Multiple Window Based Minimum Variance Spectrum Estimation for Multidimensional Random Fields

Tsung-Ching Liu and Barry D. Van Veen, *Member, IEEE*

Abstract—Spectrum analysis is viewed as the problem of estimating the power of a process contained within narrow bands. This view leads naturally to consideration of filter-based methods for estimating spectra. We consider multiple window based estimators where the power in a band is estimated as the average of the powers estimated at the outputs of multiple filters. The filters are designed using a linearly constrained minimum variance criterion commonly employed in adaptive beamforming. This results in filters that automatically adjust their sidelobes to minimize leakage of energy from outside the band of interest. Expressions for the bias and variance of the power estimates are derived assuming the sample covariance matrix estimate is utilized to estimate the data covariance matrix and that the data is independent and identically Gaussian distributed. These expressions lead to the definition of a performance factor that indicates the degree of variance reduction obtained via multiple window processing. Lastly, we present a technique for obtaining increased suppression of energy leaking through the filter sidelobes at the expense of the response fidelity to energy in the band. Simulations are presented to illustrate the effectiveness of the technique.

I. INTRODUCTION

SPECTRUM analysis is the problem of determining the power of a process in narrow frequency bands. This perspective leads to the consideration of methods which pass the signal through narrow-band filters and use the power at the filter outputs as the spectrum estimate. A large number of spectrum estimation methods fit into this framework [1]–[6]. The paper by Thomson [3] is particularly relevant to the work presented here. Thomson estimates power as the average of several windowed Fourier transforms of the data. This is equivalent to averaging the outputs of several narrow-band filters, where the frequency response of each filter is determined by the corresponding window. The purpose of averaging is to reduce the variance of the estimated power. Averaging techniques had been used for some time prior to Thomson (see [5], [6]); however, the windows proposed by Thomson result in much greater variance reduction than other

methods with similar resolution. Bronez [4] has recently extended Thomson's approach to spectrum estimation problems which do not involve uniformly spaced sampling. These methods are commonly referred to as multiple window methods.

This paper presents a multiple window spectrum estimation procedure in which the windows are data dependent. The windows are chosen such that the corresponding filters have small sidelobes in regions of the spectrum exterior to the band of interest that have large levels of power. This greatly reduces sidelobe leakage and thus gives a less biased estimate of power. Thomson recognized the sidelobe leakage problem and suggested an adaptive weighting procedure in which the windows with higher sidelobes were deemphasized if there was significant energy exterior to the band of interest. This has the effect of reducing the number of windows, with a corresponding loss in variance reduction by averaging. Here the number of windows is held constant to maintain variance reduction properties, and the response of each window is modified to minimize sidelobe leakage. This approach is essentially a multiple window equivalent to Capon's [2] minimum variance estimator. The disadvantage of this method is the requirement of multiple data vectors for covariance matrix estimation.

In Section II we derive a general framework for quadratic estimators of the power spectrum based on the generalized spectral representation for homogeneous random fields. Bias and variance properties of quadratic estimators are discussed. The multiple window minimum variance (MWMV) estimator is developed in Section III. The MWMV estimator minimizes the power at the filter outputs subject to a set of linear constraints on the filter responses in the band of interest. Methods for designing the linear constraints are presented. A statistical analysis of the mean and variance of the MWMV estimator is given in Section IV. We show that at best the MWMV estimator reduces the variance by the number of windows, while at worst the variance is equivalent to that obtained with single window processing. In Section V a method is presented for designing constraints that trade response fidelity in the band of interest for additional reductions in sidelobe leakage. Simulations are presented in Section VI for data collected at a sensor array. The goal here is to

Manuscript received August 24, 1990; revised April 2, 1991. This work was supported in part by the Army Research Office under Grant DAAL03-89-K-0141 and by the National Science Foundation under Award MIP-8958559.

The authors are with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706.

IEEE Log Number 9105657.

estimate the incident power in a specified frequency band as a function of direction. Section VII contains a summary.

II. QUADRATIC SPECTRUM ESTIMATORS

In this section we develop a class of spectrum estimators based on quadratic forms of the data. The section begins by describing a homogeneous random field model for the data. This model is very general and allows application of this work to a wide variety of spectrum estimation problems. The basic quadratic estimator is then introduced and its properties examined. Notationally, boldface lowercase and uppercase symbols denote vector and matrix quantities, respectively. The superscript t denotes transpose and superscript h denotes complex conjugate transpose. $\text{tr } A$ denotes trace of the matrix A .

A. The Homogeneous Random Field Model

Let $x(t) = x(t_1, \dots, t_n)$ denote a random field represented over the n -dimensional space $\mathcal{R}^n$ of points $t = (t_1, \dots, t_n)^t$. t represents variables over which we observe the data. In the case of a time series, we have $n = 1$ and $t = t_1$ represents the "time" variable. If $x(t)$ corresponds to a spatially propagating field, then n may vary from one to four since we can use an array of sensors to sample the field in from one to three spatial dimensions in addition to time. Here t represents the temporal and spatial variables. In an image processing system t represents the two spatial coordinates and time if a sequence of images is of interest. The field $x(t)$ is said to be homogeneous if its mean $\mathcal{E}(x(t))$ is constant and its correlation function $\mathcal{E}(x(t_1)x^*(t_2)) = R(t_1, t_2)$ depends only on the vector $\tau = t_1 - t_2$, so that $R(t_1, t_2) = R(t_1 - t_2)$.

Every homogeneous random field $x(t)$ admits a generalized spectral representation [7] of the form

$$x(t) = \int_{\mathcal{K}^n} e^{ik^t t} Z(dk). \quad (1)$$

Here $k = (k_1, \dots, k_n)^t$ are the independent variables for the spectrum of the field, $\mathcal{K}^n$ is the n -dimensional space on which these parameters are defined, $dk = dk_1, \dots, dk_n$ is a volume element in $\mathcal{K}^n$, and $Z(\Delta)$ is a complex random orthogonal increments measure on $\mathcal{K}^n$. If the data is observed over time, then $k = \omega$ represents temporal frequency. If a propagating field is of interest, then k represents spatial frequencies. We will refer to k as "frequency" with the understanding that this term is used in a general sense. We assume the field is zero mean Gaussian distributed so $\mathcal{E}(Z(dk)) = 0$ and $\mathcal{E}(Z(dk)Z^*(dk')) = S(k) dk$ where $S(k)$ is the power spectrum of the process.

Let x be an N -dimensional vector whose elements are the random field $x(t)$ sampled at t_i , $i = 1, 2, \dots, N$. This data vector has spectral representation

$$x = \begin{bmatrix} x(t_1) \\ x(t_2) \\ \vdots \\ x(t_N) \end{bmatrix} = \int_{\mathcal{K}^n} \begin{bmatrix} e^{ik^t t_1} \\ e^{ik^t t_2} \\ \vdots \\ e^{ik^t t_N} \end{bmatrix} Z(dk). \quad (2)$$

Define

$$\phi(k) = \begin{bmatrix} e^{ik^t t_1} \\ e^{ik^t t_2} \\ \vdots \\ e^{ik^t t_N} \end{bmatrix}$$

as the "response vector." The motivation for this definition is as follows. Consider the inner product of the data vector with an N -dimensional weight vector w

$$\begin{aligned} y &= w^h x \\ &= \int_{\mathcal{K}^n} w^h \phi(k) Z(dk) \\ &= \int_{\mathcal{K}^n} W(k) Z(dk). \end{aligned} \quad (3)$$

The variance of y is given by

$$\begin{aligned} P &= \mathcal{E}(|y|^2) \\ &= \int_{\mathcal{K}^n} \int_{\mathcal{K}^n} W(k) W^*(k') \mathcal{E}(Z(dk)Z^*(dk')) \\ &= \int_{\mathcal{K}^n} |W(k)|^2 S(k) dk. \end{aligned} \quad (4)$$

P is the power in a process with power spectrum $|W(k)|^2 S(k)$. $W(k)$ has the effect of filtering the process and its frequency response is given by the inner product of w and $\phi(k)$.

B. Quadratic Spectrum Estimators

We view spectrum estimation as the problem of estimating the power in narrow frequency bands of a process. Let $\mathcal{G}^n$ represent the band in which we wish to estimate the power. Our goal is to compute the integrated spectrum $\int_{\mathcal{G}^n} S(k) dk$. Thus, $\mathcal{G}^n$ defines the desired resolution of the estimator. The integrated spectrum is obtained by passing the process through an ideal bandpass filter having passband $\mathcal{G}^n$ and computing the power at the filter output. If $W(k)$ in (4) represents the response of an ideal bandpass filter with passband $\mathcal{G}^n$, then P in (4) is the desired quantity. P can be rewritten as the quadratic form

$$\begin{aligned} P &= w^h \left[\int_{\mathcal{K}^n} \phi(k) \phi^h(k) S(k) dk \right] w \\ &= w^h R_x w \end{aligned} \quad (5)$$

where the data covariance matrix is given by $R_x = \int_{\mathcal{K}^n} \phi(k) \phi^h(k) S(k) dk = \mathcal{E}(xx^h)$.

