidimensional, stationary vector Gaussian random process, with zero mean and with stationary covariance kernel $R(t)$. Assume that each sample $\mathbf{x}(t)$ is a vector of length p , and that the process is defined for all t in a hypercube, $-U \leq t \leq U$.

$$\begin{aligned} E[\mathbf{x}(t)] &= \mathbf{0} \quad \text{for } -U \leq t \leq U \\ E[\mathbf{x}(t_1)\mathbf{x}^H(t_2)] &= R(t_1 - t_2) \quad \text{for } -U \leq t_1, t_2 \leq U. \end{aligned} \quad (4.1)$$

We will assume that $R(t)$ is known for all t . (The derivation for real-valued $\mathbf{x}(t)$ would be slightly different, but the results would look the same.) Let us define $\mathbf{X}(\omega)$ as the multidimensional vector Fourier Transform of $\mathbf{x}(t)$

over the known domain

$$\mathbf{X}(\boldsymbol{\omega}) = \int_{-U}^U \mathbf{x}(t) e^{-j\boldsymbol{\omega}^T t} dt. \quad (4.2)$$

Also, let $P(\boldsymbol{\omega})$ be the Fourier Transform of $R(t)$

$$P(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} R(t) e^{-j\boldsymbol{\omega}^T t} dt. \quad (4.3)$$

Now suppose we measure a selected set of samples of this random process, $\mathbf{x}_n = \mathbf{x}(t_n)$, where the samples t_n may be uniformly or nonuniformly distributed in the domain $[-U, U]$. The vector of samples \mathbf{x} will be a zero mean Gaussian random variable with covariance matrix R_x having elements $[R_x]_{k,l} = R(t_k - t_l)$. The covariance matrix R_x is positive definite and symmetric. If the samples t_n are uniformly spaced on a rectilinear grid and organized in raster scan fashion, then R_x will also be block Toeplitz.

Let us consider forming estimates of P selected frequency samples $\boldsymbol{\omega}_k$ of the Fourier Transform $\mathbf{X}(\boldsymbol{\omega}_k)$ using the given L data samples \mathbf{x}_n . We will form each estimate y_k of $\mathbf{X}(\boldsymbol{\omega}_k)$ by taking an appropriate linear combination of the data samples

$$y_k = \sum_n T_{k,n} \mathbf{x}_n. \quad (4.4)$$

Let $T_k(\boldsymbol{\omega})$ be the transform of the k th set of coefficients

$$T_k(\boldsymbol{\omega}) = \sum_n T_{k,n} e^{j\boldsymbol{\omega}^T t_n}. \quad (4.5)$$

Applying Parseval's theorem to (4.4), we can show that

$$y_k = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} T_k(\boldsymbol{\omega}) \mathbf{X}(\boldsymbol{\omega}) d\boldsymbol{\omega}. \quad (4.6)$$

Thus, y_k will be a good approximation to $\mathbf{X}(\boldsymbol{\omega}_k)$ if we choose coefficients $T_{k,n}$ such that $T_k(\boldsymbol{\omega})$ is a good approximation to an impulse centered at $\boldsymbol{\omega}_k$. Furthermore, because the samples y_k are linear combinations of Gaussian random variables, \mathbf{x}_n , they will also be Gaussian random variables with mean and covariance

$$\begin{aligned} \mathbf{m}_{y_k} &= E[y_k] = \mathbf{0} \\ [R_y]_{k,l} &= E[y_k y_l^H] \\ &= \sum_{n,m} T_{k,n} R(t_n - t_m) T_{l,m}^H \\ &= \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} T_k(\boldsymbol{\omega}) P(\boldsymbol{\omega}) T_l^H(\boldsymbol{\omega}) d\boldsymbol{\omega}. \end{aligned} \quad (4.7)$$

Thus, if the filters $T_k(\boldsymbol{\omega})$ are designed with minimal overlap, then the correlation between different samples y_k and y_l will be minimal. In other words, the samples y_k will be nearly independent zero-mean Gaussian random variables with variance related to the power spectrum sample $P(\boldsymbol{\omega}_k)$.

