Proceedings of the Institute of Acoustics

A "TRUE" MAXIMUM LIKELIHOOD METHOD FOR ESTIMATING FREQUENCY WAVENUMBER SPECTRA

Arthur B. Baggeroer

Massachusetts Institute of Technology, Cambridge, MA USA 02139

1. INTRODUCTION

Adaptive array processing is used in many fields where high spatial resolution and sidelobe control are needed. These include geophysical exploration, seismic arrays for earthquakes, physical oceanography, sonars, radio astronomy, microwave radar for air traffic control, and phased array communication satellites. In all these fields the number of sensors in the arrays is increasing rapidly as the instrumentation technology becomes less expensive and easier to implement. Arrays with over 1000 sensors are now being deployed or considered for a number of advanced systems. In addition, the array geometries are becoming more complicated because of structural constraints and/or availability of sites, so irregular geometries with nonuniform interelement spacing must be incorporated.

2. ADAPTIVE ARRAYS, SNAPSHOTS AND SAMPLE COVARIANCES

The array processing adapts to the signal environment by using "snapshots" of the ambient field to characterize its spatial structure usually in terms of the NxN cross spectral covariance matrix, $\mathbf{S}_{ij}(f)$ of the N sensor signals. This estimate is the sample covariance matrix $\hat{\mathbf{S}}(f)$ formed by averaging the outer product of the L snapshots, $\mathbf{X}^l(f), l = 1, L$, or

$$\hat{\mathbf{S}}(f) = \frac{1}{L}\sum_{l=1}^{L} X^l(f)X^l(f)^H \tag{1}$$

where the superscript H denotes hermetian. When the snapshots are Gaussian random vectors, the elements of the sample covariance matrix have a complex Wishart joint probability density of dimension N and degrees of freedom L, which is a matrix generalization of the χ^2 density. [1]

In many applications the number of snapshots is often limited by the short term stationarity of the ambient field relative to the requirements of the adaptive processing. For example, the signals may be transient such as earthquake seismology or the result of an active signal such as in radar. The statistical character of the sample covariance introduces several important issues relevant to adaptive array processing when the number of sensors is large and the number of snapshots small.

- For an arbitrary sensor covariance matrix there are $N(N+1)/2$ separate elements to be estimated with the only constraint that the matrix be positive semidefinite. The relationship among the elements can be further constrained if homogeneity is applied; however, this is not easy to introduce. [2].

- The rank of the sample covariance is $min(L, N)$. Many adaptive array algorithms require the inverse of the sample covariance matrix. When $L < N$ the rank is not full and the matrix is singular and not invertible. This has lead to *adhoc* fixes such as diagonal loading (adding a scaled identity matrix to the diagonal), eigenvalue threshholding (performing a singular value decomposition and putting a lower bound on the eigenvalues), or subspace decomposition (using an SVD to transform to a subspace with finite eigenvalues). While all of these are *adhoc*, they do work and are used extensively.

- Even when $L \geq N$ and the rank of the sample covariance is full, the number of snapshots required for accurate processing is large. For example, Brennan and Reed [3] demonstrated that $L > 3N$ was required to obtain reliable predictions of the receiver operating characteristics (ROC's) in the high performance regime,*i.e.* low false alarm, high detection probabilities.

- Virtually all adaptive arrays are used with signals which propagate within some medium, yet few if any of the adaptive array algorithms published to date take advantage of the physical constaints which the propagation medium imposes upon the covariance matrix to reduce the number of snapshots required for characterizing it. Within the set of positive semidefinite matrices the subset which is "close" to a matrix satisfying a propagation constraint is much more restricted (an example of such a constraint might be restrictint is to be the transform of a positive function over a support corresponding to propagating wavenumbers).

The essential points summarizing these items are that (1) the sample covariance often has insufficient rank or accuracy for adaptive algorithms with a large number of sensors and (2) the physical constraints imposed by the wave equation upon the covariance have not been exploited.