It is well known that an ideal bandpass filter cannot be constructed given a finite length data record, so we choose w such that $W(k)$ approximates the ideal response

$$|W(k)|^2 \approx \begin{cases} 1 & \text{for } k \in \mathcal{G}^n \\ 0 & \text{for } k \notin \mathcal{G}^n. \end{cases} \quad (6)$$

The approximation results in two types of errors when the integrated spectrum is computed using (5). There is an *interior bias* resulting from the approximation error interior to the band of interest, i.e., $k \in \mathcal{G}^n$, and an *exterior bias* resulting from the approximation error exterior to the band of interest, i.e., $k \notin \mathcal{G}^n$. The interior bias causes an error due to nonequal weighting of spectral energy in the band $\mathcal{G}^n$ while the exterior bias results in leakage of energy from outside $\mathcal{G}^n$. Thomson [3] refers to these errors as "local" and "broad-band" biases, respectively.

An additional source of error in using P to determine the power in $\mathcal{G}^n$ arises due to the fact that R_x is not known and must be estimated from the available data. For example, given M data vectors x_i , $i = 1, 2, \dots, M$ one common estimate of R_x is the sample covariance matrix $\hat{R}_x = (1/M) \sum_{i=1}^M x_i x_i^h$. Here $\hat{\cdot}$ denotes estimated quantities. Let $\hat{P}$ be the power estimate when R_x is replaced by $\hat{R}_x$. $\hat{R}_x$ is a matrix of random variables so $\hat{P}$ is a random variable. The quality with which $\hat{P}$ represents P is often measured in terms of bias and variance. Note that the interior and exterior biases defined in the previous paragraphs are deterministic effects. We shall use the term "bias" alone to represent statistical bias, i.e., $\mathcal{E}(\hat{P}) - P$.

The variance properties of quadratic estimators are often unsatisfactory, so averaging techniques are used to reduce variance. Examples of averaging methods include time domain averaging [5], frequency domain averaging [6], and orthogonal window averaging [3], [4]. The basic idea is to obtain estimates of power that are statistically independent and then average them to produce a new estimate with lower variance than the individual estimates. Almost all averaging techniques can be placed in a multiple window framework with different methods using different criteria for choosing the windows. The term "filter" is more appropriate here than "window"; however, we use "window" for consistency with previous work. Define $y_i = w_i^h x$ as the output of the i th filter and the multiple window power P_m as

$$\begin{aligned} P_m &= \mathcal{E} \left(\sum_{i=1}^r |y_i|^2 \right) \\ &= \int_{\mathcal{K}^n} \sum_{i=1}^r |W_i(k)|^2 S(k) dk \\ &= \text{tr } W^h R_x W \end{aligned} \quad (7)$$

where $W = [w_1, w_2, \dots, w_r]$. Note that (7) implies $\sum_{i=1}^r |W_i(k)|^2$ must approximate the magnitude response of an ideal bandpass filter

$$\sum_{i=1}^r |W_i(k)|^2 \approx \begin{cases} 1 & \text{for } k \in \mathcal{G}^n \\ 0 & \text{for } k \notin \mathcal{G}^n \end{cases} \quad (8)$$

in order to determine the power in the band $k \in \mathcal{G}^n$.

The exterior bias " B_e " for a multiple window estimator is given by

$$B_e = \int_{k \notin \mathcal{G}^n} \sum_{i=1}^r |W_i(k)|^2 S(k) dk. \quad (9)$$

Define $B_i(k)$, k in $\mathcal{G}^n$ to be the interior bias on $\mathcal{G}^n$

$$B_i(k) = \sum_{i=1}^r |W_i(k)|^2 - 1, \quad \text{for } k \in \mathcal{G}^n. \quad (10)$$

The cumulative interior bias B_i is given by

$$B_i = \int_{\mathcal{K}^n} B_i(k) S(k) dk \quad (11)$$

and is dependent on the spectrum $S(k)$ in $\mathcal{G}^n$. $S(k)$ is usually unknown so we measure the bias using $B_i(k)$.

The degree to which averaging reduces variance is dependent on the degree of independence among the y_i and the relative values of $\mathcal{E}(|y_i|^2)$. The y_i are Gaussian random variables and thus are independent if they are uncorrelated. The correlation between y_i and y_j is given by

$$\begin{aligned} \mathcal{E}(y_i y_j^*) &= \int_{\mathcal{K}^n} \int_{\mathcal{K}^n} W_i(k) W_j^*(k) \mathcal{E}(Z(dk) Z^*(dk)) \\ &= \int_{\mathcal{K}^n} W_i(k) W_j^*(k) S(k) dk. \end{aligned} \quad (12)$$

Assuming the variances $\mathcal{E}(|y_i|^2)$ are fixed, the greatest variance reduction is obtained when all y_i are uncorrelated, i.e.,

$$W^h \left[\int_{\mathcal{K}^n} \phi(k) \phi^h(k) S(k) dk \right] W = W^h R_x W = \Lambda \quad (13)$$

where Λ is a diagonal matrix with diagonal elements $\lambda_i = \mathcal{E}(|y_i|^2)$. The maximum variance reduction occurs when all the λ_i are equal.

We now state a general criterion for choosing W in the multiple window spectrum estimation problem. We seek filters with averaged gain of unity on the frequency band of interest, with sidelobe structures that minimize leakage, and that maximize the variance reduction obtained by averaging. That is, W should be chosen so that $\hat{P}_m$ is unbiased, (13) is satisfied to maximize the variance reduction obtained by averaging, and both B_e and $|B_i(k)|^2$ are in some sense minimized. The estimator proposed in the following section is designed to minimize B_e subject to a constraint on $|B_i(k)|^2$. We show that it approximately satisfies (13) and, in Section IV, that it is unbiased (after a simple scaling operation).

Traditional windows that are used for multiple window spectrum estimation are usually designed to give uncorrelated y_i . Time domain averaging [5] and frequency domain averaging [6] methods approximately satisfy (13); however, the width of $\sum_{i=1}^r |W_i(k)|^2$ as a function of k increases proportional to the number (r) of windows used. This results in poor resolution. The width of $\sum_{i=1}^r |W_i(k)|^2$ expands at a much slower rate for the windows proposed by Thomson [3] and Bronez [4]. Their windows satisfy (13) when $S(k)$ is constant for $k \in \mathcal{G}^n$ and $k \notin \mathcal{G}^n$ (not necessarily the same constant). However, the sidelobes of Thomson's and Bronez's windows are fixed and do not satisfy (13) or minimize leakage when $S(k)$, $k \notin \mathcal{G}^n$ is not constant. The method proposed in this paper leads to win-

dows that approximately satisfy (13) in a more general sense than those of [3] and [4] while maintaining about the same width of $\sum_{i=1}^r |W_i(\mathbf{k})|^2$ and minimizing leakage of energy through the window sidelobes.

III. THE MULTIPLE WINDOW MINIMUM VARIANCE ESTIMATOR

The multiple window minimum variance (MWMV) estimator basically originates with the minimum variance spectrum estimator of Capon [2] (often termed Capon's MLM) and is closely related to linearly constrained minimum variance beamforming [8], [9]. The term "minimum variance" is misleading since it does not refer to minimization of statistical variance associated with the estimator $\hat{P}_m$, but instead refers to minimization of power P_m . We use "minimum variance" for consistency with the beamforming literature and interpretations of Capon's work. Note that (7) implies

$$\begin{aligned} P_m &= \int_{\mathcal{G}^n} \sum_{i=1}^r |W_i(\mathbf{k})|^2 S(\mathbf{k}) d\mathbf{k} \\ &\quad + \int_{\mathbf{k} \notin \mathcal{G}^n} \sum_{i=1}^r |W_i(\mathbf{k})|^2 S(\mathbf{k}) d\mathbf{k} \\ &= \int_{\mathcal{G}^n} S(\mathbf{k}) d\mathbf{k} + \int_{\mathcal{G}^n} \left(\sum_{i=1}^r |W_i(\mathbf{k})|^2 - 1 \right) \\ &\quad \cdot S(\mathbf{k}) d\mathbf{k} + B_e \\ &= \int_{\mathcal{G}^n} S(\mathbf{k}) d\mathbf{k} + \int_{\mathcal{G}^n} B_i(\mathbf{k}) S(\mathbf{k}) d\mathbf{k} + B_e. \end{aligned} \quad (14)$$

The first term in (14) is the desired quantity, the power in the band $\mathcal{G}^n$, and the second term is B_i . Multiple linear constraints on $\mathbf{W}$ are utilized to constrain the interior bias $B_i(\mathbf{k})$. Thus, minimization of P_m is equivalent to minimization of B_e .