It would be convenient for many estimation problems if we could approximate this actual distribution with one

in which the frequency component estimates y_k were modeled as exactly independent random variables. Hopefully, the variance of y_k in this separable approximation would then be a good estimate of $P(\boldsymbol{\omega}_k)$. The separability of the model, moreover, will often simplify further analysis of the expected characteristics of the process.

To find a good separable model, we can apply the CE and RCE methods discussed in the previous sections. Collect the samples y_k into a vector \mathbf{y} , and the elements $T_{k,n}$ into a matrix T . Assume that $P = L$ and that T is invertible. In the following subsections, we will compute the CE and RCE separable approximations, and show that the corresponding variances of y_k can be treated as estimates of the power spectrum.

A. Reverse Cross-Entropy Spectral Estimation

The RCE estimate is somewhat easier to interpret, so we will treat it first. Each y_k is constructed by filtering the data \mathbf{x}_n through a filter $T_{k,n}$ as in (4.4). The RCE separable approximation models the density $\hat{q}_k(y_k)$ as the marginal density of y_k . If we define $\hat{P}_{RCE}(\boldsymbol{\omega}_k)$ as the variance of y_k , then (3.28) shows that

$$\text{RCE: } \hat{q}_k(y_k) = N(\mathbf{0}, \hat{P}_{RCE}(\boldsymbol{\omega}_k))$$

where

$$\hat{P}_{RCE}(\boldsymbol{\omega}_k) = [R_y]_{k,k} = \sum_{n,m} T_{k,n} R(t_n - t_m) T_{k,m}^H. \quad (4.8)$$

To show that this is a reasonable power spectrum estimate, note that (4.7) implies that

$$\hat{P}_{RCE}(\boldsymbol{\omega}_k) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} T_k(\boldsymbol{\omega}) P(\boldsymbol{\omega}) T_k^H(\boldsymbol{\omega}) d\boldsymbol{\omega}. \quad (4.9)$$

Thus, the Reverse Cross-Entropy method is equivalent to a classical windowing approach for estimating the power spectrum. If the $T_k(\boldsymbol{\omega})$ are good bandpass filters centered at $\boldsymbol{\omega}_k$, then the estimate formed from the linear combination of correlations in (4.8) equals an average of the power spectrum values around $\boldsymbol{\omega}_k$. For proper scaling in the RCE method, it seems reasonable to scale the filters $T_k(\boldsymbol{\omega})$ such that if the actual power spectrum is flat, $P(\boldsymbol{\omega}) = P_0$, then $\hat{P}_{RCE}(\boldsymbol{\omega}) = P_0$ also. Applying this requirement to (4.9) gives the normalization constraints

$$\begin{aligned} &\frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} T_k(\boldsymbol{\omega}) P_0 T_k^H(\boldsymbol{\omega}) d\boldsymbol{\omega} \\ &= \sum_n T_{k,n} P_0 T_{k,n}^H = P_0 \quad \text{for all } P_0. \end{aligned} \quad (4.10)$$

A particularly good choice for the filters $T_{k,n}$ would be rectangularly windowed complex exponentials

$$T_{k,n} = \frac{1}{\sqrt{L}} e^{-j\boldsymbol{\omega}_k^T t_n} \mathbf{1} \quad (4.11)$$

where the normalization factor $1/\sqrt{L}$ is chosen to satisfy (4.10). With this choice

$$\hat{P}_{RCE}(\omega_k) = \frac{1}{L} \sum_{n,m} R(t_n - t_m) e^{-j\omega_k^T(t_n - t_m)}$$

and so the RCE power spectral density estimate is identical to a classical Bartlett (triangular) windowed periodogram estimate.

The difficulty with the RCE power spectrum is that, in general, the filters $T_k(\omega)$ overlap, so that $R_y = TR_x T^H$ is far from diagonal. Leakage between the different filters implies that even if all the power in the actual power spectrum is concentrated near ω_k , our $\hat{P}_{RCE}(\omega)$ will be non-zero over a wide range of frequencies. This is the familiar problem with sidelobe leakage in classical periodogram estimation.

