POWER SPECTRUM ESTIMATION BY MEANS OF RELATIVE-ENTROPY MINIMIZATION WITH UNCERTAIN CONSTRAINTS

Rodney W. Johnson and John E. Shore

Information Technology Division
Naval Research Laboratory
Washington, D. C. 20375

ABSTRACT

This paper presents a new multisignal spectrum-analysis method based on the relative-entropy principle ("principle of minimum cross entropy"). The method simultaneously estimates a number of power spectra given (1) an initial ("prior") estimate of each, (2) imprecise values of the autocorrelation function of their sum, and (3) estimates of the error in measurement of the autocorrelation values. One application is to the separate estimation of the spectra of a signal and independent additive noise, based on imprecise measurements of the autocorrelations of the signal plus noise. A derivation is sketched, beginning with a discussion of relative-entropy minimization subject to quadratic constraints that require certain linear constraints to hold approximately, to within a specified error bound. The new method is an extension of multisignal relative-entropy spectrum analysis (with exact autocorrelations). The two methods are compared, and connections with previous related work are indicated. Mathematical properties of the new method are discussed, and an illustrative numerical example is presented.

I. INTRODUCTION

Relative-Entropy Spectrum Analysis (RESA) is an extension of Burg's Maximum-Entropy Spectral Analysis (MESA) [1,2] that was introduced by Shore [3]. Like MESA, it estimates a spectrum from values of the autocorrelation function. RESA, however, also takes into account prior information in the form of an *initial* estimate of the spectrum. Multisignal RESA (MRESA), introduced by Shore and Johnson [4], simultaneously estimates the power spectra of several signals when an initial estimate for each spectrum is available and new information is obtained in the form of values of the autocorrelation function of the sum. The resulting *final* estimates are the solution of a constrained minimization problem: they are consistent with the autocorrelation information and otherwise as similar as possible to the respective initial estimates in a precisely defined information-theoretic sense. MRESA has recently been extended by Johnson, Shore, and Burg to incorporate weighting factors associated with each initial spectrum estimate to allow for the fact that initial estimates may not be uniformly reliable [5].

The autocorrelation values were treated in [3,4,5] as exactly given. Usually, however, these are estimated or measured values subject to error. We here show how to incorporate an error bound to allow for uncertainty in autocorrelation values.

The principle of minimum relative entropy (MRE) provides a general, information-theoretic method of inference when information about an unknown probability density $q^\dagger$ consists of an *initial* estimate p and expected values

$$\bar{f}_r = \int f_r(x) q^\dagger(x) dx \quad (1)$$

for known $f_k(x)$ and $\bar{f}_r$, $r=0, \dots, M$. The principle states that one should choose the *final* estimate q that satisfies

$$H(q, p) = \min_q H(q^\dagger, p),$$

where H is the *relative entropy* (cross-entropy, discrimination information, directed divergence, I-divergence, K-L number, etc.),

$$H(q, p) = \int q(x) \log \frac{q(x)}{p(x)} dx, \quad (2)$$

and where $q^\dagger$ varies over the set of densities that satisfy the *constraints* (1). The final estimate has the form

$$q(x) = p(x) \exp \left\{ -\alpha - \sum_r \beta_r f_r(x) \right\}, \quad (3)$$

where the β_r and α are Lagrangian multipliers determined by (1) (with $q^\dagger$ replaced by q) and by the normalization constraint

$$\int q(x) dx = 1. \quad (4)$$

Properties of MRE solutions and conditions for their existence are discussed in [6,7]. As a general method of statistical inference, MRE was first introduced by Kullback [8], has been advocated in various forms by others [9,10,11], and has been applied in various fields (for a list of references, see [12]).

Informally speaking, of the densities that satisfy the constraints, MRE selects the one that is closest to p in the sense measured by relative entropy. In more formal terms, MRE can be justified on the basis of the information-theoretic properties of relative entropy [8], or on the basis of consistency axioms for logical inference [12].