The MWMV criterion is formally stated as follows:

$$\min_{\mathbf{W}} \text{tr } \mathbf{W}^h \mathbf{R}_x \mathbf{W}, \quad \text{subject to } \mathbf{U}^h \mathbf{W} = \mathbf{G} \quad (15)$$

where $\mathbf{U}$ is the $N \times L$ constraint matrix and $\mathbf{G}$ is the $L \times r$ response matrix. An approach for determining $\mathbf{U}$ and $\mathbf{G}$ is discussed in the following paragraphs. Solving (15) for $\mathbf{W}$ and substituting in (7) gives

$$P_m = \text{tr } \mathbf{G}^h (\mathbf{U}^h \mathbf{R}_x^{-1} \mathbf{U})^{-1} \mathbf{G}. \quad (16)$$

A. Constraint Design

The constraints are design to control the interior bias $B_i(\mathbf{k})$. Recall $B_i(\mathbf{k}) = \sum_{i=1}^r |W_i(\mathbf{k})|^2 - 1$ for $\mathbf{k} \in \mathcal{G}^n$ so we desire

$$\text{tr } \mathbf{W}^h \phi(\mathbf{k}) \phi^h(\mathbf{k}) \mathbf{W} \approx 1 \quad \text{for } \mathbf{k} \in \mathcal{G}^n. \quad (17)$$

Equation (17) indicates that the component of $\mathbf{W}$ in the space spanned by $\{\phi(\mathbf{k}), \mathbf{k} \in \mathcal{G}^n\}$ must be controlled. The constraint $\mathbf{U}^h \mathbf{W} = \mathbf{G}$ forces the coordinates of $\mathbf{W}$ in the space spanned by the columns of $\mathbf{U}$ to equal $\mathbf{G}$ and thus

controls the component of $\mathbf{W}$ in the space spanned by the columns of $\mathbf{U}$. This implies that the columns of $\mathbf{U}$ and $\{\phi(\mathbf{k}), \mathbf{k} \in \mathcal{G}^n\}$ must span the same space. Define

$$\mathbf{H} = \int_{\mathcal{G}^n} \phi(\mathbf{k}) \phi^h(\mathbf{k}) d\mathbf{k} \quad (18)$$

and let $\mathbf{H}$ have eigen-decomposition $\tilde{\mathbf{U}} \tilde{\boldsymbol{\Lambda}} \tilde{\mathbf{U}}^h$. The eigenvectors in $\tilde{\mathbf{U}}$ corresponding to nonzero eigenvalues provide an orthonormal basis for the space spanned by $\{\phi(\mathbf{k}), \mathbf{k} \in \mathcal{G}^n\}$. Therefore, the constraint matrix $\mathbf{U}$ must consist of the eigenvectors in $\tilde{\mathbf{U}}$ corresponding to the $L(\leq N)$ "significant" eigenvalues of $\mathbf{H}$. The meaning of "significant eigenvalue" can be clarified by noting that $\tilde{\mathbf{u}}_i^h \mathbf{H} \tilde{\mathbf{u}}_i = \tilde{\lambda}_i$ where $\tilde{\mathbf{u}}_i$ and $\tilde{\lambda}_i$ are the i th eigenvector and eigenvalue, respectively. This implies

$$\max_{\mathbf{k} \in \mathcal{G}^n} |\tilde{\mathbf{u}}_i^h \phi(\mathbf{k})|^2 \leq \tilde{\lambda}_i. \quad (19)$$

That is, the size of the eigenvalue determines the maximum possible projection of $\{\phi(\mathbf{k}), \mathbf{k} \in \mathcal{G}^n\}$ onto the space spanned by the associated eigenvector. Thus, the components of $\mathbf{W}$ in the space spanned by eigenvectors corresponding to very small or zero eigenvalues do not need to be controlled because they do not have a significant effect on the response $\mathbf{W}^h \phi(\mathbf{k}), \mathbf{k} \in \mathcal{G}^n$.

The number of significant eigenvalues can be determined without actually computing the eigenstructure of $\mathbf{H}$ for several special cases [3], [9]. If the data is from a uniformly sampled time series and $\mathcal{G}^n$ is a symmetric frequency band about zero, then the eigenvectors of $\mathbf{H}$ are the discrete prolate spheroidal sequences and the number of significant eigenvalues is given by the time bandwidth product [10], [11]. The number of significant eigenvalues of $\mathbf{H}$ is also given by a time bandwidth product for data received at sensor arrays under certain conditions [9]. In general the eigenvalues of $\mathbf{H}$ must be numerically evaluated and compared to a threshold to determine how many are significant.

The dependence of the response on $\mathbf{G}$ is made evident by decomposing $\mathbf{W}$ into constrained and unconstrained components. Let the columns of $\tilde{\mathbf{U}}$ be a basis for the $N - L$ dimensional space orthogonal to the space spanned by the columns of $\mathbf{U}$, i.e., $\mathbf{U}^h \tilde{\mathbf{U}} = \mathbf{0}$ and $[\mathbf{U} \tilde{\mathbf{U}}]$ is nonsingular. A $\mathbf{W}$ that satisfies the constraint must be of the form $\mathbf{W} = \mathbf{U}\mathbf{G} + \tilde{\mathbf{U}}\mathbf{W}_o$ where $\mathbf{W}_o$ is $(N - L) \times r$. Clearly, $\mathbf{U}^h \mathbf{W} = \mathbf{G}$ since $\mathbf{U}^h \mathbf{U} = \mathbf{I}$ and $\mathbf{U}^h \tilde{\mathbf{U}} = \mathbf{0}$. The components of $\mathbf{W}$ in the space spanned by the columns of $\tilde{\mathbf{U}}$ are determined by $\mathbf{W}_o$ and are unconstrained. This type of decomposition is known in the beamforming community as the generalized sidelobe canceller (GSC) [12]. Now, since the columns of $\mathbf{U}$ and $\{\phi(\mathbf{k}), \mathbf{k} \in \mathcal{G}^n\}$ span the same space, the columns of $\tilde{\mathbf{U}}$ are orthogonal to $\{\phi(\mathbf{k}), \mathbf{k} \in \mathcal{G}^n\}$, i.e., $\tilde{\mathbf{U}}^h \phi(\mathbf{k}) = \mathbf{0}$ for $\mathbf{k} \in \mathcal{G}^n$, and $\mathbf{W}^h \phi(\mathbf{k}) = \mathbf{G}^h \mathbf{U}^h \phi(\mathbf{k})$ for $\mathbf{k} \in \mathcal{G}^n$. The problem is to choose $\mathbf{G}$ so that $\text{tr } \mathbf{W}^h \phi(\mathbf{k}) \phi^h(\mathbf{k}) \mathbf{W} = \phi^h(\mathbf{k}) \mathbf{W} \mathbf{W}^h \phi(\mathbf{k}) = \phi^h(\mathbf{k}) \mathbf{U} \mathbf{G} \mathbf{G}^h \mathbf{U}^h \phi(\mathbf{k}) \approx 1$ for $\mathbf{k} \in \mathcal{G}^n$.

A natural approach is to choose $\mathbf{G}$ to minimize the mean

square error over $k \in \mathcal{G}^n$

$$\min_G \int_{\mathcal{G}^n} |\phi^h(k) U G G^h U^h \phi(k) - 1|^2 dk. \quad (20)$$

A closed form expression for the solution to (20) is not evident so numerical optimization techniques must be utilized. However, a good approximate solution is obtained by any G which satisfies $GG^h = (1/N) \text{diag}(1, 1, \dots, 1, 0, 0, \dots, 0)$.

Let the number of nonzero elements on the diagonal of GG^h be equal to the number of windows (r) and define $\Phi = [\phi(k_1), \phi(k_2), \dots, \phi(k_T)]$ where $k_i, i = 1, \dots, T$ densely sample $\mathcal{G}^n$. The constraint is rewritten in terms of Φ as $[\Phi^h U G G^h U^h \Phi]_{ii} \approx 1$. Note that $H \approx \Phi \Phi^h \Delta k$. This implies the eigenvectors of H , the columns of $\tilde{U}$, correspond to the left singular vectors of Φ . Define the singular value decomposition of Φ as

$$\Phi = \tilde{U} \Sigma V^h. \quad (21)$$

The squares of the singular values correspond to the eigenvalues of H . Now if $GG^h = (1/N) \text{diag}(1, 1, \dots, 1, 0, 0, \dots, 0)$, then

$$\begin{aligned} \phi^h(k_i) U G G^h U^h \phi(k_i) &= [\Phi^h U G G^h U^h \Phi]_{ii} \\ &\approx \frac{1}{N} [V \Sigma_q^h \Sigma_q V^h]_{ii}, \quad i = 1, \dots, T \end{aligned} \quad (22)$$

where Σ_q is the rank $q = \min(r, L)$ approximation to Σ , that is, Σ_q is obtained by zeroing the $q + 1^{\text{th}}$ through N^{th} diagonal elements of Σ . If q is chosen so that the discarded singular values are sufficiently small, then $V \Sigma_q^h \Sigma_q V^h \approx \Phi^h \Phi$ and

$$\begin{aligned} \phi^h(k_i) U G G^h U^h \phi(k_i) &\approx \frac{1}{N} [\Phi^h \Phi]_{ii} \\ &= \frac{1}{N} \phi^h(k_i) \phi(k_i) = 1, \quad i = 1, \dots, T. \end{aligned} \quad (23)$$

L is chosen to represent the number of the significant eigenvalues of H , and also represents the number of significant singular values of Φ . Thus, the best response approximation is obtained by choosing $r = L$, so that $q = L$.