B. Cross-Entropy Spectral Estimation

In the CE method, the component density $\hat{q}_k(y_k)$ is estimated as equal to the conditional density of y_k given that all the remaining components are exactly equal to their *a posteriori* mean. If we define $\hat{P}_{CE}(\omega_k)$ as the variance of y_k in this separable approximation, then (3.28) implies that

$$\text{CE: } \hat{q}_k(y_k) = N(\mathbf{0}, \hat{P}_{CE}(\omega_k))$$

where

$$\begin{aligned} \hat{P}_{CE}(\omega_k) &= [R_y^{-1}]_{k,k}^{-1} \\ &= \left[\sum_{n,m} [T^{-1}]_{n,k}^H [R_x^{-1}]_{n,m} [T^{-1}]_{m,k} \right]^{-1}. \end{aligned} \quad (4.12)$$

The form of this estimate is highly reminiscent of a windowed form of Capon's MLM estimation method, in which columns of the inverse filter matrix T^{-1} replace the directional vectors in Capon's method. A simple interpretation of this effect is difficult, however, since the k th column of T^{-1} will depend, in a nonlinear fashion, on the choice of all the filters $T_l(\omega)$, and the elements of R_x^{-1} depend nonlinearly on all the given correlations $R(t_n - t_m)$.

The key to understanding this formula is to recognize that this is the conditional variance of y_k given the remaining components. $\hat{P}_{CE}(\omega_k)$ is thus the expected power in y_k which cannot be predicted from knowledge of the other frequency estimates. This estimator therefore tends to minimize leakage from other frequency bands. A side effect of this leakage cancellation, however, is that leakage from the k th filter into adjacent filters will be used to partially predict the power in y_k , thus artificially reducing the power estimate $\hat{P}_{CE}(\omega_k)$ below its proper value. This argument supports our observation in (3.14), which in this context implies that

$$\hat{P}_{CE}(\omega_k) \leq \hat{P}_{RCE}(\omega_k). \quad (4.13)$$

It would be appropriate to choose different filter scaling for the CE method to compensate for this effect. In particular, if $P(\omega)$ is flat, $P(\omega) = P_0$, then $R(t) = P_0 \delta(t)$ and R_x is block diagonal. If we require that $\hat{P}_{CE}(\omega) = P_0$ in this case, then formula (4.12) suggests that we scale T such that

$$\sum_n [T^{-1}]_{n,k}^H P_0^{-1} [T^{-1}]_{n,k} = P_0^{-1} \quad \text{for all } P_0. \quad (4.14)$$

A particularly useful set of filters $T_{k,n}$ for the CE method is found by designing them so that they do not overlap at their center frequencies

$$T_k(\omega_l) = \sqrt{L} \delta_{k,l} I \quad \text{for all } k, l \quad (4.15)$$

where \sqrt{L} is a scaling factor chosen to satisfy (4.14). The advantage of this idea is that if the signal is actually a sum of complex exponentials at frequencies ω_k with gains X_k

$$x(t_n) = \sum_k X_k e^{j\omega_k^T t_n}, \quad (4.16)$$

then it is easy to show that

$$y_k = \sqrt{L} X_k. \quad (4.17)$$

Thus, each filter output would only reflect the amplitude of the frequency component at ω_k , and would be insensitive to frequency components at the other frequencies ω_l . Another nice property of this method is that if we write (4.15) in matrix form, then we find that

$$TW = \sqrt{L} I \quad (4.18)$$

where W is a block DFT matrix, $[W]_{k,l} = e^{j\omega_l^T t_k} I$. Substituting the resulting formula for T^{-1} into (4.12)

$$\hat{P}_{CE}(\omega_k) = L \left[\sum_{n,m} [R_x^{-1}]_{n,m} e^{-j\omega_k(t_n - t_m)} \right]^{-1}. \quad (4.19)$$

For this choice of filters, the CE power spectrum is exactly equal to Capon's MLM estimator. As is well known, this MLM method exhibits less sidelobe leakage than the periodogram (RCE) estimates.

