CIRCUITS SYSTEMS SIGNAL PROCESS
VOL. 6, NO. 4, 1987

Estimation of Structured Covariances with Application to Array Beamforming*

D. A. Gray,[1] *B. D. O. Anderson,*[2] *and P. K. Sim*[3]

Abstract. The estimation of covariance matrices which are structured, for example, of Toeplitz type, from measurement data is considered. The problem is considered in the context of array beamforming, and various methods of estimation are derived and compared, such comparison including consideration of the behavior of the estimate in beamforming applications.

1. Introduction

Array beamforming may be carried out in a variety of domains, e.g., in either the time or frequency domains, and using either the receiver outputs directly or other inputs derived from the receiver outputs. Here we consider frequency-domain beamforming whereby the receiver time series outputs are first Fourier transformed, or narrow-band filtered, to the frequency of interest. The narrow-band receiver outputs are then used to estimate the cross-spectral matrix, from which the power incident upon the array from any given direction can also be estimated. The use in adaptive beamforming [1], [2] of the estimated cross-spectral matrix has been shown [3] to result in considerable improvement in the ability to reject interfering noise. Also, in the theory of maximum entropy beamforming [4] the cross-spectral, or spatial covariance, matrix plays an important role in determining the parameters of the wave-number power spectrum. Furthermore, in a number of bearing estimation algorithms that have recently been proposed [5], the eigenvalues and eigenvectors of the cross-spectral matrix play a central role.

* Received June 8, 1986. The work of the first author was carried out as a Defence Fellow on leave from the Australian Department of Defence, whose support is gratefully acknowledged. The other two authors were also supported by the Department of Defence.

[1] Weapons System Research Laboratory, G.P.O. Box 2151, Adelaide, SA 5001, Australia.

[2] Department of Systems Engineering, Research School of Physical Sciences, Australian National University, G.P.O. Box 4, Canberra, ACT 2601, Australia.

[3] Department of Electronic and Communication Science, La Trobe University, Bundoora, Victoria 3083, Australia.

For many types of arrays and noise fields, the exact cross-spectral matrix has a particular structure, e.g., in a noise field consisting of a superposition of plane waves for a linear array of equispaced receivers, it is Toeplitz [4]. This is due to the fact that the spatial covariance function is a function only of the spatial separation between two receivers. In such cases we refer to the cross-spectral matrix as a structured covariance matrix. The use of structured covariances in beamforming can result in improved performance since new beamforming algorithms can be developed. Furthermore, the use of structured covariances often allows efficient implementation of some of the standard algorithms, e.g., in maximum entropy beamforming the use of a Toeplitz matrix allows efficient algorithms [6], [7] for the solution of the linear equations determining the spectral parameters. Thus, when estimating the cross-spectral matrix of receiver outputs, it can be an advantage to constrain the estimate to have the same structure as the exact cross-spectral matrix, i.e., to estimate a structured covariance matrix. The problem of estimating a structured covariance matrix is the main focus of this paper.

As discussed above, when the noise field is a superposition of an arbitrary number of independent plane wave arrivals, the exact cross-spectral matrix for a linear array of equispaced receivers is Hermitian and Toeplitz [8]; i.e., the elements along any diagonal are equal. These ideas are reviewed in some detail in Section 2, so that we go on in Section 3 to consider the estimation of spatial covariance matrices with the Toeplitz structure. In Section 3 we first discuss a maximum likelihood method for estimating a Toeplitz structured covariance. Such an estimator has already been discussed in [9] where it was shown to lead to improved maximum entropy estimates of the power spectrum. The relationship of maximum likelihood and minimum cross-entropy estimation is considered. Some new estimators for a Toeplitz structured covariance are also proposed. These new estimators fall into two main groups. The first are based upon various measures—Kullback, divergence, and Bhattacharyya—of the difference between two hypotheses. The second group of new estimators is derived using various measures of the distance between two matrices. The resulting equations for the new estimators are all similar. Some iterative methods for the solution of the equations defining the estimators are discussed in Section 4. The probability-based estimators are closely related to optimum beamformers and, in Section 5, these relationships are discussed and some examples of the use of the probability-based estimates in optimum beamforming are presented.

2. Review of beamforming and occurrence of cross-spectral matrix

Consider a plane wave incident upon an array of K receivers as illustrated in Figure 1. For frequency-domain beamforming, the narrow-band receiver

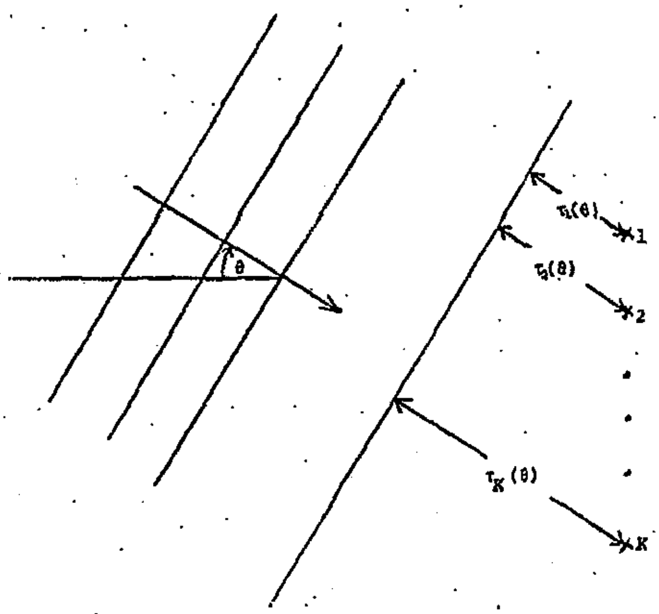

Figure 1. Plane wave incident on an array.

outputs are multiplied by phase factors proportional to the relative time delays, $\tau_j(\theta)$, as shown in Figure 1. Thus, if $x_j(f)$ is the output of the jth receiver, then the output of a *conventional beamformer steered in the direction* θ is given by

$$\mathbf{v}^H(\theta, f)\mathbf{x}(f),$$

where[4]

$$\mathbf{x}^T(f) = [x_1(f), x_2(f), \ldots, x_K(f)]$$

and for the example of a line array of equispaced receivers

$$v_j(\theta, f) = \exp(ik_z(j-1)d),$$

where $k_z = 2\pi(\sin\theta/\lambda)$, d is the separation of adjacent receivers, and λ is the wavelength. The mean output power of a conventional beamformer is given by

$$\langle|\mathbf{v}^H(\theta, f)\mathbf{x}(f)|^2\rangle,$$

[4] H denotes the complex transpose of a vector or matrix.

where $\langle\ \rangle$ denotes the ensemble average. The above expression can be shown to reduce to

$$P_c = \mathbf{v}^H(\theta, f) R(f) \mathbf{v}(\theta, f), \tag{2.1}$$

where

$$R(f) = \langle \mathbf{x}(f) \mathbf{x}^H(f) \rangle \tag{2.2}$$

and is the cross-spectral matrix of receiver outputs. Thus the mean output power of a conventional beamformer is an Hermitian form of the cross-spectral matrix.

In practice, this matrix is often estimated from a number of realizations of the receiver outputs by replacing the ensemble average with a simple linear average, e.g., given M samples of the receiver outputs

$$\mathbf{x}^{(1)}(f), \mathbf{x}^{(2)}(f), \ldots, \mathbf{x}^{(M)}(f),$$

the cross-spectral matrix of receiver outputs is estimated as

$$\begin{aligned} S &= \hat{R}(f) \\ &= \frac{1}{M} \sum_{i=1}^{M} \mathbf{x}^{(i)}(f) \mathbf{x}^{(i)H}(f). \end{aligned} \tag{2.3}$$

In general, this estimate will not be structured.

A generalization of (2.1) is to replace $\mathbf{v}$ by a different complex vector, $\mathbf{w}$, resulting in a mean output power of the form

$$\mathbf{w}^H R(f) \mathbf{w}.$$

In *optimum beamforming* (also known as maximum likelihood beamforming) $\mathbf{w}$ is chosen such that the output power is minimized whilst the response to a plane wave from a desired direction, θ, say, is kept fixed. In this case the mean output power is given by

$$P_0 = (\mathbf{v}^H(\theta, f) R^{-1}(f) \mathbf{v}(\theta, f))^{-1}. \tag{2.4}$$

Again, in practice, the above expression is estimated by replacing $R(f)$ by an estimate such as given by (2.3).