In most applications where constraints have been imposed, they are in the form of a set of discrete signals corresponding to a set of finite sources. This leads to a separable covariance matrix. The literature on this problem is quite extensive, and each field has its own set algorithms for approaching the problem of a finite number of directional signals against a background of spatially white (uncorrelated among sensors) noise. Many recent techniques involve eigenvector methods often in the context of MUSIC based approaches. *e.g.* [4], [5] In a large number of applications, especially geophysical ones, the signals field is spatially continuous and not well modelled by a discrete noise field. (Even discrete sources become continuous when the array apertures are large and approaching the coherence limits introduced by scattering within the propagation medium.)

Imposing *a priori* physical constraints on the structure of the covariance matrix for the general problem of discrete and continuous components has been attempted in a few limited cases for the problem of directional wavenumber spectrum estimation of sea surfaces in physical oceanography [6]. In this work a norm on the error matrix of the sample covariance and the transformation of an estimated wavenumber spectrum is minimized over the spectrum subject to constraints of the wavenumber support and *a priori* information, *e.g.* a bias for isotropicity. The minmization leads to an iterative algorithm which is verified against simulated and field data. While the algorithm worked, the number of snapshots needed for a specified confidence level was not discussed.

3. A "TRUE" MAXIMUM LIKELIHOOD METHOD FOR ESTIMATING FREQUENCY WAVENUMBER SPECTRA

The physical constraints of the propagation medium should reduce the number of snaphsots required for adaptive array processing since they provide *a priori* restrictions. Most importantly, it very desirable to be able to use adaptive processing when the number of snapshots is less than the number of sensors, which is often the situation with large arrays. Snyder and Miller [7] found this to be the case for stationary time series when they imposed a Toeplitz constraint in conjunction with an EM (estimate-measure) approach to spectrum estimation. Similar concepts can be applied to a variety of array processing problems if one imposes the constraints of the wave equation on the structure of the sensor covariance matrix. For problems in a homogeneous medium this can be done by requiring that the matrix have the form [1]

$$S_{ij}(f) = \sigma_n^2(f)\delta_{ij} + \int_{\Omega(\mathbf{k})} d\mathbf{k} P(f,\mathbf{k}) E_i(f,\mathbf{k}) E_j^H(f,\mathbf{k})$$

where

- $\sigma_n(f)$ is a "white noise" level associated with sensor noise and/or a sensitivity parameter;
- $P(f,\mathbf{k})$ is the directional wave spectrum of the ambient field;
- $\Omega(f,\mathbf{k})$ is the support of the wavenumber field at frequency f imposed by the wave equation;
- $E_i(f,\mathbf{k})$ is the sensor response to a signal in the wave field at frequency f and wavenumber $\mathbf{k}$.

Note that the physical constraints are imposed directly in terms of both the allowable wavenumber region and the sensor response. For example, (i) $\Omega(f,\mathbf{k})$ can be specified to include both propagating (real wavenumber) and evanescent (complex wavenumber) waves, (ii) $E_i(f,\mathbf{k})$ can accommodate omnidirectional sensors as well as sensors with directional sensitivity and vector sensors such as three component accelerometers and polarized instruments, and (iii) multimode propogation can be accommodated by specifying $\Omega(f,\mathbf{k})$ to include regions of $\mathbf{k}$ space corresponding to different propagation modes.

This physical model can be used to formulate a "true" maximum likelihood estimate of the directional wave spectrum, $P(f,\mathbf{k})$ and the white noise level σ_n^2. We assume that each "snapshot" is a Gaussian random vector drawn from an ensemble with the above sensor covariance matrix. The maximum likelihood estimate is given by

$$\arg \max_{P(f,\mathbf{k}),\sigma_n^2} \prod_{l=1}^{L} \frac{1}{(\pi \mathbf{S}(f))^N} e^{-\mathbf{X}^l(f)^H \mathbf{S}^{-1}(f) \mathbf{X}^l(f)}$$

This can be manipulated to the form

$$\arg \min_{P(f,\mathbf{k}),\sigma_n^2} L[Tr(\hat{\mathbf{S}}(f)\mathbf{S}(f)) - \ln \det[\mathbf{S}]]$$

[1]The important structure in the field is that it be represented as a superposition of uncorrelated sources. For homogeneous fields, a wavenumber spectrum provides this. More generally, one can consider a superposition of uncorrelated sources in a space whose propagation to the receivers are represented in terms of the Green's function of the medium such as is done for matched field processing. [8]

This is essentially the same result given by Burg and Luenberger [9]; however, they did not impose the propagation constraints. There are several important aspects of this maximum likelihood formulation.