V. FILTER DESIGN—UNIFORM SAMPLING

When the samples t_n and ω_k are selected on uniform rectilinear grids, with the number of time samples equal to the number of frequency samples, $L = P$, and the power spectrum is band-limited with the Nyquist frequency, then it is easy to design good bandpass filters $T_k(\omega)$. In particular, it is easy to show that the rectangular windowed filters suggested in (4.11) for the RCE method are identical to the "nonoverlapping" filters suggested in (4.15) for the CE method. In fact, this T matrix is orthonormal, with $T^{-1} = T^H = W/\sqrt{L}$. The formula for the CE method becomes identical to the triangular Bartlett periodogram estimate, and the RCE method is identical to Capon's MLM method.

An interesting interpretation of these separable approximations results if we transform the separable densities $\hat{q}_{CE}(y)$ and $\hat{q}_{RCE}(y)$ back into densities on the variable $x = T^{-1}y$: $\hat{q}_{CE}(x)$ and $\hat{q}_{RCE}(x)$. The original density $p(x)$

$= N(\mathbf{0}, R_x)$ has a block Toeplitz covariance matrix R_x . Using the fact that $T^{-1} = T^H$, the separable approximations can be shown to correspond to

$$\hat{q}_{CE}(\mathbf{x}) = N(\mathbf{0}, T^H \Lambda_{CE} T) \quad (5.1)$$

and

$$\hat{q}_{RCE}(\mathbf{x}) = N(\mathbf{0}, T^H \Lambda_{RCE} T) \quad (5.2)$$

where Λ_{CE} and Λ_{RCE} are block diagonal matrices with diagonal block elements $\hat{P}_{CE}(\omega_k)$ and $\hat{P}_{RCE}(\omega_k)$, respectively. The formulas for the covariance matrices in (5.1) are thus in Jacobian normal form. The columns of T^H , which are complex exponentials, are the eigenvectors of the matrix, and the diagonal elements of the Λ , which are uniformly spaced power spectrum sample estimates, are the eigenvalues. The resulting covariance matrices are thus different block circulant approximations to the original Toeplitz covariance matrix R_x , in which each block row is equal to the previous block row rotated right one entry. This is particularly fortuitous, since it is well known that as the size L of a noncirculant Toeplitz matrix R_x goes to infinity, the eigenvalues of R_x will approach uniformly spaced samples of the power spectrum, and the eigenvectors will approach uniformly spaced complex exponentials [10], [11]. In the limit as $L \rightarrow \infty$, we would therefore expect both separable approximations to be identical to the original model.

VI. FILTER DESIGN—NONUNIFORM SAMPLING

Achieving good spectral estimates with nonuniformly spaced samples t_n , unfortunately, is much more difficult [12], [13]. The fundamental difficulty is that when the t_n are distributed in a grossly nonuniform manner, then it is difficult, if not impossible, to design good bandpass filters with nonzero coefficients $T_{k,n}$ located at t_n . Furthermore, since the Nyquist theorem does not apply, it is hard to select a natural band-limited assumption for the problem, and therefore it is not even obvious how to choose the frequencies ω_k . Another problem is that we will not be able to apply our elegant interpretation of these methods as circulant matrix approximations to Toeplitz covariance matrices, since the nonuniformly sampled covariance matrix R_x will not be Toeplitz.

We have tried a large number of filter design strategies for various nonuniform grids t_n , but have not had much success in achieving good results for both the CE and RCE methods simultaneously. A key difficulty is that in the nonuniform sampling case, the filters in (4.11) are quite different from those in (4.15). Experimental evidence suggests that the rectangular filters suggested in (4.11) for the RCE method work acceptably for RCE, but they lead to an ill-conditioned T^{-1} matrix and give poor results for CE. Similarly, the nonoverlapped filters suggested in (4.15) for CE work well for CE, but generally lead to an ill-conditioned T matrix and give poor results for RCE.

Other design strategies we have tried include using warped sinc functions to interpolate reasonable bandpass filters $T_k(\omega)$, and iterative design techniques which iter-

ate between the time and frequency domains to find a low-pass filter shape with nonzero coefficients in the appropriate places. It is possible to find filters which work well in either the CE or the RCE methods; for example, the nonoverlapping filter designs in (4.15) work well for CE and give the same result as Capon's MLM method. Unfortunately, we could not find filters which work well in both methods. This is clearly a topic for further research.