B. MWMV Properties

The MWMV estimator approximately satisfies (13) when $S(k), k \in \mathcal{G}^n$ is constant and B_e is small. Note that

$$\begin{aligned} W^h R_x W &= W^h \left[\int_{\mathcal{G}^n} \phi(k) \phi^h(k) S(k) dk \right] W \\ &+ W^h \left[\int_{k \notin \mathcal{G}^n} \phi(k) \phi^h(k) S(k) dk \right] W. \end{aligned} \quad (24)$$

The second term is related to the exterior bias B_e . If we assume that the minimization makes B_e small, then $W^h \phi(k)$ is small for k with significant $S(k)$, $k \notin \mathcal{G}^n$, and the second term is negligible. Thus,

$$W^h R_x W \approx W^h \left[\int_{\mathcal{G}^n} \phi(k) \phi^h(k) S(k) dk \right] W. \quad (25)$$

In a previous paragraph we used the GSC decomposition to show that $W^h \phi(k) = G^h U^h \phi(k)$ for $k \in \mathcal{G}^n$. Substituting UG for W in (25) yields

$$W^h R_x W \approx G^h U^h \left[\int_{\mathcal{G}^n} \phi(k) \phi^h(k) S(k) dk \right] UG. \quad (26)$$

If $S(k)$ is constant on $k \in \mathcal{G}^n$, then we have

$$W^h R_x W \approx G^h U^h H UG. \quad (27)$$

The columns of U are the eigenvectors of H so $G^h U^h H UG$ is diagonal if G is diagonal and the MWMV outputs are approximately independent. The MWMV outputs will not in general be independent if B_e is large or if $S(k), k \in \mathcal{G}^n$ is not constant. Equation (27) also indicates that no additional variance reduction is obtained by choosing more windows than constraints ($r > L$). If $r > L$, then the $L + 1^{\text{st}}$ through r^{th} diagonal elements of $G^h U^h H UG$ are zero which indicates that the outputs of the $L + 1^{\text{th}}$ through r^{th} windows do not contain any power associated with $k \in \mathcal{G}^n$.

The multiple window methods of Thomson and Bronez provide independent outputs if $S(k)$ is constant for $k \in \mathcal{G}^n$, and is also constant for $k \notin \mathcal{G}^n$. Assuming $S(k), k \notin \mathcal{G}^n$, is constant is unrealistic for many problems. Independent outputs are obtained in the MWMV method if $S(k), k \in \mathcal{G}^n$ is constant and B_e is small. The class of $S(k)$ for which B_e is small is much larger than the class of $S(k)$ which is constant on $k \notin \mathcal{G}^n$ because B_e is minimized. Thus, the MWMV method offers potentially greater variance reduction due to averaging.

The MWMV estimator for a uniformly sampled time series is equivalent to Thomson's multiple window estimator if $R_x = I$ and G is diagonal. In this case $W = UG$ and the columns of U contain the discrete prolate spheroidal sequences (modulated by a complex sinusoid if a frequency band other than that centered at zero is of interest). The estimate $\hat{P} = (1/M) \sum_{i=1}^M \text{tr} G^h U^h x_i x_i^h UG$ represents a weighted combination of the projections of the data onto the columns of U , with the diagonal elements of G determining the weights.

Thomson's "adaptive weighting" procedure weights the projections in order to reduce the exterior bias at the expense of the interior bias. These weights correspond to the diagonal elements of GG^h . They tend to be either zero or one and thus determine the number of windows which are used. Note that the MWMV estimator adapts the shape of the windows based on the data, and hence is capable of averaging more windows for the same level of exterior bias.

IV. STATISTICAL ANALYSIS OF THE MWMV ESTIMATOR

In practice $\mathbf{R}_x$ is unknown and must be estimated from the available data. The ML estimate of $\mathbf{R}_x$ given M data vectors $\mathbf{x}_i$, $i = 1, 2, \dots, M$ which are independent and identically distributed with $\mathcal{P}(\mathbf{x}_i) = \mathcal{N}(0, \mathbf{R}_x)$ is given by the sample covariance matrix $\hat{\mathbf{R}}_x = (1/M) \sum_{i=1}^M \mathbf{x}_i \mathbf{x}_i^H$. $M \hat{\mathbf{R}}_x$ is complex Wishart distributed [13]. Denote the estimate of P_m in (16) when $\mathbf{R}_x$ is replaced by $\hat{\mathbf{R}}_x$ as $\hat{P}_m$. The mean and variance of $\hat{P}_m$ are derived by Capon and Goodman in [14] for the single window, single linear constraint case. Here we derive the mean and variance for multiple window, multiple linear constraint estimators. Our results reduce to Capon and Goodman's as a special case.

Theorem 4.1: The mean and variance of the MWMV estimator $\hat{P}_m$ are

$$\begin{aligned} \mathcal{E}(\hat{P}_m) &= \frac{M - N + L}{M} \sum_{i=1}^r \lambda_i \\ &= \frac{M - N + L}{M} P_m \end{aligned} \quad (28)$$

and

$$\text{var}(\hat{P}_m) = \frac{M - N + L}{M^2} \sum_{i=1}^r \lambda_i^2 \quad (29)$$

where $\lambda_1, \dots, \lambda_r$ are the eigenvalues of $\mathbf{G}^h(\mathbf{U}^h \mathbf{R}_x^{-1} \mathbf{U})^{-1} \mathbf{G}$, $\mathbf{U}$ is $N \times L$, $\mathbf{G}$ is $L \times r$ and M is the number of data vectors used to estimate $\mathbf{R}_x$.

Proof: The derivation of mean and variance are based on the characteristic function of $\hat{P}_m$. The proof utilizes several properties of Wishart distributed matrices which are stated in the following paragraph. Although many of the results in the references are derived for real random matrices, equivalent results also apply to the complex case [15].

If $\mathbf{S} = \sum_{i=1}^M \mathbf{x}_i \mathbf{x}_i^H$ where $\mathbf{x}_i$ are N -dimensional Gaussian i.i.d. random vectors with covariance $\mathbf{R}_x$ then $\mathbf{S}$ is complex Wishart distributed with distribution denoted as $\mathcal{P}(\mathbf{S}) = CW(\mathbf{R}_x, N, M)$. If Ψ is an $N \times L$ rank L matrix, then

- 1) $\mathcal{P}(\Psi^H \mathbf{S} \Psi) = CW(\Psi^H \mathbf{R}_x \Psi, L, M)$ [16].
- 2) $\mathcal{P}((\Psi^H \mathbf{S}^{-1} \Psi)^{-1}) = CW((\Psi^H \mathbf{R}_x^{-1} \Psi)^{-1}, L, M - N + L)$ [16].
- 3) The characteristic function of $\text{tr } \mathbf{S}$ is

$$\phi_{\text{tr } \mathbf{S}}(\nu) = \frac{\sigma_1^{-M} \sigma_2^{-M} \dots \sigma_N^{-M}}{(\sigma_1^{-1} - i\nu)^M (\sigma_2^{-1} - i\nu)^M \dots (\sigma_N^{-1} - i\nu)^M}$$

where σ_i are eigenvalues of $\mathbf{R}_x$ [13].

- 4) Given a Wishart distributed matrix $\mathbf{C}$ such that $\mathcal{P}(\mathbf{C}) = CW(\mathbf{B}, r, K)$, there exists a decomposition $\mathbf{C} = \sum_{i=1}^K \mathbf{z}_i \mathbf{z}_i^H$ where the $\mathbf{z}_i$, $i = 1, \dots, K$ are r -dimensional Gaussian i.i.d. random vectors with covariance $\mathbf{B}$ [17].