- The sample covariance matrix, $\hat{\mathbf{S}}(f)$, is not inverted, so one does not need to resort to the various *adhoc* strategies to do so.
- The function has a nonlinear dependence upon the arguments for the minimization, so a closed form solution is not generally available.
- While a closed form solution is not available, the first and second variations can be derived using matrix calculus which suggests a numerical algorithm for the minimization.
- There are no constraints upon the array, so irregular geometries with nouniform interelement spacings can be used.

The variations have a remarkably simple form which have a very intuitive interpretation. For the first variation on $P(\mathbf{k})$ one obtains [2]

$$\Delta(\mathbf{k}) = [\frac{\frac{1}{L}\sum_{l=1}^{L}|\mathbf{W}_{mv}^{H}(\mathbf{k})\mathbf{X}^{l}|}{P_{mv}(\mathbf{k})} - 1]P_{mv}^{-1}(\mathbf{k})\delta P(\mathbf{k})$$

where

- $\mathbf{W}_{mv}(\mathbf{k}) = \mathbf{S}^{-1}\mathbf{E}(\mathbf{k})/\mathbf{E}(\mathbf{k})^{H}\mathbf{S}^{-1}\mathbf{E}(\mathbf{k})$ is the minimum variance beamformer for a field at wavenumber $\mathbf{k}$ with covariance $\mathbf{S}$;
- $P_{mv}(\mathbf{k}) = [\mathbf{E}(\mathbf{k})^{H}\mathbf{S}^{-1}\mathbf{E}(\mathbf{k})]^{-1}$ is the minimum variance output power (often referred to as the maximum likelihood method (MLM), or Capon, estimate.[11]

Since $P(\mathbf{k}) \geq 0$, the conditions for a stationary point must incorporate this constraint. This leads to either $\Delta(\mathbf{k})$ or $P(\mathbf{k}) = 0$ for all $\mathbf{k}$. There is a similar expression for the variation with respect to σ_n^2. The second variation can also be specified analytically so that one can use one of the several methods for multiparameter optimization [10] This was used to derive an update algorithm based upon discretizing the wavenumber function and using Newton's method for the iteration.

The intuitive interpretation of $\Delta(\mathbf{k})$ follows. $\mathbf{W}_{mv}^{H}(\mathbf{k})\mathbf{X}^{l}$ represents the output of the minimum variance filter at wavenumber $\mathbf{k}$ when snapshot $\mathbf{X}^{l}$ is the input. As such it passes all components at $\mathbf{k}$ and minimizes the all others. The sum is an average over all snaphsots. This average is compared to what is predicted by the model power output $P_{mv}(\mathbf{k})$ in order to determine the direction of the increase for $\delta P(\mathbf{k})$. The stationarity condition is applying what one would intuitively expect of an algorithm - making a beamformer measurement, comparing it to a model, adjusting the model, recalculating the beamformer.

The approach derived has considerable similarity to that of Snyder and Miller [7] using an estimate-measure formulation; however, there are several important differences: i) the formulation is for spatial signals and in corporates all the constraints of the wave equation, ii) the role of the white, or sensor, noise is included explicitly [3], iii) the iteration uses variational expressions so a

[2] The notation for the frequency dependence has been suppressed for simplicity.

[3] Spatial white noise cannot be represented by establishing a floor in the wave number spectrum since its support is usually finite in the $\mathbf{k}$ domain

number of update algorithms can be employed.

