ISSPA 87, signal processing, theories, implementations and Applications.
Editor: B. Boashash, Brisbane, Australia
24-28 August, 1987.

The Estimation of Structured Covariances for use in Array Beamforming

D. A. GRAY
Senior Research Scientist, Weapons Systems Research Laboratory, Defence Science and Technology Organisation, Department of Defence, GPO Box 2151, Adelaide, SA 5001

B. D. O. ANDERSON
Professor, Department of Systems Engineering, Research School of Physical Sciences, Australian National University, GPO Box 4, Canberra, ACT 2601

P. K. SIM
Lecturer, Department of Electronic and Communication Science, La Trobe University, Bundoora, Victoria, 3083

1. INTRODUCTION

In frequency-domain array beamforming the output power can be written as an Hermitian form of the cross-spectral matrix of the receiver outputs for both conventional and optimum beamforming algorithms [1-3]. Furthermore, in a number of bearing estimation algorithms [4-5] the eigenvalues and eigenvectors of the cross-spectral matrix play a central role.

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 and for a linear array of equispaced receivers, it is Toeplitz [6]. 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. 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.

Although the results of this paper can be extended to arbitrary arrays attention here is restricted to linear arrays of equispaced receivers. In this case the problem becomes that of estimating an Hermitian Toeplitz matrix. First the maximum likelihood estimator of such a covariance matrix is discussed [7-9] and its relation to minimum cross-entropy estimation is outlined.

A number of new estimators are also proposed [9]; these are based upon various information theoretic measures of the differences between two hypotheses. Non-linear equations for the new estimators are derived and some sample illustrations are given to illustrate both statistical properties of these estimators and to compare their performance when used in beamforming applications.

The estimators are shown to be closely related to optimum beamformers in interesting but surprising ways.

2. NOTATION AND REVIEW OF BEAMFORMING

For frequency-domain beamforming, the narrowband receiver outputs are multiplied by phase factors proportional to the relative time delay, $\tau_j(\theta)$, of the jth receiver and the output of a conventional beamformer, p_c, steered in the direction θ is given by

$$p_c = \underline{v}^H(\theta,f)\ R(f)\ \underline{v}(\theta,f)$$

where

$$R(f) = \langle \underline{x}(f)\underline{x}^H(f) \rangle$$

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 realisations of the receiver outputs by replacing the ensemble average with a simple linear average, e.g. given M samples of the receiver outputs

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

the cross-spectral matrix of receiver outputs is estimated as

$$S = \hat{R}\ (f)$$

$$= \frac{1}{M} \sum_{i=1}^{M} \underline{x}^{(i)}(f)\underline{x}^{(i)H}(f).$$

In general, this estimator will not be structured.

3. ESTIMATION OF FINITE DIMENSIONAL TOEPLITZ MATRICES

Here we consider a number of different estimators of a finite K-dimensional Toeplitz matrix, R, from M independent realisations of the receiver outputs.

3.1 Maximum likelihood estimation

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

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

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 maximised is given by

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

where S is given by

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

and is the maximum likelihood estimator of the unstructured covariance matrix. Of course, the R maximising L is the structured covariance matrix estimate. Maximising results in the following set of equations

$$\mathrm{Tr}(V^{\rho}R^{-1}) = \mathrm{Tr}(V^{\rho}R^{-1}SR^{-1}), \ \rho = 0,\pm1,\ldots\pm(K-1)$$

for R. The matrix V is defined as

$$V = \begin{bmatrix} 0 & 1 & 0 & . & . & . & 0 \\ 0 & 0 & 1 & 0 & & & . \\ & & . & & . & & . \\ . & & & . & & . & . \\ . & & & & . & . & . \\ . & & & & & . & 0 \\ & & & & & 0 & 1 \\ 0 & . & . & . & . & 0 & 0 \end{bmatrix}$$

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

Conditions for the existence of a solution are, at present, not well defined.