Application of property 2 implies that $M(\mathbf{U}^h \hat{\mathbf{R}}_x^{-1} \mathbf{U})^{-1}$ is distributed as

$$\mathcal{P}(M(\mathbf{U}^h \hat{\mathbf{R}}_x^{-1} \mathbf{U})^{-1}) = CW((\mathbf{U}^h \mathbf{R}_x^{-1} \mathbf{U})^{-1}, L, M - N + L). \quad (30)$$

Subsequently, using property 1

$$\begin{aligned} &\mathcal{P}(M \mathbf{G}^h (\mathbf{U}^h \hat{\mathbf{R}}_x^{-1} \mathbf{U})^{-1} \mathbf{G}) \\ &= CW(\mathbf{G}^h (\mathbf{U}^h \mathbf{R}_x^{-1} \mathbf{U})^{-1} \mathbf{G}, r, M - N + L). \end{aligned} \quad (31)$$

Let $\hat{\mathbf{A}} = M \mathbf{G}^h (\mathbf{U}^h \hat{\mathbf{R}}_x^{-1} \mathbf{U})^{-1} \mathbf{G}$. The power estimate $\hat{P}_m$ is

$$\begin{aligned} \hat{P}_m &= \text{tr } \mathbf{G}^h (\mathbf{U}^h \hat{\mathbf{R}}_x^{-1} \mathbf{U})^{-1} \mathbf{G} \\ &= \frac{1}{M} \text{tr } \hat{\mathbf{A}}. \end{aligned} \quad (32)$$

Property 3 cannot be directly applied to determine the characteristic function of $\text{tr } \hat{\mathbf{A}}$ because the derivation of property 3 utilizes the $\sum_{i=1}^M \mathbf{x}_i \mathbf{x}_i^H$ decomposition of $\mathbf{S}$. $\hat{\mathbf{A}}$ is not expressed as the sum of outer products of i.i.d. Gaussian random vectors. However, property 4 guarantees that a set of i.i.d. Gaussian random vectors whose outer product sums to $\hat{\mathbf{A}}$ exists and thus permits application of property 3. Therefore, the characteristic function of $\text{tr } \hat{\mathbf{A}}$ is

$$\phi_{\text{tr } \hat{\mathbf{A}}}(\nu) = \frac{\lambda_1^{-K} \lambda_2^{-K} \dots \lambda_r^{-K}}{(\lambda_1^{-1} - i\nu)^K (\lambda_2^{-1} - i\nu)^K \dots (\lambda_r^{-1} - i\nu)^K} \quad (33)$$

where λ_i are the eigenvalues of $\mathbf{G}^h (\mathbf{U}^h \mathbf{R}_x^{-1} \mathbf{U})^{-1} \mathbf{G}$ and $K = M - N + L$.

Finally, we complete the proof using the identities

$$\begin{aligned} \mathcal{E}(\hat{P}_m) &= \frac{1}{M} \mathcal{E}(\text{tr } \hat{\mathbf{A}}) \\ &= \frac{1}{M} \left(-i \frac{d}{d\nu} \ln \phi_{\text{tr } \hat{\mathbf{A}}}(\nu) \Big|_{\nu=0} \right) \end{aligned} \quad (34)$$

and

$$\begin{aligned} \text{var}(\hat{P}_m) &= \frac{1}{M^2} \text{var}(\text{tr } \hat{\mathbf{A}}) \\ &= \frac{1}{M^2} \left(-\frac{d^2}{d\nu^2} \ln \phi_{\text{tr } \hat{\mathbf{A}}}(\nu) \Big|_{\nu=0} \right). \end{aligned} \quad (35)$$

From (33) we have

$$\begin{aligned} \ln \phi_{\text{tr } \hat{\mathbf{A}}}(\nu) &= \ln \frac{\prod_{i=1}^r (\lambda_i^{-1})^K}{\prod_{i=1}^r (\lambda_i^{-1} - i\nu)^K} \\ &= \ln \prod_{i=1}^r (\lambda_i^{-1})^K - \ln \prod_{i=1}^r (\lambda_i^{-1} - i\nu)^K \\ &= K \sum_{i=1}^r \ln \lambda_i^{-1} - K \sum_{i=1}^r \ln (\lambda_i^{-1} - i\nu). \end{aligned} \quad (36)$$

Substituting (36) into (34) we obtain

$$\begin{aligned} \mathcal{E}(\text{tr } \hat{\mathbf{A}}) &= -K \sum_{i=1}^r \frac{(-i)(-i)}{\lambda_i^{-1} - i\nu} \Big|_{\nu=0} \\ &= K \sum_{i=1}^r \lambda_i \end{aligned} \quad (37)$$

which implies

$$\begin{aligned} \mathcal{E}(\hat{P}_m) &= \frac{K}{M} \sum_{i=1}^r \lambda_i \\ &= \frac{M - N + L}{M} \sum_{i=1}^r \lambda_i. \end{aligned} \quad (38)$$

Now $\sum_{i=1}^r \lambda_i = \text{tr } G^h (U^h R_x^{-1} U)^{-1} G = P_m$, so

$$\mathcal{E}(\hat{P}_m) = \frac{M - N + L}{M} P_m. \quad (39)$$

Substituting (36) into (35) gives

$$\begin{aligned} \text{var}(\text{tr } \hat{A}) &= K \sum_{i=1}^r \frac{-(-i)(-i)}{(\lambda_i^{-1} - i\nu)^2} \Big|_{\nu=0} \\ &= K \sum_{i=1}^r \lambda_i^2. \end{aligned} \quad (40)$$

Thus,

$$\text{var}(\hat{P}_m) = \frac{M - N + L}{M^2} \sum_{i=1}^r \lambda_i^2. \quad \square \quad (41)$$

Estimation of R_x introduces a bias which depends only on M , N , and L . They are known so the bias is easily removed by multiplying $\hat{P}_m$ by $M/(M - N + L)$. Clearly, $\hat{P}_m$ is unbiased as M increases without bound. The variance of the unbiased estimator is

$$\text{var} \left(\frac{M}{M - N + L} \hat{P}_m \right) = \frac{1}{M - N + L} \sum_{i=1}^r \lambda_i^2. \quad (42)$$

The variance given in (42) decreases as M increases and/or $N - L$ decreases. As M increases without bound the variance goes to zero so the estimate is consistent.

It is convenient for comparison purposes to normalize the variance with respect to the mean

$$\text{var} \left(\frac{\hat{P}_m}{\mathcal{E}(\hat{P}_m)} \right) = \frac{1}{M - N + L} \frac{\sum_{i=1}^r \lambda_i^2}{\left(\sum_{i=1}^r \lambda_i \right)^2}. \quad (43)$$

We define the performance factor, pf, as

$$\text{pf} = \frac{\left(\sum_{i=1}^r \lambda_i \right)^2}{\sum_{i=1}^r \lambda_i^2}. \quad (44)$$

pf is a measure of the degree of variance reduction obtained by multiple window processing and can be viewed as the effective number of degrees of freedom employed by the multiple window estimator. It is straightforward to show that pf is bounded by $1 \leq \text{pf} \leq r$. The upper bound is attained when all the λ_i are equal and the lower bound when all the λ_i except one are zero. In the single window case $r = 1$ and $\text{pf} = 1$. At best multiple window

processing reduces the variance by the number of windows while at worst the variance is equivalent to the single window case.

pf can be directly evaluated for several special cases. If $R_x = I$ and $G = (1/\sqrt{N}) [I_r^h]$ (this G satisfies $GG^h = (1/N) \text{diag}(1, 1, \dots, 1, 0, \dots, 0)$) then $G^h(U^h R_x^{-1} U)^{-1} G = G^h G = (1/N) I_r$ and $\text{pf} = r$ where r is the number of ones in GG^h . A more general case results if $S(k)$ is constant over $\mathcal{G}^n$ and B_e is small in addition to $G = (1/\sqrt{N}) [I_r^h]$. Recall that λ_i are also the eigenvalues of $W^h R_x W$ which is approximately given by $G^h U^h H U G$ (see (27)). U contains eigenvectors of H so the λ_i are $1/N$ times the eigenvalues of $H(\lambda_i)$. Although in general these must be evaluated numerically, if the data is from a uniformly spaced time series and $\mathcal{G}^n$ is a symmetric frequency band about zero, then the first r eigenvalues of H are approximately equal where r is the time bandwidth product [3], [10], [11] and the variance is reduced by a factor of r .

The normalized variance is also inversely proportional to L , the number of constraints. Increasing the number of constraints reduces normalized variance and is equivalent to reducing the degrees of freedom in W that can be used to minimize exterior bias B_e . Reducing degrees of freedom in the MWMV estimator is analogous to partially adaptive beamforming [18]. It is straightforward to adapt partially adaptive beamformer design techniques [18]–[20] to design partially adaptive MWMV estimators. The disadvantage associated with reducing the number of degrees of freedom is an increase in the exterior bias B_e .

V. TRADING INTERIOR BIAS FOR EXTERIOR BIAS

It is known in the context of beamforming that the level of interference cancellation can be increased by allowing distortion of the desired signal [21], [22]. This is also true in the MWMV estimator: the exterior bias is reduced by allowing an increase in the interior bias. The soft constraint philosophy proposed in [21] is directly applicable to the MWMV estimator. However, the soft constraint replaces the linear constraints with quadratic constraints and changes the basic form of the estimator to the extent that the statistical analysis of the previous section is no longer applicable. In this section we consider optimizing G to sacrifice interior bias for a gain in exterior bias. This approach retains the basic form of the MWMV estimator.