4. AN EXAMPLE

An algorithm implementing the "true" maximum likelihood formulation was implemented on a PC (IBM-AT class machine) to examine its convergence properties and compare it to existing algorithms. A field consisting of sensor noise, 3D isotropic (black body) noise[4] and three directional signals was synthesized. A receiving array consisting of a 30 element, circular array with $3/2\lambda$ radius was used. An algorithm based upon Newton's method was implemented to iterate to a stationary point for the "true" maximum likelihood estimate. Figures 1 and 2 show the directional spectrum estimates generated using the minimum variance, or Capon, estimator, and the proposed algorithm, respectively. [5] There are several improvements in the estimate of the directional spectrum using the new algorithm.

- It detects one of the directional signals whereas the minimum variance estimate does not.
- 3D isotropic noise measured on a 2D surface of the circular array has frequency wavenumber function given by $c/[1-(|\mathbf{k}|\lambda/2\pi)^2]^{1/2}$ which should diverge at $|\mathbf{k}| = 2\pi/\lambda$. The minimum variance algorithm does not indicate even a ridge for this divergence whereas the new algorithm has a ring in this region suggesting the spectral peak.
- The maximum likelihood algorithm decays much more rapidly to zero in the region outside the propagating region, *i.e.* $|\mathbf{k}| > 2\pi/\lambda$ which is the correct value for the wavenumber function in this region.

While the algorithm has been tested on a very limited number of cases, the work to date suggests that it does converge and give superior estimates for a finite number of snapshots and irregular array geometries.

5. SUMMARY

The "true" maximum likelihood method incorporates a finite number of snaphots, arbitrary array geometries and *a priori* physical constraints so it provides a consistent framework for estimating frequency wavenumber spectra. It is an implicit method since it estabishes a set of equations for the maximum as well as the local gradient about it. These conditions have intuitive interpretations in terms of the minimum variance fromula; however, they cannot be solved for explicitly nor can the global optimality be established. Nevertheless, an algorithm has been devised which apprears to converge relatively quickly and has modest computational requirements. Moreover, the resolution is superior to the minimum variance algorithm, the only method now capable of dealing with arrays of arbitrary geometries. There is the attractive possibility of using the minimum variance estimate to initialize the maximum likelihood algorithm.

While some preliminary analysis and implementation has been done on the "true" maximum likelihood algorithm, there are several important issues which remain for future investigation.

[4] Black body noise has a spatial covariance $S(\delta\mathbf{x}) = sinc(2\pi|\delta\mathbf{x}|/\lambda)$. It is not spatially white.

[5] The wavenumber spectra are plotted *vs* $\mathbf{k}\lambda/2\pi$, or normalized slowness. All propagating signals are therefore within the unit circle.

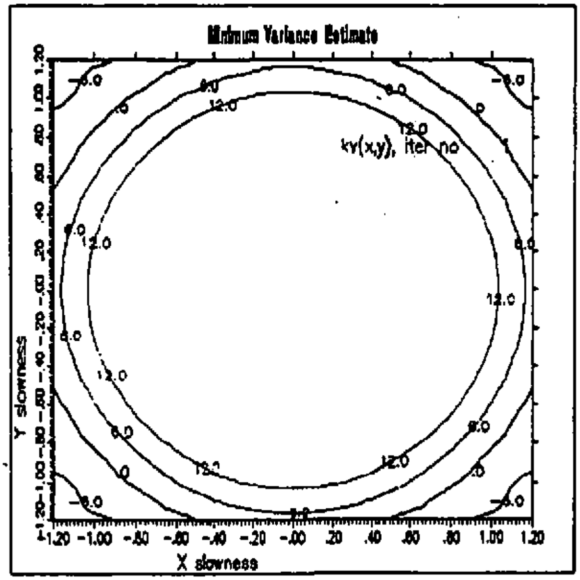

Figure 1: Directional Spectrum Estimate Generated With Minimum Variance Estimator