VII. CONCLUSIONS

In this paper, we have applied two different cross-entropy criteria to the power spectrum estimation problem. In each case, given L samples of data, we build L linearly independent filters to estimate the signal transform at L different frequency samples ω_k . Provided these filters are designed to be good bandpass filters with minimal overlap, their outputs will be good estimates of the signal transform at ω_k , and their outputs will be nearly independent Gaussian random variables with covariance related to the power spectrum. In our cross-entropy methods, we force two different "optimal" separable approximations to the actual model, in which these filter outputs are approximated as being exactly independent. One of these methods yields an estimate of the variance of each filter output which is similar to a windowed periodogram estimate. The other method yields an estimate of the variance which is similar to Capon's MLM spectral estimate. In the case of uniformly sampled data, uniform frequency samples, and rectangular filters, we get exactly a triangular window Bartlett spectral estimate and Capon's MLM estimate. The solutions in this case can be viewed as building circulant approximations to the original Toeplitz covariance matrix R_x .

The key to getting good spectral estimates from the CE and RCE methods is to choose the transform matrix T carefully. In designing T , we need to keep several goals in mind. The samples y_k must be accurate and robust estimates of $X(\omega_k)$, and the variance of each y_k must be an accurate estimate of the power spectrum $P(\omega_k)$. At the same time, the y_k must be chosen to be as independent of each other as possible, so that R_y is almost block diagonal, and so that our separable approximation is valid.

One difficulty with both methods is that, given L samples of x_n , they estimate at most L samples from the power spectrum, y_k . This appears to be a fundamental statement about the achievable spectral resolution: given L data samples, only L independent frequency samples can be estimated. In many applications, however, it would be useful to interpolate between these known samples to get a continuous power spectrum estimate at all frequencies. A related difficulty is involved in choosing the set of L frequencies of interest, ω_k . If we change only one of these frequencies, and then redesign T for the new set, we may get somewhat different estimates of the power at all the frequencies.

For the special case of uniform spatial and frequency samples with rectangular windows, these issues are easily resolved. Fortunately, in this case, the spectral estimation

formulas at frequency ω_k depend only on ω_k , and do not explicitly refer to the remaining $L - 1$ samples. Simply by replacing ω_k with a general frequency parameter ω , we can convert these formulas into a continuous power spectrum estimate.

For nonuniformly sampled data, unfortunately, it is difficult to achieve all the design goals for these methods. Designing good bandpass filters on a nonuniform grid, for example, is often difficult or impossible. The aliasing that occurs cannot be simply expressed as a periodic summation at the Nyquist frequency, but is a more complex distortion of the spectrum. Picking a good set of frequencies to force to be independent is difficult. Theorems are needed to show that a separable approximation is asymptotically correct in the limit as the size of the nonuniformly sampled, non-Toeplitz covariance matrix R_x goes to infinity. Perhaps a better method might force selected samples of the Laplace transform to be independent random variables, where the samples are not necessarily located on the imaginary axis. Although we have tried several strategies for designing nonuniform filters for grossly nonuniformly sampled data, we have been unable to get good spectral estimates from both methods with the same filter matrix T . The key to good performance, undoubtedly, will be adaptively design the T filter matrix using the given covariance matrix R_x in a manner similar to that used by Capon in deriving MLM. Much work remains to be done on modifying this idea for the cross-entropy spectral estimation techniques.

APPENDIX A DERIVATION OF CE METHOD

To solve (3.3), let us focus on the i th density $q_i(\mathbf{y}_i)$. We can rewrite (3.3) as

$$H(q, p) = \int q_i(\mathbf{y}_i) \log \frac{q_i(\mathbf{y}_i)}{\phi_i(\mathbf{y}_i)} d\mathbf{y}_i + \text{other terms} \quad (\text{A.1})$$

where the other terms do not depend on q_i , and where:

$$\log \phi_i(\mathbf{y}_i) = \int \log p(\mathbf{y}_i | \{\mathbf{y}_j, j \neq i\}) \prod_{\substack{j=1 \\ j \neq i}}^P q_j(\mathbf{y}_j) d\mathbf{y}_j. \quad (\text{A.2})$$