The *maximum entropy method* (MEM) of beamforming the outputs of a linear array of equispaced receivers maximizes the expression

$$\int_{-k_0}^{k_0} \ln S(k_z)\, dk_z$$

subject to the constraint

$$\int_{-k_0}^{k_0} S(k_z) \exp(ik_z pd) = r_p \qquad \text{for} \quad p = 0, \pm 1, \ldots, \pm(K-1) \tag{2.5}$$

and where $k_0 = \pi/d$ and r_p is a covariance between receivers. More precisely, when the receivers are equispaced, r_p is the covariance between any two receivers separated by a distance of dp, and is given by

$$r_p = \langle x_l^*(f) x_{l+p}(f) \rangle.$$

From the above expression the elements of the pth diagonal of the cross-spectral matrix of narrow-band receiver outputs are all equal and given by r_p. The solution of the above optimization problem is

$$S(k_z)=\frac{P_{K-1}}{2d}\left(\left|\sum_{p=0}^{K-1}-\alpha_p\exp(ik_zpd)\right|^2\right)^{-1}, \tag{2.6}$$

where P_{K-1} and the α_p satisfy the equation

$$\begin{bmatrix} r_0 & r_1 & \cdots & r_{K-1} \\ r_{-1} & r_0 & r_1 & \vdots \\ \vdots & & & \vdots \\ r_{-(K-1)} & & & r_0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \alpha_1 \\ \vdots \\ \alpha_{K-1} \end{bmatrix} = \begin{bmatrix} P_{K-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2.7}$$

The matrix on the left of the above equation is a Toeplitz structured covariance matrix and efficient methods [6], [7] based upon this structure can be used to solve the above equation.

A further example of the use of the cross-spectral matrix of receiver outputs is in the theory of estimating the *arrival direction* of a plane wave incident upon the array. Recent work by a number of authors [5], [11], [12] has indicated that the maximum of the beam pattern of the eigenvector corresponding to the largest eigenvalue is a useful estimate of the arrival direction of the strongest plane wave signal. In a different approach, other authors [13] have indicated how zeros of the beam pattern of the eigenvector corresponding to the minimum eigenvalue can also be used to estimate the arrival directions of plane wave signals. The above methods can be sensitive to data errors such as phase errors and more recent work has suggested that stable eigenvector methods are achievable by weighting the eigenvalues [14].

In the above applications, i.e., conventional, optimum, and maximum entropy beamforming, and also in the theory of bearing estimation, the cross-spectral matrix of receiver outputs plays a central role. As discussed earlier this matrix is not known *a priori* and must be estimated from samples of the receiver outputs. A natural estimator of the cross-spectral matrix is given by the matrix S, defined by (2.3) above, but for linear arrays of equispaced receivers the Toeplitz structure of the true cross-spectral matrix may not be reflected in the estimated matrix S.

Since the true cross-spectral matrix is a structured covariance it seems reasonable to estimate not S, but a matrix with the same structure as the exact cross-spectral matrix. For example, estimates of the spatial covariance, r_p, can readily be obtained from S by averaging along the pth diagonal i.e.,

$$\hat{r}_p=\frac{1}{K}\sum_{i=1}^{K-|p|}S_{i,i+p}. \tag{2.8}$$

This estimator has the advantage of producing a positive definite Toeplitz structured matrix but, as discussed in [15], can introduce distortion when used to estimate the maximum entropy power spectrum. In the following section we consider some alternative methods for estimating Toeplitz structured covariances.

3. Estimation of finite-dimensional Toeplitz matrices

In this section we consider the problem of estimating a finite K-dimensional Toeplitz matrix, R, from M independent realizations of the receiver outputs. Estimation of the entries of R by the maximum likelihood method is first considered on the assumption that $r_0, r_{\pm 1}, \ldots, r_{\pm(K-1)}$ are independent. This approach has already received some attention [9], [16] and has been shown to lead to improved maximum entropy estimates of the frequency power spectrum.

By defining an estimate of the entropy of a random process, as a function of one, or a number, of realizations of that random process, we can estimate the covariance matrix R by requiring that this entropy estimator is minimized. It is shown in Section 3.2 that this procedure is equivalent to maximum likelihood estimation.

In Sections 3.3 and 3.4 we propose a number of new methods for estimating R and discuss conditions for the existence of unique solutions. These new estimators fall into two main classes. The first class is derived from measures of the difference between two hypotheses when the underlying processes are Gaussian. Such measures, discussed in [17] for signal selection problems, are used in Section 3.3 to derive estimates of the Toeplitz covariance matrix. The second class, discussed in Section 3.4, is based upon minimizing a measure of the distance between two matrices, one of which is R and the other is S, given by (2.3) and a function of the M independent realizations.

3.1. *Maximum likelihood estimation*

We first discuss existing work [9] on this problem, deriving the equations to be solved and discussing the question of the existence of a solution.

Let $\mathbf{x}^{(m)}$, $m = 1, 2, \ldots, M$, denote M independent sample realizations of the vector of complex receiver outputs. Assuming these are Gaussian the joint probability density function is given [18] by

$$p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M) = (\pi)^{-MK} |R|^{-M} \prod_{m=1}^{M} \exp(-\mathbf{x}_m^{\mathrm{H}} R^{-1} \mathbf{x}_m),$$

where R is the assumed Toeplitz covariance matrix and $|A|$ denotes the determinant of a matrix A. Ignoring additive and multiplicative constants, the function, L, to be maximized is given by

$$L = -\ln|R| - \mathrm{Tr}(R^{-1}S), \tag{3.1}$$

where S is given by

$$S=\frac{1}{M}\sum_{m=1}^{M}\mathbf{x}^{(m)}\mathbf{x}^{(m)\mathrm{H}},$$

and is the maximum likelihood estimator of the unstructured covariance matrix [18]. Of course, the R maximizing L is the structured covariance matrix estimate.

Using the identities

$$\frac{\partial \ln|R|}{\partial\alpha}=\mathrm{Tr}\left(R^{-1}\frac{\partial R}{\partial\alpha}\right)$$

and

$$\frac{\partial R^{-1}}{\partial\alpha}=-R^{-1}\frac{\partial R}{\partial\alpha}R^{-1}$$

allows (3.1) to be maximized with respect to the real and imaginary parts of r_p for $p=0, 1, \ldots, K-1$. The result is the following set of equations:

$$\mathrm{Tr}(V^pR^{-1})=\mathrm{Tr}(V^pR^{-1}SR^{-1}) \tag{3.2}$$

for R. The matrix V is defined as

$$V=\begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0\\ 0 & 0 & 1 & 0 & \cdots & \vdots\\ \vdots & & & \ddots & \ddots & \vdots\\ \vdots & & & & & 0\\ \vdots & & & & 0 & 1\\ 0 & \cdots & \cdots & \cdots & 0 & 0 \end{bmatrix}$$

with $V^{-p}=(V^{\mathrm{T}})^p$ where superscript T denotes transpose and $V^0=I$.

Conditions for the existence of a solution are, at present, not well defined. It is however relevant to consider the Hessian of L, and its possible sign definiteness. Take as the independent variables a_p, b_p, where $r_p=a_p+ib_p=r^*_{-p}$, $r_0=a_0$. Set $U_p=V^p+V^{-p}$, $p\geq 0$, $T_p=V^p-V^{-p}$, $p\geq 1$. Then

$$\frac{\partial^2 L}{\partial a_p\,\partial a_\nu}=\mathrm{Tr}(R^{-1}U_pR^{-1}U_\nu)-2\,\mathrm{Tr}(R^{-1}U_pR^{-1}U_\nu R^{-1}S),$$

$$\frac{\partial^2 L}{\partial a_p\,\partial b_\nu}=i\,\mathrm{Tr}(R^{-1}U_pR^{-1}T_\nu)-i\,\mathrm{Tr}(R^{-1}U_pR^{-1}T_\nu R^{-1}S)$$
$$-i\,\mathrm{Tr}(R^{-1}T_\nu R^{-1}U_pR^{-1}S),$$

$$\frac{\partial^2 L}{\partial b_p\,\partial b_\nu}=-\mathrm{Tr}(R^{-1}T_pR^{-1}T_\nu)+2\,\mathrm{Tr}(R^{-1}T_pR^{-1}T_\nu R^{-1}S).$$

(We remark that the right-hand member in the second equation is real, despite the appearance of i.)