The exterior bias is strongly dependent on the external spectrum $S(k)$, $k \notin \mathcal{G}^n$. In order to incorporate this dependence in the design of G we minimize the average exterior bias over a range of likely scenarios for the external spectrum. A similar philosophy has proved to be very effective in partially adaptive beamformer design [18], [19]. The likely scenarios may be generated based on physical constraints and other *a priori* knowledge, or could be defined based on a preliminary spectrum estimate. Let a vector θ of parameters represent the external spectrum, denoted as $S_\theta(k)$, $k \notin \mathcal{G}^n$, and assume the set of likely external spectra is described by $\theta \in \mathcal{E}$. For example, in a sensor array

processing problem θ could represent the directions and power levels of sources of interference. Ξ would then represent the likely range of power levels and directions from which the interferers arrive. Define

$$\mathbf{R}_I(\theta) = \int_{\mathbf{k} \in \mathcal{G}^n} \phi(\mathbf{k}) \phi^h(\mathbf{k}) S_\theta(\mathbf{k}) d\mathbf{k} \quad (45)$$

so that

$$\mathbf{R}_x = \int_{\mathbf{k} \in \mathcal{G}^n} \phi(\mathbf{k}) \phi^h(\mathbf{k}) S_\theta(\mathbf{k}) d\mathbf{k} + \mathbf{R}_I(\theta). \quad (46)$$

It can be shown using the GSC decomposition that the exterior bias defined in (9) is expressed as

$$\begin{aligned} B_e(\theta) = & \text{tr}(\mathbf{G}^h \mathbf{U}^h \mathbf{R}_I(\theta) \mathbf{U} \mathbf{G}) \\ & - \mathbf{G}^h \mathbf{U}^h \mathbf{R}_I(\theta) \bar{\mathbf{U}} (\bar{\mathbf{U}}^h \mathbf{R}_I(\theta) \bar{\mathbf{U}})^{-1} \bar{\mathbf{U}}^h \mathbf{R}_I(\theta) \mathbf{U} \mathbf{G} \end{aligned} \quad (47)$$

where $\bar{\mathbf{U}}$ represents a basis for the $N - L$ dimensional space orthogonal to the columns of $\mathbf{U}$. We define the "average" exterior bias $B_{e_{av}}$ over Ξ as

$$B_{e_{av}} = \int_{\theta \in \Xi} B_e(\theta) d\theta. \quad (48)$$

The goal is to choose $\mathbf{G}$ to minimize $B_{e_{av}}$ subject to a constraint on the maximum interior bias. Constraining the magnitude squared of $B_i(\mathbf{k})$ (see (10)) leads to an expression which is quartic in $\mathbf{G}$ and presents considerable analytic obstacles. Instead of using this expression, we consider constraining a somewhat simpler function of $\mathbf{G}$ that is closely related to the interior bias

$$f(\mathbf{k}, \mathbf{G}) = \|\mathbf{G}^h \mathbf{U}^h \phi(\mathbf{k}) - \mathbf{G}_o^h \mathbf{U}^h \phi(\mathbf{k})\|^2, \quad \mathbf{k} \in \mathcal{G}^n. \quad (49)$$

Here $\mathbf{G}_o$ is the $\mathbf{G}$ that minimizes the interior bias. $f(\mathbf{k}, \mathbf{G})$ measures the squared error between the response associated with $\mathbf{G}$ and the response that minimizes the interior bias. This is actually more restrictive than constraining the interior bias since $f(\mathbf{k}, \mathbf{G})$ measures both magnitude and phase errors while $B_i(\mathbf{k})$ only measures magnitude errors. In choosing $\mathbf{G}$ to minimize $B_{e_{av}}$ we constrain the average value of $f(\mathbf{k}, \mathbf{G})$ on $\mathbf{k} \in \mathcal{G}^n$

$$\int_{\mathbf{k} \in \mathcal{G}^n} f(\mathbf{k}, \mathbf{G}) d\mathbf{k} \leq e. \quad (50)$$

The optimization problem for $\mathbf{G}$ is written as

$$\min_{\mathbf{G}} \text{tr} \mathbf{G}^h \mathbf{Q} \mathbf{G} \quad \text{subject to } \text{tr}(\mathbf{G}_o - \mathbf{G})^h \mathbf{\Pi} (\mathbf{G}_o - \mathbf{G}) \leq e \quad (51)$$

where

$$\begin{aligned} \mathbf{Q} = & \int_{\theta \in \Xi} (\mathbf{U}^h \mathbf{R}_I(\theta) \mathbf{U} \\ & - \mathbf{U}^h \mathbf{R}_I(\theta) \bar{\mathbf{U}} (\bar{\mathbf{U}}^h \mathbf{R}_I(\theta) \bar{\mathbf{U}})^{-1} \bar{\mathbf{U}}^h \mathbf{R}_I(\theta) \mathbf{U}) d\theta \end{aligned} \quad (52)$$

and

$$\mathbf{\Pi} = \mathbf{U}^h \left[\int_{\mathbf{k} \in \mathcal{G}^n} \phi(\mathbf{k}) \phi^h(\mathbf{k}) d\mathbf{k} \right] \mathbf{U}. \quad (53)$$

This constrained minimization problem is solved using the method of Lagrange multipliers. Define the Lagrangian

$$L(\lambda, \mathbf{G}) = \text{tr} \mathbf{G}^h \mathbf{Q} \mathbf{G} + \lambda \text{tr}(\mathbf{G}_o - \mathbf{G})^h \mathbf{\Pi} (\mathbf{G}_o - \mathbf{G}). \quad (54)$$

Forcing the derivative of $L(\lambda, \mathbf{G})$ with respect to $\mathbf{G}$ equal to zero gives

$$2\mathbf{Q}\mathbf{G} - 2\lambda\mathbf{\Pi}(\mathbf{G}_o - \mathbf{G}) = \mathbf{0}. \quad (55)$$

The solution to (55) is given as a function of λ as

$$\mathbf{G}(\lambda) = \lambda(\mathbf{Q} + \lambda\mathbf{\Pi})^{-1}\mathbf{\Pi}\mathbf{G}_o. \quad (56)$$

The largest possible λ is chosen such that $\mathbf{G}(\lambda)$ satisfies the constraint

$$\text{tr}(\mathbf{G}_o - \mathbf{G}(\lambda))^h \mathbf{\Pi} (\mathbf{G}_o - \mathbf{G}(\lambda)) \leq e. \quad (57)$$

It is easy to show that λ and e are inversely related. At one extreme, $\lambda \rightarrow 0$, we have

$$\lim_{\lambda \rightarrow 0} e = \text{tr} \mathbf{G}_o^h \mathbf{\Pi} \mathbf{G}_o \quad (58)$$

while at the other extreme, $\lambda \rightarrow \infty$, we have

$$\lim_{\lambda \rightarrow \infty} e = 0 \quad (59)$$

since

$$\lim_{\lambda \rightarrow 0} \mathbf{G}(\lambda) = \lim_{\lambda \rightarrow 0} \lambda(\mathbf{Q} + \lambda\mathbf{\Pi})^{-1}\mathbf{\Pi}\mathbf{G}_o = \mathbf{0} \cdot \mathbf{Q}^{-1}\mathbf{\Pi}\mathbf{G}_o = \mathbf{0} \quad (60)$$

and

$$\lim_{\lambda \rightarrow \infty} \mathbf{G}(\lambda) = \lim_{\lambda \rightarrow \infty} \lambda(\mathbf{Q} + \lambda\mathbf{\Pi})^{-1}\mathbf{\Pi}\mathbf{G}_o = \mathbf{G}_o. \quad (61)$$

In general an iterative algorithm is needed to solve (57) for λ . However, approximate solutions for λ can be obtained for either λ very small or very large. It is shown in the Appendix that for sufficiently large λ (small e)

$$\lambda = \sqrt{\frac{\text{tr} \mathbf{G}_o^h \mathbf{Q} \mathbf{\Pi}^{-1} \mathbf{Q} \mathbf{G}_o}{e}}. \quad (62)$$

Similarly, if λ is sufficiently small (e large), then λ is the largest positive root of the quadratic equation

$$\gamma - \beta\lambda + \alpha\lambda^2 = e. \quad (63)$$

The constants γ , β , and α are given in the appendix as functions of $\mathbf{Q}$, $\mathbf{\Pi}$, and $\mathbf{G}_o$.