However if

$$2S - R > 0$$

then any quadratic form derived from the Hessian is negative-definite. [7,8,9].

Hence if $||S-R_{opt}||$ 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 [8] for other discussions on the existence of solutions.

3.2 Minimum Entropy Estimation

In [9] it is argued that

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

can be interpreted as the amount of information contained in the data sequences

$\underline{x}^{(1)}, \underline{x}^{(2)}, \ldots, \underline{x}^{(M)}$ or alternatively as an estimator of the entropy of a set of data samples.

The minimum entropy method implies that we choose our unknown parameters, in this case r_ρ the elements of R, such that $\hat{H}_{K,M}$ is minimised.

An interesting property of this is that it is equivalent to maximising the log of the likelihood since

$$\hat{H}_{K,M} = -\ln[p(\underline{x}(1), \underline{x}(2),\ldots, \underline{x}(M))]$$

as a consequence of the independence of the $\underline{x}^{(j)}$. (Note that for dependent $\underline{x}^{(j)}$, the above 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 [10], this holds for any parametrisation of the probability density function and not just for the parametrisation using the r_ρ.

3.3 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 [11] for the mean amount of information for discriminating in favour of H_R against H_S is denoted as I(R:S) and is given by

$$I(R:S) = \int \ln\left\{\frac{p_R(x)}{p_S(x)}\right\} p_R(x)dx.$$

For Gaussian variables with zero mean this reduces to

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

We choose the elements of R by minimising I(R:S) i.e. minimising the mean amount of information for discriminating in favour of H_R against H_S. Setting the gradient of the above expression to zero implies

$$\mathrm{Tr}[V^{\rho}(R^{-1} - S^{-1})] = 0, \quad \rho = 0, \pm1,\ldots,\pm(K-1)$$

The relevant quadratic form derived from the Hessian can be shown to be positive definite for all positive definite R. Thus a unique positive definite solution to the above equation can always be obtained.

3.4 Divergence

A symmetric 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

$$J(R:S) = I(R:S) + I(S:R)$$

Kullback [11] 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 minimise J(R:S).

For Gaussian densities it can be shown 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

$$\mathrm{Tr}[V^{\rho}(S^{-1} - R^{-1}SR^{-1})] = 0, \quad \rho = 0,\pm1,\ldots,\pm(K-1)$$

The quadratic form obtained from the Hessian is sign definite, and thus a unique solution can always be obtained.

3.5 Bhattacharyya Distance

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

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

and lies between zero and unity. In this application we choose the r_ρ such that B is minimised.

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

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

The quadratic form obtainable from the associated Hessian matrix can be shown to be positive definite when a sufficiently large number of integrations are used.

3.6 Least Squares

The least squares estimate of the structured covariance is obtained from S by averaging along the appropriate diagonal. In [9] it is shown that this minimises the Frobenius norm of the difference between R and S. This estimate is not always guerenteed to be positive definite.

4. RELATIONSHIP TO OPTIMUM BEAMFORMERS

It is shown here that, when 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.

We first state without proof the following lemma. [9].

<u>Lemma</u>: Let A be an arbitrary matrix and $\underline{v}(\theta,f)$ be a steering vector as defined above. Then

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

where

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

(a) Maximum Likelihood

Taking A as $\hat{R}^{-1} - \hat{R}^{-1}S\hat{R}^{-1}$ and using the lemma results in the following equation

$$\underline{w}^H\hat{R}\underline{w} = \underline{w}^H S\underline{w}$$

where

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

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

(b) Kullback

Taking A as $\hat{R}^{-1} - S^{-1}$ and using the lemma implies

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

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 $S^{-1} - \hat{R}^{-1}S\hat{R}^{-1}$ and using the lemma implies

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

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 $\hat{R}$, the divergence estimate of the noise power, is the true cross-spectral matrix.

(d) Bhattacharyya

The corresponding equation for the Bhattacharyya estimator is

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

This does not have a ready physical interpretation but as shown in [9] the divergence and Bhattacharyya estimates of R will often be identical. Thus we may expect the optimum output powers to be close.