Let $z^T=(x_0\ \ x_1\ \ \ldots\ \ x_{K-1}\ \ y_1\ \ \ldots\ \ y_{K-1})$, $X=\sum_{p=0}^{K} x_p U_p = X^T$, $Y=\sum_{p=1}^{K-1} y_p T_p = -Y^T$, and $Z=X+iY$. Then

$$z^T\begin{bmatrix}\dfrac{\partial^2 L}{\partial a_p\,\partial a_v} & \dfrac{\partial^2 L}{\partial a_p\,\partial b_v}\\ \dfrac{\partial^2 L}{\partial a_p\,\partial a_v} & \dfrac{\partial^2 L}{\partial b_p\,\partial b_v}\end{bmatrix}z$$

$$\begin{aligned}&=\mathrm{Tr}(R^{-1}XR^{-1}X)-2\,\mathrm{Tr}(R^{-1}XR^{-1}XR^{-1}S)\\ &\quad+2i\,\mathrm{Tr}(R^{-1}XR^{-1}Y)-2i\,\mathrm{Tr}(R^{-1}XR^{-1}YR^{-1}S)\\ &\quad-2i\,\mathrm{Tr}(R^{-1}YR^{-1}XR^{-1}S)-\mathrm{Tr}(R^{-1}YR^{-1}Y)\\ &\quad+2\,\mathrm{Tr}(R^{-1}YR^{-1}YR^{-1}S)\\ &=\mathrm{Tr}[R^{-1}Z^H R^{-1}Z(I-2R^{-1}S)]\\ &=\mathrm{Tr}[R^{-1/2}ZR^{-1/2}(I-2R^{-1/2}SR^{-1/2})R^{-1/2}Z^H R^{-1/2}].\end{aligned}$$

Clearly, if

$$2R^{-1/2}SR^{-1/2}-I>0$$

or

$$2S-R>0$$

then any quadratic form derived from the Hessian is negative definite.

Hence if $\|S-R_{\mathrm{opt}}\|$ is small relative to S we will have concavity near a solution. However, the above condition may not be necessary and it may be that weaker conditions than above still allow the Hessian to be negative definite. See also [9] and [16] for other discussions on the existence of solutions.

3.2. *Minimum entropy estimation*

In Section 2 we noted how the principle of maximum entropy could be used to extrapolate a sequence of K known covariance lags. Here we apply a similar principle to the estimation of the covariance lags, i.e., choose R as that function of the observed data which minimizes the entropy.

For example, the entropy defined by

$$H_K=-\int p(x)\ln[p(x)]\,dx$$

is, for a K-dimensional Gaussian random process, given by

$$H_K=\ln[\det(R)]+K\ln\pi+K. \tag{3.3}$$

Thus, given two estimates for R, we might choose the one which, when substituted in (3.3), gave the smaller estimate of H_K, i.e., the one with the smaller determinant. However, since such a method is sensitive to scaling problems and does not explicitly involve the observed data, we must modify the approach.

Generalizing the concept of the entropy of a single event [19], an expression for the entropy or information content of a given sample of data points $\mathbf{x}$ is

$$\hat{H}_K = -\ln[p(\mathbf{x})], \tag{3.4}$$

where we use $\hat{}$ to indicate that this can be interpreted as an estimate of the entropy of the random process. For M blocks of independent realizations of the time series $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(M)}$, an appropriate generalization of (3.4) is

$$\hat{H}_{K,M} = -\frac{1}{M} \sum_{j=1}^{M} \ln[p(\mathbf{x}^{(j)})]. \tag{3.5}$$

Normally (3.5) would be interpreted as the amount of information contained in the data sequences $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(M)}$. The interpretation of $\hat{H}_{K,M}$ as an estimator is only semantically different to that above, but does have the advantage of introducing concepts such as bias, consistency, etc. For example, $\hat{H}_{K,M}$ is an unbiased estimator of H_K as $M \to \infty$. Leaving aside different interpretations of (3.5) we note that it allows us to incorporate our *a priori* information regarding the time series in the expression for $p(\mathbf{x})$. Furthermore, if $p(\mathbf{x})$ is composed of additive random processes then the concept of cross-entropy [20] can be used and the minimization of entropy proposed can be related to the minimization of cross-entropy as discussed in [21].

The minimum entropy method then implies that we choose our unknown parameters, in this case the r_μ, such that $\hat{H}_{K,M}$ is minimized. An interesting property of this is that it is equivalent to maximizing the log of the likelihood, since

$$\hat{H}_{K,M} = -\frac{1}{M} \ln[p(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(M)})] \tag{3.6}$$

as a consequence of the independence of the $\mathbf{x}^{(j)}$. Note that for dependent $\mathbf{x}^{(j)}$, (3.6) can be taken as the definition of the entropy estimator. Thus *this form of minimum entropy estimation is equivalent to maximum likelihood estimation.* As noted in [21], this holds for any parametrization of the probability density function and not just for the parametrization using the r_μ. As noted in [9], it can be shown that this problem is equivalent to

$$\min \det(R) \tag{3.7}$$

subject to

$$\mathrm{Tr}[R^{-1}S] = K. \tag{3.8}$$

Note that (3.7), apart from some additive constants, is the expression for entropy discussed at the beginning of this section and the reformulation of the problem has overcome the lack of a suitable constraint explicitly involving the data, by requiring (3.8) to be satisfied. A further insight into (3.8) can be seen for the following expression for the sample bias of the estimator $\hat{H}_{K,M}$:

$$\hat{H}_{K,M} - H_K = (K - \mathrm{Tr}((R^{-1}S)),$$

which holds for any R and follows from (3.3). Requiring (3.8) to hold is equivalent to requiring the sample bias to be zero.

3.3. *Probability-based estimators*

A number of new criteria for estimating R are proposed in this subsection and equations for the resulting estimators are derived. Conditions for the solvability of these equations are also given. The new estimators are all derived by minimizing different measures of the differences between two sets of hypotheses when the processes are Gaussian. The first hypothesis, with probability density function (pdf) $p_R(\mathbf{x})$, involves an unknown Toeplitz covariance matrix R, whilst the second, with pdf $p_S(\mathbf{x})$, involves the covariance matrix S, the unstructured maximum likelihood estimate of R. The elements of R are determined by minimizing various measures of the differences between $p_R(\mathbf{x})$ and $p_S(\mathbf{x})$. The measures considered are Kullback, divergence, and Bhattacharyya.

3.3.1. *Kullback information measure.* Let H_R be the hypothesis that the covariance matrix is R and H_S the hypothesis that it is S. The Kullback measure [22] for the mean amount of information for discriminating in favor of H_R against H_S is denoted as $I(R; S)$ and is given by

$$I(R; S) = \int \ln\left(\frac{p_R(\mathbf{x})}{p_S(\mathbf{x})}\right) p_R(\mathbf{x})\, d\mathbf{x}.$$

For Gaussian variables with zero mean this reduces to

$$I(R; S) = \ln(\det(S)) - \ln(\det(R)) + \mathrm{Tr}(RS^{-1}) - K. \tag{3.9}$$

We choose the elements of R by minimizing $I(R; S)$, i.e., minimizing the mean amount of information for discriminating in favor of H_R against H_S. Setting the gradient of the above expression to zero implies

$$\mathrm{Tr}[V^\rho(R^{-1} - S^{-1})] = 0, \qquad \rho = 0, \pm 1, \ldots, \pm(K-1). \tag{3.10}$$

The relevant quadratic form derived from the Hessian is

$$\mathrm{Tr}(Z^H R^{-1} Z R^{-1}),$$

which is positive definite for all positive definite R. Thus a unique positive definite solution to (3.10) can always be obtained.

Finally, as a brief aside, we note that, for Gaussian densities, minimizing $I(S; R)$ is equivalent to maximum likelihood. (Compare (3.9) with R and S interchanged, and (3.1).)