If e is large, then the interior bias will be large and the estimate of power for $\mathbf{k} \in \mathcal{G}^n$ will be severely biased. If the shape of $S(\mathbf{k})$, $\mathbf{k} \in \mathcal{G}^n$ is assumed known, then this bias can be removed by normalizing the estimated power. Let $\sigma(\mathbf{k})$ represent the assumed shape of $S(\mathbf{k})$, $\mathbf{k} \in \mathcal{G}^n$ normalized so that $\int_{\mathbf{k} \in \mathcal{G}^n} \sigma(\mathbf{k}) d\mathbf{k} = 1$. Given $\mathbf{G}$, the bias is equalized by dividing the estimated power by the constant

$$c = \text{tr} \int_{\mathbf{k} \in \mathcal{G}^n} \mathbf{G}^h \mathbf{U}^h \phi(\mathbf{k}) \phi^h(\mathbf{k}) \mathbf{U} \mathbf{G} \sigma(\mathbf{k}) d\mathbf{k}. \quad (64)$$

Fig. 1. Sensor array/delay line structure and block diagram of the MWMV spatial spectrum estimator. The array has 8 sensors in a linear equally spaced geometry with 6 temporal samples per sensor channel.

Errors will result if the actual and assumed spectral shapes differ.

Normalization also changes the exterior bias, but does not change the ratio of exterior to interior bias. The goal of optimizing G is to obtain significantly greater reductions in exterior bias than the increase in interior bias. If this is accomplished, then the normalized estimate will have less exterior bias than an estimate based on G_o .

VI. SIMULATION

The MWMV method is employed in this section to estimate the spatial spectrum of a temporally broad-band propagating field which is sampled by an array of sensors. Eight sensors are arranged in a linear equally spaced configuration with intersensor spacing of one half the wavelength corresponding to the highest frequency present. The sensor outputs are sampled in time with six samples per sensor as depicted in Fig. 1. Thus, the dimension of the data vector x is 48 and t represents spatial position and time. k represents spatial direction θ measured in degrees relative to the perpendicular to the array and temporal frequency f . $\mathcal{K}^n$ represents directions from $-60^\circ \leq \theta \leq 60^\circ$ and frequencies normalized on the interval $-0.5 \leq f \leq 0.5$. All the simulations assume the environment contains two independent sources with equal power arriving from $+5$ and -5 degrees in uncorrelated noise. The ratio of source to white noise power is 15 dB and the sources of energy have constant spectral density on the frequency interval $0.2 \leq |f| \leq 0.4$.

The band $\mathcal{G}^n$ over which power is estimated is

$$\mathcal{G}^n = \{(\theta, f): 0.2 \leq |f| \leq 0.4, \theta = \theta_o\} \quad (65)$$

where a different set of MWMV weights is designed for each direction θ_o of interest. The number of constraints is chosen as $L = 6$ and $r = 6$ windows are used in the MWMV estimates. A single window minimum variance estimate is also presented for comparison purposes. The multiple and single window estimators use the same constraint matrix U , but different response matrices G . The single window $G = g$ was chosen to approximate unit gain and linear phase on $\mathcal{G}^n$ using the least squares procedure described in [9] while the multiple window estimator employed $G = I$ as suggested in Section III. This

results in single and multiple window estimators with approximately the same level of interior bias.

Fig. 2 illustrates representative responses of the single and multiple window estimators as a function of frequency f in the direction $\theta_o = 40^\circ$. The deviation of the response from unity is about the same for both estimators. The power computed using the true covariance matrix is depicted in Fig. 3 for the single and multiple window minimum variance methods. The power obtained using Thomson's approach is also plotted in Fig. 3. The deviation of the power estimates from the true spectrum is a result of exterior bias since the interior bias is close to zero and the true covariance matrix is used. The exterior biases of the MWMV and single window methods are similar. Thomson's power estimate has much larger values of exterior bias due to leakage of energy through sidelobes and two sources cannot be distinguished.

Fig. 4 plots the performance factor pf to illustrate the degree of variance reduction gain associated with the MWMV method. At angles far away from the point sources the MWMV estimator obtains near the maximum variance reduction factor (r). In these regions the MWMV filter sidelobes effectively null out the point sources. Thus, B_e is small. $S(k)$, $k \in \mathcal{G}^n$ is also constant, so, as discussed in Section III, the MWMV satisfies (13). Near the point sources B_e increases and the degree of variance reduction decreases.

Ten independent simulations were conducted using independent Gaussian distributed random vectors with covariance corresponding to the true covariance matrix. Fifty-two data vectors are used in each simulation to estimate R_x and the estimates are normalized by $M/(M - N + L)$ to eliminate bias. Figs. 5 and 6 depict single and multiple window minimum variance estimates, respectively. Variance reduction is evident over all angles, but is particularly large at angles away from the point sources as predicted by Fig. 4.

Fig. 3 indicates significant exterior bias when θ_o is in the vicinity of the source directions. The exterior bias can be reduced using the methods for optimizing G presented in Section V. In the next example the vector θ of parameters represents the directions of two point sources in uncorrelated noise. The range of directions $\mathcal{E}$ is chosen to

Fig. 2. Temporal frequency responses of the windows corresponding to single and multiple window minimum variance estimators in the direction of interest. In this example the direction of interest is $\theta_o = -40^\circ$.

Fig. 3. Spatial spectrum estimates based on the true covariance matrix. Plots show the MWMV estimate, the single window minimum variance estimate, and Thomson's estimate for two broad-band sources located at $\pm 5^\circ$ in -15 dB white noise. Thomson's adaptive weighting procedure is not used.

be an interval of $\pm 11^\circ$ centered on the direction of interest θ_o . $\mathcal{E}$ is chosen to be centered on θ_o because the exterior bias is greatest when a point source is near θ_o . Four hundred eighty-four possible arrangements of these two sources in $\mathcal{E}$ are used to design $\mathbf{G}$ for each θ_o . The power computed using the true covariance matrix for the optimized $\mathbf{G}$ is normalized as described in Section V and compared to that for $\mathbf{G} = \mathbf{I}$ in Fig. 7. The actual spectral shape is constant so normalization eliminates all interior bias due to optimization of $\mathbf{G}$. The approach is clearly effective in reducing exterior bias.

The performance factor is plotted in Fig. 8 for the estimator based on the optimized $\mathbf{G}$. The degree of variance reduction is less than that obtained with $\mathbf{G} = \mathbf{I}$ even though B_e is smaller. Recall that in Section III we show that maximum variance reduction requires B_e small, $S(\mathbf{k})$, $\mathbf{k} \in \mathcal{G}^n$ constant, and $\mathbf{G}$ diagonal. The optimization process yields $\mathbf{G}$ that are not diagonal. Ten independent simulations based on random data are depicted in Fig. 9. The increase in variance and decrease in B_e are evident.

Fig. 4. Reduction in variance of the MWMV estimator relative to single window estimation as predicted by the performance factor defined in (44).

Fig. 5. Ten independent simulations of the single window LCMV estimator. Fifty-two independent data vectors are used in the sample covariance matrix estimate. Each spectrum estimate is normalized by $M/(M - N + L)$.

Fig. 6. Ten independent simulations of the MWMV estimator. Fifty-two independent data vectors are used in the sample covariance matrix estimate. Each estimate is normalized by $M/(M - N + L)$.

VII. SUMMARY

A multiple window spectrum estimator is proposed for data obtained from a homogeneous random field. The es-

Fig. 7. Plots of the MWMV estimate using an optimized G and $G = (1/\sqrt{48})I$ based on the true covariance matrix. Two broad-band sources are located at $\pm 5^\circ$ in -15 dB white noise.

Fig. 8. Reduction in variance of the MWMV estimator using an optimized G relative to single window processing as predicted by the performance factor defined in (44).

Fig. 9. Ten independent simulations of the MWMV estimator with an optimized G . Fifty-two independent data vectors are used in the sample covariance matrix estimate. Each spectrum estimate is normalized by $M/(M - N + L)$.

timator consists of a bank of filters whose output powers are averaged to estimate the power in a narrow band of the process. The filter sidelobes are chosen to minimize

leakage of energy from outside the band of interest. The bias and variance properties of the MWMV estimator are compared to the multiple window estimators of Thomson [3] and Bronez [4].

The MWMV method requires an estimate of the data covariance matrix. The mean and variance of the MWMV estimator are derived assuming the sample covariance matrix estimate is employed and the data vectors are independent and identically Gaussian distributed. A performance factor which reflects the degree of variance reduction obtained using multiple window is defined in terms of the estimator variance. At best the variance is reduced by the number of windows and at worst the variance is not changed by using multiple windows.