Substituting the formula for the conditional density [derived from (2.6)] into this and evaluating the expectations over $\mathbf{y}_1, \dots, \mathbf{y}_{i-1}, \mathbf{y}_{i+1}, \dots, \mathbf{y}_p$ gives

$$\log \phi_i(\mathbf{y}_i) = -\xi(\mathbf{y}_i - \bar{\boldsymbol{\mu}}_i)^H [R_y^{-1}]_{i,i} (\mathbf{y}_i - \bar{\boldsymbol{\mu}}_i) + \text{constant} \quad (\text{A.3})$$

where

$$\bar{\boldsymbol{\mu}}_i = \mathbf{m}_{y_i} - [R_y^{-1}]_{i,i}^{-1} \prod_{\substack{j=1 \\ j \neq i}}^P [R_y^{-1}]_{i,j} (E[\mathbf{y}_j | q_j] - \mathbf{m}_{y_j})$$

$$\xi = \begin{cases} \frac{1}{2} & \text{if } \mathbf{y} \text{ is real} \\ 1 & \text{if } \mathbf{y} \text{ is complex} \end{cases} \quad (\text{A.4})$$

and where $E[\cdot | q_j]$ is the expectation with respect to density q_j . The minimum of $H(q, p)$ over q_i must occur at the value which minimizes (A.1) subject to the constraint that $\int q_i(\mathbf{y}_i) d\mathbf{y}_i = 1$. Equation (A.1), however, is strictly convex in q_i , and thus it is easy to show that the unique minimum is

$$\hat{q}_i(\mathbf{y}_i) = K_i \phi_i(\mathbf{y}_i) \quad (\text{A.5})$$

where K_i is a normalization constant. Formula (A.3) then implies that the $\hat{q}_i(\mathbf{y}_i)$ are Gaussian

$$\hat{q}_i(\mathbf{y}_i) = N(\bar{\boldsymbol{\mu}}_i, [R_y^{-1}]_{i,i}^{-1}). \quad (\text{A.6})$$

All that remains is to solve for the $\bar{\boldsymbol{\mu}}_i$ values. Note that $\bar{\boldsymbol{\mu}}_i = E[\mathbf{y}_i | \hat{q}_i]$. Substituting this into (A.4), multiplying both sides by $[R_y^{-1}]_{i,i}$, and rearranging, gives

$$\sum_{j=1}^P [R_y^{-1}]_{i,j} (\bar{\boldsymbol{\mu}}_j - \mathbf{m}_{y_j}) = \mathbf{0} \quad \text{for } i = 1, \dots, P. \quad (\text{A.7})$$

Combining terms into matrices and vectors in the obvious way,

$$R_y^{-1} (\bar{\boldsymbol{\mu}} - \mathbf{m}_y) = \mathbf{0}. \quad (\text{A.8})$$

Since R_y^{-1} is assumed to be full rank,

$$\bar{\boldsymbol{\mu}} = \mathbf{m}_y. \quad (\text{A.9})$$

The *a posterior* means $\bar{\boldsymbol{\mu}}_i$ must therefore exactly equal to the *a priori* means \mathbf{m}_{y_i} , and thus the probability density estimate \hat{q}_{CE} which minimizes the Shore-Johnson cross-entropy is as given in (3.4).

APPENDIX B CROSS-ENTROPY FOR THREE UNKNOWNNS

Consider a problem with three real-valued unknowns with covariance matrix

$$R_y = \text{Cov} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \rho_{13} \\ 0 & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}. \quad (\text{B.1})$$

Then

$$R_y^{-1} = \frac{1}{1 - \rho_{23}^2 - \rho_{13}^2} \begin{pmatrix} 1 - \rho_{23}^2 & \rho_{13}\rho_{23} & -\rho_{13} \\ \rho_{13}\rho_{23} & 1 - \rho_{13}^2 & -\rho_{23} \\ -\rho_{13} & -\rho_{23} & 1 \end{pmatrix}. \quad (\text{B.2})$$

Equations (3.6) and (3.12) show that

$$H(\hat{q}_{CE}, p) = \frac{1}{2} \log \left(\frac{(1 - \rho_{23}^2)(1 - \rho_{13}^2)}{(1 - \rho_{23}^2 - \rho_{13}^2)^2} \right)$$

$$H(p, \hat{q}_{RCE}) = \frac{1}{2} \log \left(\frac{1}{1 - \rho_{23}^2 - \rho_{13}^2} \right). \quad (B.3)$$