3.3.2. *Divergence.* A symmetrical measure of the mean difference in information between two hypotheses H_R and H_S is given by the divergence, $J(R; S)$, defined as

$$J(R; S) = I(R; S) + I(S; R).$$

Note that $J(R; S)$ has all the properties of a metric except that it does not satisfy the triangle inequality. Kullback [22] interprets $J(R; S)$ as a measure of the difficulty in discriminating between H_R and H_S and here we choose the elements of R so as to minimize $J(R; S)$.

For Gaussian densities it follows from (3.9) that

$$J(R; S) = -2K + \mathrm{Tr}[R^{-1}S] + \mathrm{Tr}[S^{-1}R].$$

Setting the gradient of $J(R; S)$ to zero yields

$$\mathrm{Tr}[V^{\rho}(S^{-1} - R^{-1}SR^{-1})] = 0, \qquad \rho = 0, \pm 1, \ldots, \pm(K-1).$$

The quadratic form obtained from the Hessian is sign definite, being given by

$$\mathrm{Tr}[(Z^{H}R^{-1}Z + ZR^{-1}Z^{H})R^{-1}SR^{-1}]$$
$$= \mathrm{Tr}(R^{-1/2}ZR^{-1}SR^{-1}Z^{H}R^{-1/2}) + \mathrm{Tr}(R^{-1/2}Z^{H}R^{-1}SR^{-1}ZR^{-1/2}).$$

Thus a unique solution can always be obtained.

3.3.3. *Bhattacharyya distance.* The Bhattacharyya measure, B, of the distance between two probability densities is defined by

$$B = -\ln \int p_R^{1/2}(x) p_S^{1/2}(x)\, dx$$

and lies between zero and unity. Its use in signal selection problems has been discussed by Kailath [17]. In this application we choose the r_ρ such that B is minimized.

For Gaussian density functions with zero mean and covariances R and S, it can be shown that

$$B = -\ln 2 - \tfrac{1}{2}\{\ln(\det(R)) - \ln(\det(S))\} + \ln(\det(R+S)).$$

Setting the gradient to zero results in

$$\mathrm{Tr}[V^{\rho}(2(R+S)^{-1} - R^{-1})] = 0, \qquad \rho = 0, \pm 1, \ldots, \pm(K-1).$$

The quadratic form obtainable from the associated Hessian matrix is

$$Q = \tfrac{1}{2}\mathrm{Tr}[Z^{H}R^{-1}ZR^{-1} - 2Z^{H}(S+R)^{-1}Z(S+R)^{-1}].$$

In general, it would not appear that this form is positive definite. Note, however, that if $R=S$, then

$$Q=\tfrac{1}{4}\operatorname{Tr}[Z^{H}R^{-1}ZR^{-1}]>0,$$

which implies that if the optimum is sufficiently close to S, then we have convexity. In fact, we can show that convexity holds if

$$S+R>\sqrt{2}R.$$

For this inequality implies

$$R^{-1}>\sqrt{2}(S+R)^{-1}$$

and, for any nonzero Z,

$$Z^{H}R^{-1}Z\geq\sqrt{2}Z^{H}(S+R)^{-1}Z$$

and then $Q>0$ follows. (If $A>B>0$, and $C\geq D\geq 0$, $C\neq D$, $D\neq 0$, then $\operatorname{Tr}(AC)>\operatorname{Tr}(BD)$.)

We would expect S to be close to R when a sufficiently large number of integrations is used, so that the condition $S+R>\sqrt{2}R$ would be fulfilled.

3.4. *Matrix approximation methods*

In this section a class of estimation criteria are considered which are based upon various measures of the similarity of the two matrices R and S. As for the probability-based estimates, the elements of R are determined by minimizing a measure of the difference between R and S. Equations for the resulting estimators are derived and conditions for the existence of unique solutions are investigated.

A number of the criteria considered below have already been proposed in considering the problem of estimating the powers of a finite number of plane waves from known directions incident upon an array of receivers [23]. Where possible, the same nomenclature as in [23] is used.

In the following the norm of a matrix is defined by $\|A\|^2$ and is given by

$$\|A\|^2=\sum_{i,j}|a_{ij}|^2=\operatorname{Tr}(AA^{H}).$$

3.4.1. *Least squares fit.* Here the r_μ are determined by minimizing the norm of the difference between R and S, i.e., $\|R-S\|^2$. Differentiating

$$\operatorname{Tr}[(R-S)^2]$$

and equating the result to zero implies

$$\operatorname{Tr}[V^{\mu}R]=\operatorname{Tr}[V^{\mu}S].$$

Since R is Toeplitz it can easily be seen that the solution is

$$\hat{r}_{\mu}=\frac{1}{K-|\mu|}\sum_{p=1}^{K-|\mu|}(S)_{p+|\mu|,p}.$$

This is readily recognized as the unbiased estimator that is commonly used in practice.

3.4.2. *Inverse least squares.* In many applications, e.g., adaptive beamforming, it is not R but R^{-1} which is used. In this case it may be more appropriate to estimate the r_p by minimizing the distance between R^{-1} and S^{-1}.

It can readily be shown that the stationary points of

$$\mathrm{Tr}[(R^{-1}-S^{-1})^2]$$

are determined by the solutions of

$$\mathrm{Tr}[V^p(R^{-1}(R^{-1}-S^{-1})R^{-1})]=0.$$

The quadratic form definable with the Hessian is

$$\begin{aligned}Q=&2\,\mathrm{Tr}[R^{-1/2}ZR^{-1}(R^{-1}-S^{-1})R^{-1}Z^H R^{-1/2}]\\&+2\,\mathrm{Tr}[R^{-1/2}Z^H R^{-1}(R^{-1}-S^{-1})R^{-1}ZR^{-1/2}]\\&+2\,\mathrm{Tr}[R^{-2}ZR^{-2}Z^H].\end{aligned}$$

While this is not clearly positive for all Toeplitz positive definite R and all nonzero (Toeplitz) Z, it is certainly positive when R is near S, since when $R=S$, there holds

$$Q=2\,\mathrm{Tr}(R^{-2}ZR^{-2}Z^H)=2\,\mathrm{Tr}(R^{-1}ZR^{-2}Z^H R^{-1})>0.$$

3.4.3. *Approximate inverse.* If R and S are close then both $R^{-1}S$ and $S^{-1}R$ approximate the identity matrix. In this case two estimates, termed type A or B, are given by choosing the r_p such that either

$$\|I-RS^{-1}\|$$

or

$$\|I-SR^{-1}\|$$

is minimized. Differentiating the above expressions and equating the results to zero gives the following equations

$$\mathrm{Tr}[V^p((RS^{-2}+S^{-2}R)-2S^{-1})]=0 \tag{3.11}$$

and

$$\mathrm{Tr}[V^p(2R^{-1}SR^{-1}-R^{-1}S^2R^{-2}-R^{-2}S^2R^{-1})]=0, \tag{3.12}$$

respectively.

Expressing

$$R=\sum_{-K+1}^{K-1} r_p V^p,$$

(h) Approx inverse (B)

Cost function $\mathrm{Tr}[(I-SR^{-1})^2]$.

Gradient $\mathrm{Tr}[V^p(2R^{-1}SR^{-1}-R^{-1}S^2R^{-2}-R^{-2}S^2R^{-1})]$.

Quadratic form
$$\begin{aligned}\mathrm{Tr}[&-2R^{-1}ZR^{-1}SR^{-1}Z^H-2R^{-1}SR^{-1}ZR^{-1}Z^H\\&+R^{-1}ZR^{-1}S^2R^{-2}Z^H+R^{-1}S^2R^{-1}ZR^{-2}Z^H\\&+R^{-1}S^2R^{-2}ZR^{-1}Z^H+R^{-1}ZR^{-2}S^2R^{-1}Z^H\\&+R^{-2}ZR^{-1}S^2R^{-1}Z^H+R^{-2}S^2R^{-1}ZR^{-1}Z^H].\end{aligned}$$

Finally, we note that $S\to R_e$, a Hermitian Toeplitz matrix for a spatially stationary process, in the limit as $M\to\infty$. Furthermore, it can easily be verified that $R=R_e$ results in zero gradient for all the estimators discussed above. Thus the solutions to the above equation are asymptotically unbiased.