Finally, we observe that the above expressions are not unique to a linear array of equispaced receivers. By replacing V^ρ by the appropriate matrix i.e. $\frac{\partial R}{\partial r_\rho}$ the lemma can readily be generalised to arbitrary arrays. Thus, the results above hold for arbitrary arrays.

5. EXAMPLE APPLICATIONS

The narrow-band outputs from an array of equi-spaced receivers in a number of different noise fields were simulated. The noise field consisted of isotropic noise, uncorrelated receiver self noise and a plane wave signal. The frequency was such that d, the separation between adjacent receivers was either $\lambda/2$ or $\lambda/4$. Expressions for the bias of the various estimators and are given in tables 1 and 2. In

TABLE 1. Bias Factors - 400 samples

$d/\lambda = \frac{1}{4}$, $\theta = 25°$, SNR* = -10dB

K	M	Kullback	Max.Lik.
4	16	1.165	.998
	12	1.232	.997
	8	1.403	.994
	6	1.66	.993
2	16	1.03	-
	12	1.045	-
	8	1.06	-
	4	1.14	-

*Signal to uncorrelated noise ratio

Table 2. Bias factors - 400 samples

$d/\lambda = \frac{1}{2}$, $\theta = 90°$, SNR = 6dB

K	M	Kullback	Max.Lik.
4	16	1.172	1
	12	1.236	.999
	8	1.397	.999
2	16	1.029	-
	12	1.04	-
	8	1.061	-
	4	1.138	-

all cases the biases depended only on M and K i.e. the number of integrations and the number of receivers. Bearing in mind the close relationship of these estimators to the optimum beamforming using S this result is not unexpected, [12].

In figures 1 and 2 the snapshot angular spectra for are shown for the least squares, Kullback, and maximum likelihood estimates using a sample matrix inverse beamformer. Also plotted are the angular spectrum 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. Apart from the different bias factors the probability based estimators all gave roughly equivalent results. This was also observed for many other simulations. Significant differences between the least squares and the probability based estimators is apparent, particularly regarding biases in the peaks corresponding to signal arrival directions and spurious side lobes.

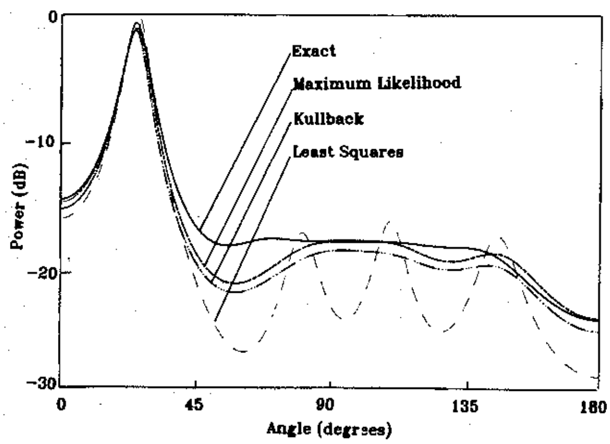

Fig 1: Angular spectra (d/λ = 0.25, SNR = 6 dB, K = 8, M = 32)

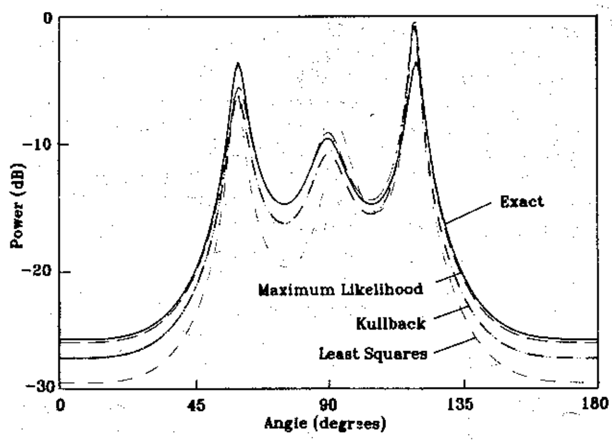

Fig 2: Angular spectra (d/λ = 0.5, SNRs = 16,10,16, K = 4, M = 8)

To illustrate further the difference between the least squares and the Kullback estimator, 350 different estimates of the angular spectrum were generated, using the least squares and Kullback estimators of the structured covariance. The mean plus or minus one standard deviation are plotted in figure 3. As can be seen the Kullback estimator has a significantly reduced variance with respect to the least squares one.