In applying MRE to multisignal spectrum estimation, we assume the existence of L independent signals with power spectra $S_i(f)$ and autocorrelations

$$R_{ir} = \int C_r(f) S_i(f) df, \quad (5)$$

where

$$C_r(f) = \cos 2\pi t_r f. \quad (6)$$

Given initial estimates $P_i(f)$ of the power spectrum of each signal S_i , and autocorrelation measurements on the sum of the signals, we wish to obtain final estimates for the S_i . In particular, we assume the measurements

$$R_r^{\text{tot}} = \sum_{i=1}^L \int C_r(f) Q_i(f) df, \quad (7)$$

for lags $r=0, \dots, M$. The resulting final estimates are

$$Q_i(f) = \frac{1}{\frac{1}{P_i(f)} + \sum_r \beta_r C_r(f)}, \quad (8)$$

where the β_r are chosen so that the Q_i satisfy the autocorrelation constraints (7) [4]. Since some initial estimates may be more reliable than others, these results have been extended recently to include a frequency-dependent weight $w_i(f)$ for each initial estimate $P_i(f)$ [5]. The larger the value of $w_i(f)$ the more reliable the initial estimate $P_i(f)$ is considered to be. With the weights included, the result (8) becomes

$$Q_i(f) = \frac{1}{\frac{1}{P_i(f)} + \frac{1}{w_i(f)} \sum_r \beta_r C_r(f)}. \quad (9)$$

II. RELATIVE-ENTROPY MINIMIZATION WITH UNCERTAIN CONSTRAINTS

In most applications of MRE, the known expected values $\bar{f}_r$ in (1) correspond to physical measurements. Such measurements usually are subject to error so that the equality in (1) is unrealistic. In the applications to spectrum estimation, for example, the expected values correspond to autocorrelations whose values are known only approximately. In this Section we extend MRE to incorporate uncertainty about the values of the $\bar{f}_r$.

We define an error vector $\mathbf{v}$ with components

$$v_r = \int f_r(x) q^\dagger(x) dx - \bar{f}_r. \quad (10)$$

A simple generalization would be to replace the set of constraints (1) with a bound on the magnitude of $\mathbf{v}$:

$$\sum_r \left[\int f_r(x) q^\dagger(x) dx - \bar{f}_r \right]^2 \leq \varepsilon^2. \quad (11)$$

However, all components v_r may not have equal uncertainty, and different components may be correlated. We therefore replace (11) with the more general constraint

$$\sum_{rs} M_{rs} v_r v_s \leq \varepsilon^2. \quad (12)$$

In matrix notation this is

$$\mathbf{v} \cdot \mathbf{M} \mathbf{v} \leq \varepsilon^2, \quad (13)$$

where $\mathbf{M}$ is any positive-definite matrix.

We assume that we are given an initial estimate p of $q^\dagger$, measured values $\bar{f}_r$ of the expectations (1) of functions f_r for a finite set of indices r , and an error estimate ε . We will first derive the form of the final estimate q under the assumption that the constraint has the form (11) and that the $\bar{f}_r$ are 0; that is, we assume a constraint

$$\sum_r \left[\int f_r(x) q^\dagger(x) dx \right]^2 \leq \varepsilon^2. \quad (14)$$

Next we show how to reduce the more general constraint (12) to this case. We conclude this section with a remark on the relation between the result with $\varepsilon > 0$ and that for "exact constraints" ($\varepsilon = 0$).

Our problem is to minimize the relative entropy $H(q, p)$ subject to the constraint (14) (with q in place of $q^\dagger$) and the normalization constraint (4). If the initial estimate satisfies the constraint (i.e., (14) holds with p in place of q), then setting $q = p$ gives the minimum. Otherwise equality holds in (14), and the criterion for a minimum is that the variation of

$$\int q(x) \log \frac{q(x)}{p(x)} dx + \lambda \sum_r \left[\int f_r(x) q(x) dx \right]^2 + (\alpha - 1) \int q(x) dx \quad (15)$$

with respect to $q(x)$ is zero for some Lagrange multipliers $\lambda > 0$, corresponding to (14), and $\alpha - 1$, corresponding to (4). (We write $\alpha - 1$ instead of α for later convenience.) With $\lambda > 0$, the criterion intuitively implies that a small change δq in q that leaves $\int q(x) dx$ fixed and decreases $H(q, p)$ must increase the error term $\sum_r \left[\int f_r(x) q(x) dx \right]^2$.