The above expressions can readily be extended to arbitrary arrays by noting that in the above expressions V^p is the derivative of R with respect to r_p. Thus for linear arrays with element spacings taking integer values, the expressions for the gradients are obtained by replacing V_p by $\partial R/\partial r_p$ which is a matrix whose ijth element is zero unless the exact spatial covariance between the ith and jth elements is r_p, in which case the ijth element is unity. For the example of a three element linear array with receivers located at 0, d, and $3d$, we have

$$\left(\frac{\partial R}{\partial r_0}\right)_{ij}=\delta_{ij},$$

$$\left(\frac{\partial R}{\partial r_1}\right)_{ij}=\delta_{i1}\delta_{j2},$$

$$\left(\frac{\partial R}{\partial r_2}\right)_{ij}=\delta_{i2}\delta_{j3},$$

and

$$\left(\frac{\partial R}{\partial r_3}\right)_{ij}=\delta_{i1}\delta_{j3}.$$

4. Solutions of the equations

At the end of the last section, we summarized the equations defining various optimal estimates. We begin by discussing the closeness of solutions of different equations.

4.1. *Closeness of different optima*

Suppose that sufficient measurements are taken that the unstructured (sample) covariance matrix S is close to R. This means that

$$R=S+\varepsilon T \tag{4.1}$$

for some T with $\|T\| \approx \|R\|$, and ε small. Neglecting second- and higher-order terms in ε, the following approximation results for maximum likelihood, Kullback, divergence, Bhattacharyya, and approximate inverse (B):

$$\mathrm{Tr}[V^p(\varepsilon S^{-1}TS^{-1})] = 0. \tag{4.2}$$

Equation (4.2) puts K complex constraints on an Hermitian εT, and (4.1), through the Toeplitz nature of R, means that knowledge of S and the first row of εT fully determines εT. Consequently, εT can be determined from (4.1) and (4.2), which constitute a linear collection of equations. This means that solutions for the five criteria listed will be close.

In fact, the divergence and Bhattacharyya solutions are more tightly related again. For with (4.1) holding, to second order in ε, the divergence gradient is

$$\mathrm{Tr}[V^p(S^{-1} - R^{-1}SR^{-1})] = 2\varepsilon S^{-1}TS^{-1} - 3\varepsilon^2 S^{-1}TS^{-1}TS^{-1}$$

and the Bhattacharyya gradient is

$$\mathrm{Tr}[V^p[2(R+S)^{-1} - R^{-1}]] = \frac{\varepsilon}{2} S^{-1}TS^{-1} - \frac{3\varepsilon^2}{4} S^{-1}TS^{-1}TS^{-1}$$

and, to within a scaling constant, these are the same.

While the 2×2 case is only of academic interest, we note that when

$$S = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix},$$

the optimum solutions for divergence and Bhattacharyya are identical, being

$$R = \sqrt{\frac{4(\alpha\gamma - \beta^2)}{(\alpha+\gamma)^2 - 4\beta^2}} \begin{bmatrix} (\alpha+\gamma)/2 & \beta \\ \beta & (\alpha+\gamma)/2 \end{bmatrix},$$

while for maximum likelihood and inverse least squares the solutions are also identical:

$$R = \begin{bmatrix} (\alpha+\gamma)/2 & \beta \\ \beta & (\alpha+\gamma)/2 \end{bmatrix}.$$

Note as $M \to \infty$ then $\alpha \to \gamma \to r_0$ and all the above estimators become equivalent.

4.2. *Methods of solution*

Here we briefly discuss a number of methods that can be used to find approximate solutions of the nonlinear equations derived in the previous sections. Let $g(S, R)$ denote any of the gradients summarized in Section 3.5.

(a) *Gradient methods*. In cases where the Hessian is single signed a gradient-based method can be used to find either the minimum or the maximum. As an example, if $r_\rho(k)$ is the kth approximation to the solution then a steepest descent method can be used to update the r_ρ, i.e.,

$$r_\rho^{(k+1)} = r_\rho^{(k)} - \mu g_\rho(S, R^{(k)}),$$

where $g_\rho(S, R^{(k)})$ is the gradient evaluated at $r_\rho = r_\rho^{(k)}$. The step size is μ, which can, in principle, be chosen to be optimum. Alternatively, a more efficient Newton-Raphson type method may be used, i.e.,

$$r_\rho^{(k+1)} = r_\rho^{(k)} - \mu \sum_\nu (H^{-1})_{\rho\nu} g_\nu(S, R^{(k)}),$$

where H^{-1} is the inverse of the Hessian evaluated at the point $\mathbf{r}^{(k)}$. Whilst being more efficient the computational loads involved in evaluating $(H^{-1})_{\rho\nu}$ may be prohibitive.

(b) *Inverse iteration*. This method was introduced in [9] to utilize the fact that although the maximum likelihood equations are nonlinear in R they are linear in S. This allows a linear Toeplitz perturbation on S, i.e., D, to be found such that the perturbed S, i.e., $S - D$, and current estimate of R satisfy the equation. Then one obtains a new value for R by adding D to the current value. One terms D an improving direction as it can be shown to have a positive projection along the gradient [9]. Thus updating R in the direction of D will tend to force R in the general direction of a minimum.

Thus, for maximum likelihood, the inverse iterative method becomes:

(1) Let $R^{(k)}$ be the kth approximation to the solution.

(2) Find an Hermitian Toeplitz matrix $D^{(k)}$ such that

$$g(S - D^{(k)}, R^{(k)}) = 0,$$

i.e.,

$$\mathrm{Tr}[V^\rho R^{(k)-1} D^{(k)} R^{(k)-1}] = \mathrm{Tr}[V^\rho (R^{(k)-1} S R^{(k)-1} - R^{(k)-1})]. \tag{4.3}$$

In component form this becomes the following set of linear equations:

$$\sum_\nu \mathrm{Tr}[V^\rho R^{(k)-1} V^\nu R^{(k)-1}] d_\nu^{(k)} = \mathrm{Tr}[V^\rho (R^{(k)-1} S R^{(k)-1} - R^{(k)-1}].$$

(3) Set

$$R^{(k+1)} = R^{(k)} + D^{(k)}.$$

5. Beamforming and examples

The estimates derived above can be used for conventional and optimum beamforming. It is shown here that, when some of the above estimators are used in optimum beamforming, the resulting output power can be interpreted in terms of the output power of an optimum beamformer using the unstructured covariance matrix S. Some examples are then presented comparing the performance of the various estimators.

5.1. *Relation to optimum beamforming*

We first consider the following lemma.

Lemma. *Let A be an arbitrary matrix and $\mathbf{v}(\theta, f)$ be a steering vector as defined in Section 2. Then*

$$\mathbf{v}^{H}(\theta, f) A \mathbf{v}(\theta, f) = \sum_{\rho=-K+1}^{K-1} z^{\rho} \operatorname{Tr}[V^{\rho} A],$$

where

$$z = \exp\left(2\pi i \frac{d \sin\theta}{\lambda}\right).$$

Proof. Since $(\mathbf{v}(\theta, f))_j = z^{j-1}$ it follows that

$$\begin{aligned} \mathbf{v}^{H}(\theta, f) A \mathbf{v}(\theta, f) &= \sum_{k,j=1}^{K} z^{k-j} A_{jk} \\ &= \sum_{\rho=-K+1}^{K-1} z^{\rho} \sum_{\nu=1}^{K-|\rho|} A_{\nu,\rho+\nu} \\ &= \sum_{\rho=-K+1}^{K-1} z^{\rho} \operatorname{Tr}[V^{\rho} A]. \end{aligned}$$ □

Now apply this lemma to the stationary points of the gradients summarized in Section 3.5.