APPENDIX A

Let $Q = \Gamma \Lambda \Gamma^h$ and use the matrix inversion formula to express $(Q + \lambda \Pi)^{-1}$ as

$$(Q + \lambda \Pi)^{-1} = (\lambda \Pi)^{-1} - (\lambda \Pi)^{-1} \Gamma (\Gamma^h (\lambda \Pi)^{-1} \Gamma + \Lambda^{-1})^{-1} \Gamma^h (\lambda \Pi)^{-1}. \quad (A.1)$$

Substituting (A.1) into (56) we obtain

$$G(\lambda) = \lambda (Q + \lambda \Pi)^{-1} \Pi G_o \\ = G_o - \Pi^{-1} \Gamma (\Gamma^h \Pi^{-1} \Gamma + \lambda \Lambda^{-1})^{-1} \Gamma^h G_o. \quad (A.2)$$

If λ is sufficiently large such that

$$[\lambda \Lambda^{-1}]_{ii} \gg [\Gamma^h \Pi^{-1} \Gamma]_{i,j}. \quad (A.3)$$

Then

$$G(\lambda) \approx G_o - \Pi^{-1} \Gamma (\lambda \Lambda^{-1})^{-1} \Gamma^h G_o \\ \approx G_o - \frac{1}{\lambda} \Pi^{-1} \Gamma \Lambda \Gamma^h G_o \\ \approx G_o - \frac{1}{\lambda} \Pi^{-1} Q G_o. \quad (A.4)$$

Substituting (A.4) into (57) implies that

$$\lambda = \sqrt{\frac{\text{tr } G_o^h Q \Pi^{-1} Q G_o}{e}}. \quad (A.5)$$

If λ is sufficiently small, then a Taylor series approximation of the matrix inverse in (A.2)

$$F(\lambda) = (\Gamma^h \Pi^{-1} \Gamma + \lambda \Lambda^{-1})^{-1} \quad (A.6)$$

provides the needed simplification. We approximate $F(\lambda)$ as

$$F(\lambda) \approx F(0) + \left. \frac{dF(\lambda)}{d\lambda} \right|_{\lambda=0} \lambda. \quad (A.7)$$

The derivative is evaluated using the property

$$\frac{d(A^{-1})}{d\lambda} = -A^{-1} \frac{dA}{d\lambda} A^{-1} \quad (A.8)$$

which results in

$$F(\lambda) \approx A^{-1}(0) + \left. \frac{d(A^{-1}(\lambda))}{d\lambda} \right|_{\lambda=0} \cdot \lambda \\ \approx (\Gamma^h \Pi^{-1} \Gamma)^{-1} - A^{-1} \Lambda^{-1} A^{-1} \Big|_{\lambda=0} \lambda \\ \approx (\Gamma^h \Pi^{-1} \Gamma)^{-1} - (\Gamma^h \Pi^{-1} \Gamma)^{-1} \Lambda^{-1} (\Gamma^h \Pi^{-1} \Gamma)^{-1} \lambda. \quad (A.9)$$

This implies that $G(\lambda)$ is expressed as

$$G(\lambda) \approx G_o - \Pi^{-1}\Gamma(C - B\lambda)\Gamma^h G_o \quad (\text{A.10})$$

where $B = (\Gamma^h \Pi^{-1} \Gamma)^{-1} \Lambda^{-1} (\Gamma^h \Pi^{-1} \Gamma)^{-1}$ and $C = (\Gamma^h \Pi^{-1} \Gamma)^{-1}$. The constraint is then a quadratic function of λ since

$$\begin{aligned} \text{tr}(G_o - G(\lambda))^h \Pi (G_o - G(\lambda)) \\ \approx \text{tr} G_o^h \Gamma (C - B\lambda) \Gamma^h \Pi^{-1} \Pi \Pi^{-1} \Gamma (C - B\lambda) \Gamma^h G_o \\ = \text{tr} (C \Gamma^h G_o G_o^h \Gamma C \Gamma^h \Pi^{-1} \Gamma - (C \Gamma^h G_o G_o^h \Gamma B^h \\ + B \Gamma^h G_o G_o^h \Gamma C^h) \lambda + B \Gamma^h G_o G_o^h \Gamma B^h \Pi^{-1} \Gamma \lambda^2). \end{aligned} \quad (\text{A.11})$$

Thus, λ is given by the largest positive root of the quadratic equation

$$\gamma - \beta\lambda + \alpha\lambda^2 = e \quad (\text{A.12})$$

where

$$\begin{aligned} \alpha &= \text{tr} B \Gamma^h G_o G_o^h \Gamma B^h \Pi^{-1} \Gamma \\ \beta &= \text{tr} (C \Gamma^h G_o G_o^h \Gamma B^h + B \Gamma^h G_o G_o^h \Gamma C^h) \end{aligned}$$

and

$$\gamma = \text{tr} C \Gamma^h G_o G_o^h \Gamma C^h \Pi^{-1} \Gamma. \quad (\text{A.13})$$

REFERENCES

- [1] A. Shuster, "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena," *Terrestr. Mag.*, vol. 3, pp. 13-41, Mar. 1898.
- [2] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408-1418, Aug. 1969.
- [3] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055-1096, Sept. 1982.
- [4] T. P. Bronez, "Spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 1862-1873, Dec. 1988.
- [5] 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, June 1967.
- [6] P. J. Daniell, "Discussion on the theoretical specification and sampling properties of autocorrelated time series," *J. Roy. Stat. Soc. (suppl.)*, vol. 8, pp. 88-90, 1946.
- [7] A. M. Yaglom, *Correlation Theory of Stationary and Related Random Functions I*. New York: Springer, 1987, pp. 329.
- [8] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926-935, Aug. 1972.
- [9] K. M. Buckley, "Spatial/spectral filtering with linearly constrained minimum variance beamformers," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 249-266, Mar. 1987.
- [10] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case," *Bell Syst. Tech. J.*, vol. 57, pp. 1371-1429, May-June 1978.
- [11] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty—III," *Bell Syst. Tech. J.*, vol. 40, pp. 1295-1336, July 1962.
- [12] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive arrays," *IEEE Trans. Antennas Propagat.*, vol. AP-30, pp. 27-34, Jan. 1982.
- [13] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Stat.*, vol. 34, pp. 152-177, Mar. 1963.
- [14] J. Capon and N. R. Goodman, "Probability distributions for estimators of the frequency-wavenumber spectrum," *Proc. IEEE*, vol. 58, pp. 1785-1786, Oct. 1970.
- [15] E. J. Kelley and K. M. Forsythe, "Adaptive detection and parameter estimation for multidimensional signal models," M.I.T. Lincoln Lab., Tech. Rep. 848, Appendix A, Apr. 19, 1989.
- [16] M. L. Eaton, *Multivariate Statistics—A Vector Space Approach*. New York: Wiley, 1983, pp. 304-312.
- [17] S. M. Srivastava and G. C. Khatri, *An Introduction to Multivariate Statistics*. New York: Elsevier North-Holland, 1979, p. 77.
- [18] B. D. Van Veen and R. A. Roberts, "Partially adaptive beamformer design via output power minimization," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1524-1532, Nov. 1987.
- [19] B. D. Van Veen, "An analysis of several partially adaptive beamformer designs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 192-203, Feb. 1989.
- [20] K. A. Byerly and R. A. Roberts, "Output power based partially adaptive array design," in *Proc. Twenty-Third Asilomar Conf. Signals, Syst., Comput.*, 1989, pp. 576-580.
- [21] B. D. Van Veen, "Minimum variance beamforming with soft response constraints," *IEEE Trans. Signal Processing*, vol. 39, no. 9, pp. 1964-1972, Sept. 1991.
- [22] D. Chazan, Y. Medan, and Elko, "Noise cancellation for hearing aids," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1697-1705, Nov. 1988.

Tsung-Ching Liu received the B.S. degree in electronic engineering from the National Taiwan Institute of Technology in 1980, the M.S. degree from the Florida Institute of Technology in 1982, and the Ph.D. degree from the University of Wisconsin-Madison in 1990, both in electrical engineering.

During military service from 1975 to 1977, he served as an Electronics Officer on a Naval destroyer in Taiwan. From 1977 to 1980 he also worked as a System Engineer for the Taipei Telephone Company and was in charge of the TI PCM transmission system in the Taipei section. From 1982 to 1984, he was a Lecturer at the Ming-Chi Institute of Technology, Taipei Hsien, Taiwan. He is currently with the R&D Group of Toshiba Corporation in Tokyo, Japan. His research interests include spectrum estimation, beamforming, statistical signal processing, digital signal processing, and digital communications.

Dr. Liu is a member of the IEEE Signal Processing Society.

Barry D. Van Veen (S'81-M'83) was born in Green Bay, WI, in 1961. He received the B.S. degree from Michigan Technological University in 1983 and the Ph.D. degree from the University of Colorado in 1986, both in electrical engineering.

In the spring of 1987 he was with the Department of Electrical and Computer Engineering at the University of Colorado, Boulder. Since August of 1987 he has been with the Department of Electrical and Computer Engineering at the University of Wisconsin-Madison and is currently an Associate Professor. His research interests include signal processing for sensor arrays, spectrum estimation, adaptive algorithms, and biomedical applications of signal processing.

Dr. Van Veen was a recipient of the 1989 Presidential Young Investigator Award.