Subtracting gives

$$H(\hat{q}_{CE}, p) - H(p, \hat{q}_{RCE})$$

$$= \frac{1}{2} \log \left(1 + \frac{\rho_{13}^2 \rho_{23}^2}{1 - \rho_{23}^2 - \rho_{13}^2} \right). \quad (B.4)$$

Thus, the difference between the cross-entropy measures of the two separable approximations can be arbitrarily large in a positive direction. If we had started instead with a covariance matrix \hat{R}_y with the formula in (B.2), then the difference would have been arbitrarily large in a negative direction. Thus, for three or more unknowns, the value of the minimum cross-entropies for the CE and RCE methods can be arbitrarily different.

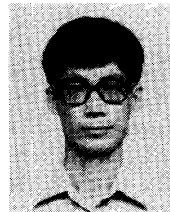
ACKNOWLEDGMENT

We would like to thank G. Duckworth and J. K. Vandiver for numerous discussions on this topic, and D. Cobra for programming assistance with filter design.

REFERENCES

[1] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*. San Francisco, CA: Holden-Day, 1968.
 [2] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*. New York: Dover, 1958.
 [3] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 70-73; reprinted in *Modern Spectrum Analysis*. D. G. Childers, Ed. New York: IEEE Press, June 1967.
 [4] J. P. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Stanford Univ., May 1975.
 [5] T. L. Marzetta, "A new interpretation for Capon's maximum likelihood method of frequency-wavenumber spectral estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 445-449, Apr. 1983.
 [6] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, pp. 661-675; reprinted in *Modern Spectrum Analysis*. D. G. Childers, Ed. New York: IEEE Press, 1978.

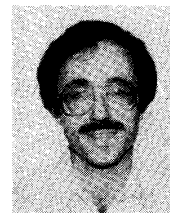
[7] B. R. Musicus, "Iterative algorithms for optimal signal reconstruction and parameter identification given noisy and incomplete data," Ph.D. dissertation, M.I.T., Dep. EECS, Aug. 1982.
 [8] S. Kullback, *Information Theory and Statistics*. New York: Wiley, 1959.
 [9] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 26-37, Jan. 1980.
 [10] R. Molten Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 725-730, Nov. 1972.
 [11] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. Berkeley, CA: University of California Press, 1958.
 [12] R. H. Jones, "Estimation of spatial wavenumber spectra and falloff rate with unequally spaced observations," *J. Astr. Sci.*, vol. 32, no. 23, pp. 260-268.
 [13] P. R. Julian and A. K. Cline, "The direct estimation of spatial wavenumber spectra of atmospheric variables," *J. Atmos. Sci.*, vol. 31, no. 6, pp. 1526-1539, Sept. 1974.



Cheng-Yuan Liu (M'86) was born in Taiwan on November 14, 1951. He received the B.S. degree in physics from National Central University in 1974, the M.S. degree in physical oceanography from National Taiwan University in 1978, and the Ph.D. degree in ocean engineering from the Massachusetts Institute of Technology in 1985.

From 1986 to 1987 he was Visiting Associate Professor in the Institute of Applied Mechanics, National Taiwan University, where he taught courses in stochastic processes and digital signal processing and did research in sonar array signal processing and signal diagnostic systems. In 1988 he joined the Faculty at the same university, where he is currently Associate Professor in the Department of Computer Science and Information Engineering. His current interests include array signal processing and neural networks.

Dr. Liu is a member of the International Neural Network Society.



Bruce R. Musicus (S'77-M'81) was born in Chicago, IL, in 1954. He received the S.B. degree from Harvard University, Cambridge, MA, in 1975, the M.S. and E.E. degrees from M.I.T. in 1979, and the Ph.D. degree from M.I.T. in 1982, all in electrical engineering and computer science.

He was an Assistant Professor of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, from 1982 to 1986, and has been an Associate Professor there since then. He has sat on the Rockwell International Career Development chair at M.I.T. His research interests include algorithms and architectures for digital signal processing, stochastic estimation, and system identification; and he has published numerous papers in these fields.