(a) *Maximum likelihood.* Taking A as $R^{-1} - R^{-1} S R^{-1}$ and using the lemma implies

$$\begin{aligned} &\mathbf{v}^{H}(\theta, f) R^{-1} \mathbf{v}(\theta, f) - \mathbf{v}^{H}(\theta, f) R^{-1} S R^{-1} \mathbf{v}(\theta, f) \\ &\quad = \sum_{\rho=-K+1}^{K+1} z^{\rho} \operatorname{Tr}[V^{\rho}(R^{-1} - R^{-1} S R^{-1})] \\ &\quad = 0. \end{aligned}$$

Multiplying the above by $(\mathbf{v}^{H}(\theta, f) R^{-1} \mathbf{v}(\theta, f))^{-1}$ results in the following equation:

$$\mathbf{w}^{H} R \mathbf{w} = \mathbf{w}^{H} S \mathbf{w}, \tag{5.1}$$

where

$$\mathbf{w} = R^{-1} \mathbf{v}(\theta, f) / \mathbf{v}^{H}(\theta, f) R^{-1} \mathbf{v}(\theta, f).$$

This is the optimum weight vector in the case where the estimated R is assumed to be the exact cross-spectral matrix of receiver outputs. Equation

(5.1) states that the sample output power of such a processor, i.e., $\mathbf{w}^H S \mathbf{w}$, is the same as the mean output using the assumed cross-spectral matrix.

(b) *Kullback.* Taking A as $R^{-1}-S^{-1}$ and using the lemma implies

$$[\mathbf{v}^H(\theta,f)R^{-1}\mathbf{v}(\theta,f)]^{-1}=[\mathbf{v}^H(\theta,f)S^{-1}\mathbf{v}(\theta,f)]^{-1}.$$

Thus, using the Kullback estimator of R results in the same expression for the estimate of the output power of an optimum beamformer as is obtained by using the unstructured covariance matrix S.

(c) *Divergence.* Taking A as $S^{-1}-R^{-1}SR^{-1}$ and using the lemma implies

$$\mathbf{v}^H(\theta,f)S^{-1}\mathbf{v}(\theta,f)=\mathbf{v}^H(\theta,f)R^{-1}SR^{-1}\mathbf{v}(\theta,f).$$

The inverse of the quantity on the left is an estimate of the output power of an optimum beamformer using the unstructured covariance S. The term on the right is the sample output power of a minimum noise beamformer under the assumption that R, the divergence estimate of the noise power, is the true cross-spectral matrix.

(d) *Bhattacharyya.* The corresponding equation for the Bhattacharyya estimator is

$$2\mathbf{v}^H(\theta,f)(R+S)^{-1}\mathbf{v}(\theta,f)=\mathbf{v}^H(\theta,f)R^{-1}\mathbf{v}(\theta,f).$$

This does not have a ready physical interpretation but we recall from Section 4.1 that, to $O(\varepsilon^2)$, the divergence and Bhattacharyya estimates of R will be identical. Thus we may expect the optimum output powers to be close.

(e) *Inverse.* Consideration of the inverse least squares, and approximate inverses (A) and (B) leads to the following equations, respectively:

$$\mathbf{v}^H(\theta,f)R^{-3}\mathbf{v}(\theta,f)=\mathbf{v}^H(\theta,f)R^{-1}SR^{-1}\mathbf{v}(\theta,f),$$

$$\mathbf{v}^H(\theta,f)(RS^{-2}+S^{-2}R)\mathbf{v}(\theta,f)=2\mathbf{v}^H(\theta,f)S^{-1}\mathbf{v}(\theta,f),$$

and

$$\mathbf{v}^H(\theta,f)(R^{-1}S^2R^{-2}+R^{-2}S^2R^{-1})\mathbf{v}(\theta,f)=2\mathbf{v}^H(\theta,f)R^{-1}SR^{-1}\mathbf{v}(\theta,f).$$

Although each quantity on the right-hand side of the above equations can be interpreted in terms of the output (or inverse) of an optimum beamformer, the expressions on the left-hand side are too complicated to allow simple interpretations.

Finally, we observe that the above expressions are not unique to linear array equispaced receivers. By replacing V'' by the appropriate matrix, i.e., $\partial R/\partial r_n$, as discussed at the end of Section 3.5, the lemma can readily be generalized to arbitrary arrays. Thus, the results above hold for arbitrary arrays.

5.2. *Examples*

Here we consider the application of the least squares and probability-based estimators to simulated data from an array of equispaced receivers. Estimation of the spatial covariance is first considered and then some examples using these structured estimates of the cross-spectral matrix to estimate the angular power spectrum of the output of an optimum beamformer are given.

(a) *Methodology*. For a given angular distribution of the signal and noise fields the outputs of an array of K receivers were simulated at the frequency of interest. This was done by generating, at the desired value of d/λ, the *exact* cross-spectral matrix of the receiver outputs appropriate to the chosen signal and noise angular distributions. Using an eigenvector decomposition of the *exact* cross-spectral matrix, the receiver outputs, i.e., the $x_j^{(m)}(f)$, $j=1,2,\ldots,K$, were simulated for M independent Gaussian realizations. The sample cross-spectral matrix S was then estimated using (2.3).

The least squares estimates of the r_p were found by averaging along the pth minor diagonal of S. The probability-based estimates were calculated by an interactive NAG library subroutine (CO5NBF). As an input to the subroutine, the gradient at each step of the iteration was calculated by using the expressions in Section 3.5. The convergence was achieved by a mixture of gradient and Newton-Raphson techniques. For some realizations, i.e., a given S, the iterative procedure failed to converge, usually because at some stage in the iteration, the iterative value of R was not positive definite. The algorithm was modified by testing the positive definiteness of R at each iteration step and successively halving the step size until the next iterative value for R was positive definite. This modification greatly improved the convergence properties of the algorithm.

For a given number of averages, M, the structured estimate of the cross-spectral matrix was substituted in the angular power spectrum formula (2.4) and the result plotted as a function of these θ, the angle relative to broadside. We term these estimates of R and of the angular spectrum, "snapshots" as they correspond to a given set of M realizations. To obtain more meaningful statistical results, a large number (typically 400) of snapshots of both the covariances and the spectra were generated and the means and standard deviations of these calculated. Some example results are discussed below.

(b) *Covariance estimates*. We discuss here some example results for an array of four receivers with $M=16$. Two noise fields were considered; taking the isotropic noise level as a reference, in noise field A the level of uncorrelated receiver self noise was set at 10 dB whilst in B it was −10 dB. A single plane wave incident upon the array from 25° or 90° with SNR either 16 or 6 dB for noise field A or B, respectively, was considered. The frequency was such that $d = \lambda/2$ or $\lambda/4$.

In all the above cases the Kullback estimator was found to be biased by a constant factor which was empirically measured as 1.175. (The bias factor was found by scaling the means of the Kullback results until they equalled those of the unbiased least squares estimator.) An example of the mean and the mean plus or minus one standard deviation of the covariance is illustrated in Figures 2 and 3, which depict the real part and the imaginary part of the estimated covariance. In the figure, the Kullback estimates are presented after scaling by 1.175 and for the chosen parameters both least squares and the scaled Kullback estimates gave almost identical results. Similarly, the Bhattacharyya estimate was found to have a bias factor of 1.08 and maximum likelihood was unbiased. In all cases considered the divergence results obtained were indistinguishable from the Bhattacharyya ones; this result follows from Section 4.1.

The bias of the Kullback results is certainly not unexpected since, from Section 5.1, the Kullback covariance estimate produces the same estimate of the optimum angular spectrum as does S. As discussed in [24] the estimates of the optimum angular power spectrum using S are biased by a

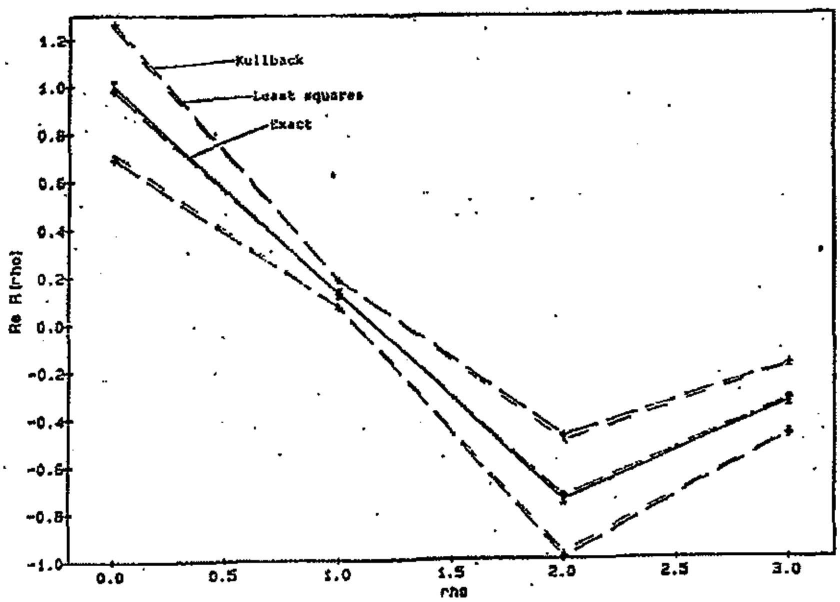

Figure 2. Real part of covariance with noise field A, SNR = 16 dB, $d/\lambda = 0.25$, $M = 16$, and $K = 4$.