Equating the variation of (15) to zero gives

$$\log \frac{q(x)}{p(x)} + \alpha + \lambda \sum_r 2 f_r(x) \int f_r(x') q(x') dx' = 0.$$

Therefore q satisfies

$$q(x) = p(x) \exp \left[-\alpha - \sum_r \beta_r f_r(x) \right] \quad (16)$$

where

$$\beta_r = 2\lambda \int f_r(x) q(x) dx. \quad (17)$$

Conversely, if q has the form (16), and if α , λ , and the β_r are chosen so that (17), the constraint (14), and the normalization condition (4) hold, then q is a solution to the minimization problem. But if (17) holds, the constraint with equality is equivalent to

$$\sum_r \left(\frac{\beta_r}{2\lambda} \right)^2 = \varepsilon^2,$$

or to

$$\lambda = \frac{1}{2\varepsilon} \|\boldsymbol{\beta}\|,$$

where we have written $\|\boldsymbol{\beta}\|$ for the Euclidean norm $(\sum_r \beta_r^2)^{1/2}$. Thus if we choose α and β_r in (16) so that (4) and

$$\varepsilon \frac{\beta_r}{\|\boldsymbol{\beta}\|} = \int f_r(x) q(x) dx \quad (18)$$

hold, then the constraint (14) will be satisfied, and we can assure that (17) holds by the choice of λ .

Next assume a constraint of the general form (12), (10), with a symmetric, positive-definite matrix $\mathbf{M}$. Then there is a matrix $\mathbf{A}$, not in general unique, such that $\mathbf{A}^t \mathbf{A} = \mathbf{M}$. Now

$$\mathbf{v} \cdot \mathbf{M} \mathbf{v} = \mathbf{v} \cdot \mathbf{A}^t \mathbf{A} \mathbf{v} = (\mathbf{A} \mathbf{v}) \cdot (\mathbf{A} \mathbf{v}) = \sum_r \left(\sum_s A_{rs} v_s \right)^2,$$

and so the constraint assumes the form

$$\sum_r u_r^2 \leq \varepsilon^2, \quad (19)$$

where

$$u_r = \sum_s A_{rs} v_s.$$

In view of (4) we may rewrite (10) as

$$v_r = \int (f_r(x) - \bar{f}_r) q(x) dx$$

and obtain

$$u_r = \int \sum_s A_{rs} (f_s(x) - \bar{f}_s) q(x) dx$$

Defining

$$g_r(x) = \sum_s A_{rs} (f_s(x) - \bar{f}_s), \quad (20)$$

we obtain

$$\sum_r \left[\int g_r(x) q(x) dx \right]^2 \leq \varepsilon^2. \quad (21)$$

from (19). Thus constraints of the form general (12) can be transformed to (21), which is of the same form as (14).

We note that (16) is identical to (3); the functional form of the solution with uncertain constraints is the same as that for exact constraints. The difference is that, for uncertain constraints, the conditions that determine the β_r have the general form (18). These conditions reduce to the exact constraint case for $\varepsilon = 0$.

III. APPLICATION TO SPECTRUM ANALYSIS

As was the case for the multisignal RESA derivations in [4] and [5], for each of the L signals we use a discrete-spectrum approximation

$$s_i(t) = \sum_{k=1}^N (a_{ik} \cos 2\pi f_k t + b_{ik} \sin 2\pi f_k t)$$

($i = 1, \dots, L$) with nonzero frequencies f_k , not necessarily uniformly spaced. We treat the coefficients a_{ik} and b_{ik} as random variables with independent, zero-mean, Gaussian initial distributions. We define random variables

$$x_{ik} = \frac{1}{2}(a_{ik}^2 + b_{ik}^2) \quad (22)$$

representing the power of process s_i at frequency f_k , and we describe the collection of L signals in terms of the joint probability density $q^\dagger(\mathbf{x})$, where $\mathbf{x} = (x_1, \dots, x_K)$ and $x_i = (x_{i1}, \dots, x_{iN})$. We express the power spectrum S as an expectation