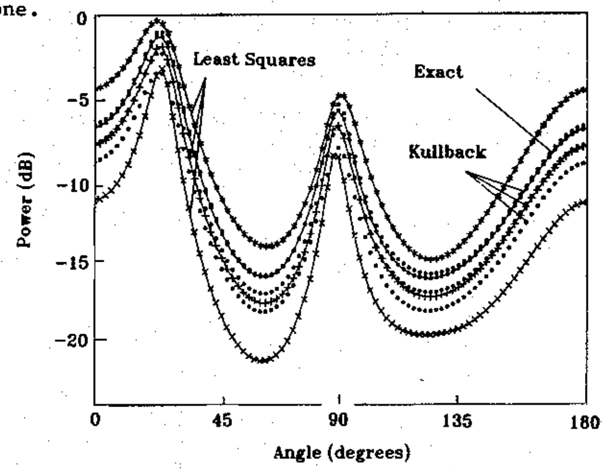

Fig 3: Angular spectra (mean ± sigma) (d/λ = 0.5, SNRs = 10,5,K = 4, M= 16)

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 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.

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 (1) M is large (typically $M > 4K$) and (2) the condition number of the observed *unstructured estimate* is *not too* large, were confirmed. In most simulations the information theoretic estimators gave almost equivalent results which confirmed further results in [9]. *When the number of* integrations was reduced (i.e. typically $M < 2K$) 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] Monzingo, R.A. and Miller, T.W., "Introduction to Adaptive Arrays". *Wiley-Interscience*, 1980.

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

[3] Gray, D.A., "Theoretical and Experimental Comparisons of the Formulation of Optimum Processors in Element, Beam and Eigen Spaces" in "Adaptive Methods *in Underwater Acoustics*" Ed. Urban, H.G., D. Reidel.

[4] Johnson, D.H., "The Application of Spectral Estimation Methods to Bearing Estimation Problems", Proc. IEEE, 70, 1982, pp.1018-1028.

[5] Byrne, C.L. and Steele, A.K., "Stable Nonlinear Methods for Sensor Array Processing", IEEE Journal of Oceanic Engineering, 10, 1985, pp.255-259.

[6] Iohvidov, I.S., "Hankel and Toeplitz Matrices and Forms", Birkhauser, 1982.

[7] Burg, J.P., Luenberger, D.G. and Wenger, D.L., "Estimation of Structured Covariances", Proc. IEEE, 70, 1982, pp.963-974.

[8] A, N.Q., "On the Uniqueness of the Maximum-Likelihood Estimate of Structured Covariance Matrices", IEEE Trans. Acoustics Speech and Signal Processing, ASSP-32, 1984, pp.1249-1251.

[9] D.A. Gray, B.D.O. Anderson, P.K. Sim, "Estimation of Structured Covariances with *Application to Array Beamforming*" to appear in Circuits, Systems and Signal Processing.

[10] Shore, J.E., "On a Relation Between Maximum Likelihood *Classification and* Minimum Relative Entropy Classification, IEEE Trans. Info. Theory, IT-30, 1984, pp.851-854.

[11] Kullback, S., "Information Theory and Statistics" Dover, New York, 1968.

[12] Capon, J. and Goodman, J.R., "Probability Distributions for Estimators of the Frequency-Wavenumber Spectrum", Proc. IEEE, 1970, 58, pp.1785-1786.