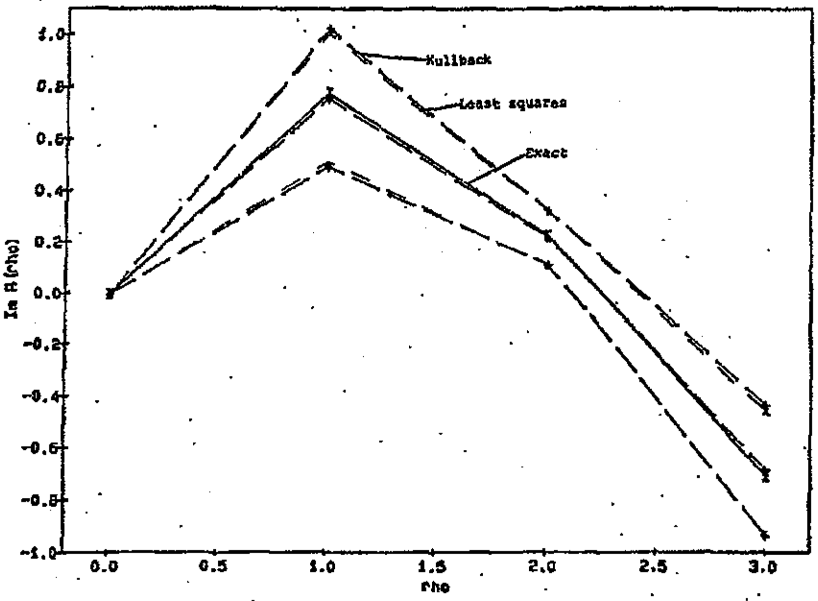

Figure 3. Imaginary part of covariance with noise field A, SNR = 16 dB, $d/\lambda = 0.25$, $M = 16$, and $K = 4$.

factor $M/M - K + 1$. From this we would expect the bias of the Kullback covariance estimates to depend only on M and K, as was found to be the case.

(c) *Estimates of angular spectrum of optimum beamformer.* To illustrate differences between the probability-based estimators and the least squares estimator the number of averages, i.e., M, was reduced, while K, the number of receivers, was increased and the complexity and strength of the signal field was increased. All of these modifications resulted in a covariance matrix with a larger condition number. Here we present some example angular snapshot spectra.

In Figures 4 and 5 the snapshot angular spectra for $M = 4$ and 16 are shown for the least squares, Kullback, Bhattacharyya, and maximum likelihood estimates. The noise field was A and d equalled $\lambda/2$. Also plotted are the angular spectra of the output of an optimum beamformer when the cross-spectral matrix is known exactly and the one derived using the unstructured estimator S. In this and all cases considered, the use of S gave identical results to the use of the Kullback estimator, thus confirming that the Kullback estimator had converged. As the number of integrations is decreased, the probability-based estimators showed an increased improvement over the least squares one. Apart from the different bias factors the

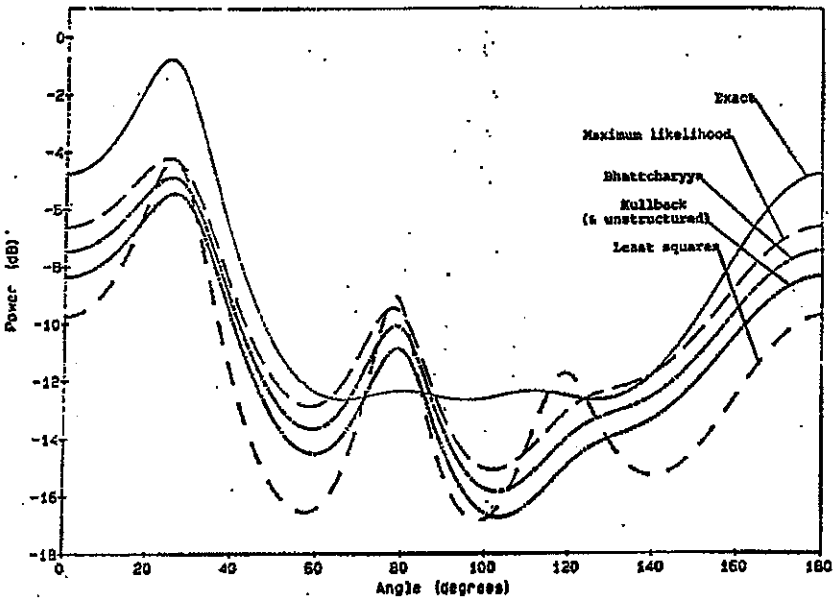

Figure 4. Snapshot angular spectra noise field A, SNR = 16 dB, $d/\lambda = 0.5$, $M = 4$, and $K = 4$.

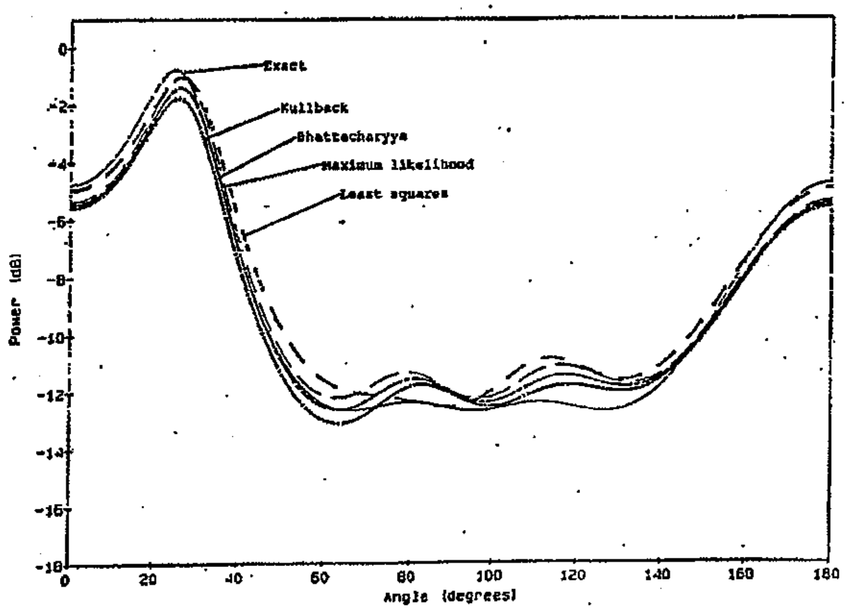

Figure 5. Snapshot angular spectra noise field A, SNR = 16 dB, $d/\lambda = 0.5$, $M = 16$, and $K = 4$.

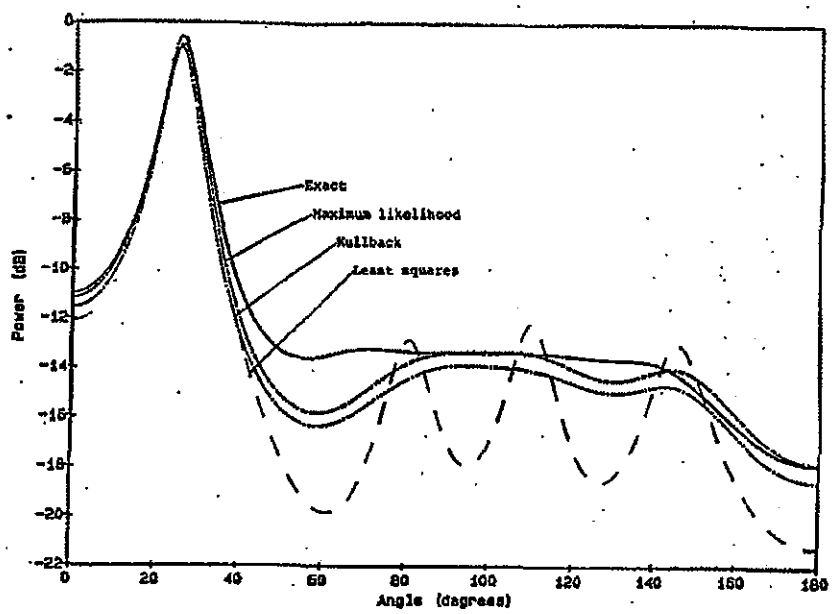

Figure 6. Snapshot angular spectra noise field B, SNR = 6 dB, $d/\lambda = 0.25$, $M = 32$, and $K = 8$.

probability-based estimators all gave roughly equivalent results. This was also observed for many other simulations.