$$S_i(f_k) = \int x_{ik} q^\dagger(\mathbf{x}) d\mathbf{x}. \quad (23)$$

In deriving results for the case of exact constraints, we chose a prior probability density p so that

$$\int x_{ik} p(\mathbf{x}) d\mathbf{x} = P_i(f_k) \quad (24)$$

held, we re-wrote the autocorrelation constraints (7) as expressions in terms of expectations of $q(\mathbf{x})$ (the MRE final probability density), we derived $q(\mathbf{x})$ from the general solution (3), and we calculated the final power spectra $Q_i(f_k)$ from (23) (with Q replacing S and q replacing q^T). The result was (8) or (9) depending on whether weights were included [4, 5].

In the case of uncertain constraints, the constraints (7) are replaced by

$$\left| \sum_i \left(\int C_r(f) Q_i(f) df - R_r^{\text{tot}} \right)^2 \right| \leq \varepsilon^2 \quad (25)$$

The derivation of the final estimates follows the steps outlined above and is a straightforward generalization of the derivations in [4] and [5]. The detailed derivation will be contained in a forthcoming paper. Here we just state that the solution forms for uncertain constraints are the same as for exact constraints, i.e. (8) or (9). For the conditions on the β_r , however, we obtain

$$\varepsilon \frac{\beta_r}{\|\beta\|} = \int C_r(f) Q(f) df - R_r^{\text{tot}} \quad (26)$$

instead of (7). In the case of the more general constraint form

$$\sum_r M_{rs} v_r v_s \leq \varepsilon^2,$$

$$v_r = \sum_i \int C_r(f) Q_i(f) df - R_r^{\text{tot}},$$

it is convenient to carry the matrix through the derivation rather than transforming the constraint functions as in (20). The result is that the final estimates again have the form (8) or (9), while the conditions (26) on the β_r are replaced by

$$\varepsilon \frac{\beta_r}{(\beta^T M \beta)^{1/2}} = \int C_r(f) Q(f) df - R_r^{\text{tot}},$$

where

$$\beta^T = M^{-1} \beta$$

IV. EXAMPLE

We shall use a numerical example from [4, 5]. We define a pair of spectra, S_B and S_A , which we think of as a known "background" component and an unknown "signal" component of a total spectrum. Both are symmetric and defined in the frequency band from -0.5 to +0.5, though we plot only their positive-frequency parts. S_B is the sum of white noise with total power 5 and a peak at frequency 0.215 corresponding to a single sinusoid with total power 2. S_B consists of a peak at frequency 0.165 corresponding to a sinusoid of total power 2. Figure 1 shows a discrete-frequency approximation to the sum $S_B + S_A$, using 100 equispaced frequencies. From the sum, six autocorrelation were computed exactly. S_B itself was used as the initial estimate P_B of S_B — i.e., P_B was Figure 1 without the left-hand peak. For P_A we used a uniform (flat) spectrum with the same total power as P_B . Figures 2 and 3 show unweighted multisignal RESA final estimates Q_B and Q_A [4]. The signal peak shows up primarily in Q_A , but some evidence of it is in Q_B as well. This is reasonable since P_B , although exactly correct, is treated as an initial estimate subject to change by the data. The signal peak can be suppressed from Q_B and enhanced in Q_A by weighting the background estimate P_B heavily [5].

In Figure 4 we show final estimates for uncertain constraints with an error bound of $\varepsilon=1$. The Euclidean distance (i.e., a constraint of the form (25)) was used. The estimates were obtained with Newton-Raphson algorithms similar to those developed by Johnson [13]. Both final estimates in Figure 4 are closer to the corresponding initial estimates than is the case in Figures 2 and 3, since the sum of the final estimates is no longer constrained to satisfy the autocorrelations. Figure 5 shows results for $\varepsilon=3$; the final estimates are even closer to the initial estimates. Because the example was constructed with exactly-known autocorrelations, it's not surprising that the exactly-constrained final estimates are better than those in Figures 4 and 5, which illustrate the more conservative

deviation from initial estimates that results from incorporating uncertain constraints.