In Figure 6 eight receivers were used and again a significant difference between the least squares and the probability-based estimators is seen. Finally, results for a complex distribution of three signals, 16, 10 and 16 dB, relative to the isotropic noise level with $d = \lambda/2$ are illustrated Figure 7. Again a significant difference between the least squares and the probability-based estimators is apparent, particularly regarding biases in the peaks corresponding to signal arrival directions.

6. Conclusions

For a linear array of equispaced receivers the exact cross-spectral matrix of receiver outputs has an Hermitian Toeplitz structure. New estimators of the cross-spectral matrix, constrained to have this structure, have been proposed. These estimators were derived using information theoretic and matrix measures of the distance between structured and unstructured estimates of the cross-spectral matrix.

Some example applications of the methods to estimating the spatial covariances of simulated data have indicated biases in some of the estimators. Further quantification of the biases is required.

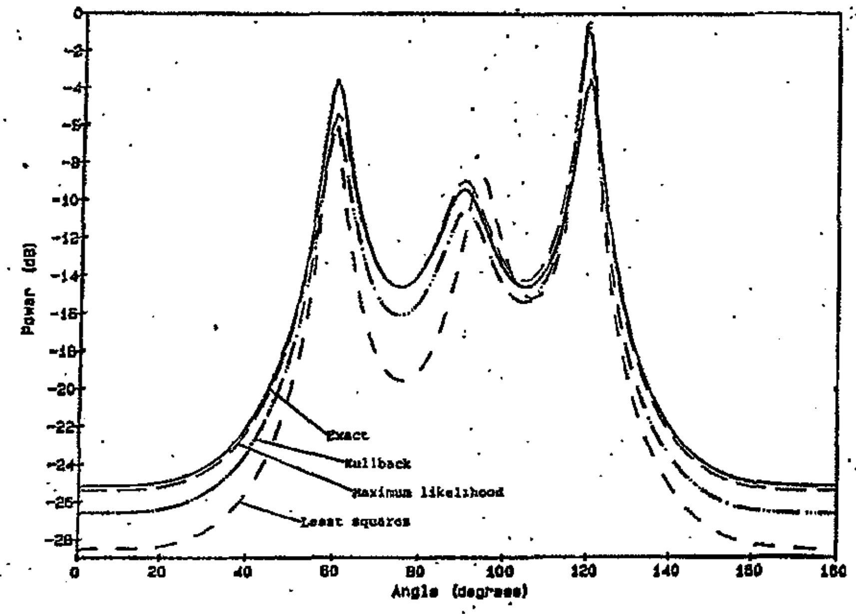

Figure 7. Snapshot angular spectra noise field B, SNR's = 16, 10, and 16 dB, d/λ = 0.5, M = 8, and K = 4.

Estimators based upon information theoretic measures have been shown to be closely related to optimum beamformers. Estimates of the optimum angular power spectrum of simulated data using the information theoretic estimators have been compared with those using unstructured estimates or the unbiased least squares estimator. Theoretical results indicating the equivalence of the methods when M is large (typically $M \geq 4$ K) with the condition number of the observed unstructured estimate not too large were confirmed. Theoretical predictions of the approximate agreement of the divergence and Bhattacharyya results and that of the Kullback and unstructured estimates were also confirmed. Indeed; in most simulations the information theoretic estimators gave almost equivalent results. When the number of integrations was reduced (i.e., typically $M \leq 2$ K) or the condition number of the unstructured estimate of the cross-spectral matrix was increased by either increasing the number of receivers or by increasing the signal to noise ratio the information-theoretic-based estimators showed significant improvements over the unbiased least squares one.

References

[1] R. A. Mozingo and T. W. Miller, *Introduction to Adaptive Arrays*, Wiley-Interscience, New York, 1980.

[2] J. E. Hudson, *Adaptive Array Principles*, Peter Peregrinus, 1981.

[3] D. A. Gray, Theoretical and experimental comparisons of the formulation of optimum processors in element, beam and eigenspaces, in *Adaptive Methods in Underwater Acoustics* (H. G. Urban, ed.), Reidel, Dordrecht.

[4] R. N. McDonaugh, Application of the maximum-likelihood and maximum-entropy method to array processing, in *Nonlinear Methods of Spectral Analysis* (S. Haykin ed.), Chap. 6, Springer-Verlag, New York, 1979.

[5] D. H. Johnson, The application of spectral estimation methods to bearing estimaton problems, *Proc. IEEE*, **70**, 1018–1028, 1982.

[6] N. Levinson, The Wiener RMS (root mean square) error in filter design, *J. Math. Phys.*, **25**, 261–278, 1947.

[7] W. F. Trench, An algorithm for the inversion of finite Toeplitz matrices, *J. Soc. Indust. Appl. Math.*, **12**, 515–522, 1964.

[8] I. S. Iohvidov, *Hankel and Toeplitz Matrices and Forms*, Birkhauser, Boston, 1982.

[9] J. P. Burg, D. G. Luenberger, and D. L. Wenger, Estimation of structured covariances, *Proc. IEEE*, 70, 963–974, 1982.

[10] A. T. Moffett, Minimum redundancy linear arrays, *IEEE Trans. Antennas and Propogation*, **16**, 172–175, 1968.

[11] D. H. Johnson and S. De Graaf, Improving the resolution of bearing in passive sonar arrays by eigenvalue analysis, *IEEE Trans. Acoust. Speech Signal Process.*, **30**, 638–647, 1982.

[12] G. Bienvenu and L. Kopp, Source power estimation associated with high resolution bearing estimation, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Atlanta, GA, 153–156, 1981.

[13] A. Cantoni and L. C. Godara, Resolving the directions of sources in a correlated field incident on an array, *J. Acoust. Soc. Amer.*, **67**, 1247–1255, 1980.

[14] C. L. Byrne and A. K. Steele, Stable nonlinear methods for sensor array processing, *IEEE J. Oceanic Engng*, **10**, 255–259, 1985.

[15] D. A. Gray, Maximum entropy estimates of the frequency wavenumber power spectrum using a linear array of equispaced receivers, Technical Report WSRL-0196, Defence Research Centre, Salisbury, Adelaide, South Australia, 1981.

[16] N. Q. A, On the uniqueness of the maximum-likelihood estimate of structured covariance matrices, *IEEE Trans. Acoust. Speech Signal Process.*, **32**, 1249–1251, 1984.

[17] T. Kailath, The divergence and Bhattacharyya distance measures in signal selection, *IEEE Trans. Comm.*, **15**, 52–60, 1967.

[18] N. R. Goodman, Statistical analysis based on a certain multivariate complex Gaussian distribution, *Ann. Math. Statist.*, **34**, 152–176, 1963.

[19] J. Aczél and Z. Daróczy, *On Measures of Information and Their Characterizations*, Academic Press, New York, 1975.

[20] J. E. Shore and R. W. Johnson, Automatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, *IEEE Trans. Inform. Theory*, 26, 26–27, 1980.

[21] J. E. Shore, On a relation between maximum likelihood classification and minimum relative entropy classification, *IEEE Trans. Inform. Theory*, 30, 851–854, 1984.

[22] S. Kullback, *Information Theory and Statistics*, Dover, New York, 1968.

[23] H. A. d'Assumpaco, Some new signal processors for arrays of sensors, *IEEE Trans. Inform. Theory*, 26, 441–453, 1982.

[24] J. Capon and J. R. Goodman, Probability distributions for estimators of the frequency-wavenumber spectrum, *Proc. IEEE*, **58**, 1785–1786, 1970.