V. DISCUSSION

It appears that Ables was the first to suggest using an uncertain constraint of the Euclidean form (25) in MESA [14]. The use of this and a weighted Euclidean constraint in MESA has been studied by Newman [15, 16], and the generalization to general, matrix constraints has been studied by Schott and McClellan [17]. Uncertain constraints have also been used in applying maximum entropy to image processing [18, 19], although with a different entropy expression [20]. The results presented herein generalize these previous results in two main respects: treatment of the multisignal case and inclusion of initial estimates.

A pleasant property of the new estimator, both in the general probability density form and in the power spectrum form, is that the functional form is unchanged from the exact constraint case. For the power spectrum estimator, this means that resulting final estimates are still all-pole spectra whenever the initial estimates are all-pole and the weights are frequency-independent.

REFERENCES

1. J. P. Burg, "Maximum entropy spectral analysis," presented at the 37th Annual Meeting Soc. of Exploration Geophysicists, Oklahoma City, Okla. (1967).
2. J. P. Burg, "Maximum Entropy Spectral Analysis," Ph.D. Dissertation, Stanford University, Stanford, California (1975). (University Microfilms No. 75-25,499)
3. J. E. Shore, "Minimum cross-entropy spectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing ASSP-29*, pp. 230-237 (Apr. 1981).
4. 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).
5. 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*, to be published.
6. I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *Ann. Prob.* 3, pp. 146-158 (1975).
7. J. E. Shore and R. W. Johnson, "Properties of cross-entropy minimization," *IEEE Trans. Inform. Theory IT-27*, pp. 472-482 (July 1981).
8. S. Kullback, *Information Theory and Statistics*, Dover, New York (1969). Wiley, New York, 1959
9. I. J. Good, "Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables," *Ann. Math. Stat.* 34, pp. 911-934 (1963).
10. E. T. Jaynes, "Prior probabilities," *IEEE Trans. Systems Science and Cybernetics SSC-4*, pp. 227-241 (1968).
11. R. W. Johnson, "Axiomatic characterization of the directed divergences and their linear combinations," *IEEE Trans. Inf. Theory IT-25*, pp. 709-716 (Nov. 1979).
12. 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, (See also comments and corrections in *IEEE Trans. Inform. Theory IT-29*, Nov. 1983) (Jan. 1980).
13. R. W. Johnson, "Algorithms for single-signal and multisignal minimum-cross-entropy spectral analysis," NRL Report 8667, Naval Research Laboratory, Washington, D.C. (1983). (submitted to *IEEE Trans. Acoustics, Speech, Signal Processing*.)
14. J. G. Ables, "Maximum entropy spectral analysis," *Astron. Astrophys. Suppl.* 15, pp. 383-393 (1974).

15. W. I. Newman, "Extension to the Maximum Entropy Method," *IEEE Trans. Inform. Theory* IT-23, pp. 89-708 (January 1977).
16. W. I. Newman, "Extension to the Maximum Entropy Method III," *Proc. First ASSP Workshop on Spectral Estimation*, pp. 3.2.1-3.2.7 (Aug. 1981).
17. J.-P. Schott and J. H. McClellan, "Maximum entropy power spectrum estimation with uncertainty in correlation measurements," *Proc. ICASSP 83*, pp. 1068-1071, IEEE (April 1983).
18. S. F. Gull and G. J. Daniell, "Image reconstruction from incomplete and noisy data," *Nature* 272, pp. 686-670 (April, 1978).
19. J. Skilling, "Maximum entropy and image processing — algorithms and applications," *Proc. First Maximum Entropy Workshop* (1981).
20. J. E. Shore, "Inversion as logical inference — theory and applications of maximum entropy and minimum cross-entropy," *Proceedings American Mathematical Society Symposium on Inverse Problems* (13 April 1983).

Fig. 1. Sum $S_B + S_A$ of original spectra.

Fig. 2. MRESA final estimate Q_B of "background" S_B .

Fig. 3. MRESA final estimate Q_A of "signal" S_A .

Fig. 4. Final estimates Q_B and Q_A with $\varepsilon=1$.

Fig. 5. Final estimates Q_B and Q_A with $\varepsilon=3$.