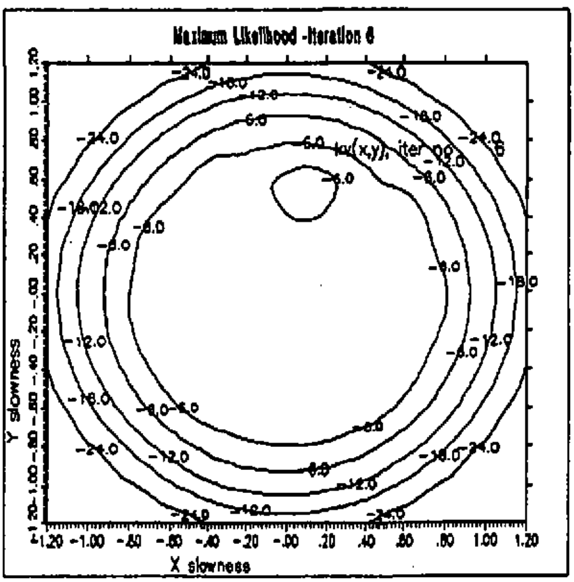

Figure 2: Directional Spectrum Estimate Generated With Proposed Algorithm (after 6 iterations)

REFERENCES

- The number of snapshots, L, required for a given array dimension, N, needs to be examined for the situation when $L \leq N$, *i.e.* the rank dificient case which often exists for large arrays. It is clear from empirical experience that arrays can work in this regime using *adhoc* methods to fill out or constrain the rank of the sample covariance, but there is no appropriate theory for this.
- A numerical implementation of the proposed algorithm requires that the wavenumber function, $P(f, \mathbf{k})$ be discretized. In principle this discretization can be arbitrarily fine which implies a very large number of degrees of freedom. In the preliminary analysis already done, the response remains stable in regions where the wavenumber function is smooth, but tends to become superdirective about directional signals. We have attempted to mitigate this by regularizing, or smoothing, the wavenumber response since the directional signals are impulsive in wavenumber space. Snyder has suggested that this is necessary in the EM approach to spectrum estimation [12]. The discretization methodology *versus* the natural resolution of the array, the regularization, and the number of snapshots all need to resolved.
- There are a number of applications involving transient data where "short term" directional spectra are desired. For example, Capon's original paper was concerned with earthquake data which are transient. This is often called velocity analysis in the geophysical literature.[13], [14], [15],[16] In these applications the wavenumber support can often be limited by *a priori* information. The proposed algorithm can incorporate this whereas others cannot. The application to transient data and short term directional spectra is an important application which needs to be explored.
- The proposed algorithm can be extended to inhomogeneous field applications. These are commonly called matched field methods and have evolved from sonar and geophysics, but are now finding application in a number of other areas where the inhomogeneity of the field is an important constraint. The extension of the analyses to inhomogeneous field applications is one of the important research topics.

6. ACKNOWLEDGEMENT

This work was supported in part by the Ocean Acoustics and the Arctic Acoustics Programs of the Office of Naval Research and in part by the Green Foundation while the author was on a sabbatical leave at the Scripps Institute of Oceanography.

References

[1] T.W. Anderson, "*An Introduction to Multivariate Statistical Analysis.*", 2nd Ed, John Wiley, 1984.

[2] S. Lang, J. McClellan, "Multidimensional MEM Spectral Estimation", *IEEE Trans of ASSP*, ASSP-30, No. 6, Dec 1982, pp 880-887.

[3] L. Reed, J. Mallett, L. Brennan, "Rapid Convergence Rate in Adaptive Arrays", *IEEE Trans on Aerospace and Electronic Systems*, AES-10, No. 6, Nov 1974, pp 853-863.

[4] R. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation", *IEEE Trans on Antennas and Propagation*, AP-34, No. 3, March 1986, pp 276-280.

REFERENCES

[5] M. Wax, T. Kailath, "Optimum Localization of Multiple Sources by Passive Arrays", *IEEE Trans of ASSP*, ASSP-31, No. 5, Oct 1983, pp 1210-1218.

[6] R. Long, "The Statistical Evaluation of Directional Spectrum Estimates Derived from Pitch/Roll Bouy Data", *J. Phys Oceanography*, Vol. 10, pp 944-952.

[7] M. Miller, D. Snyder, " The Role of Likelihood and Entropy in Incomplete-Data Problems: Applicatins to Estimating Point-Process Intensities and Toeplitz Constrained Covariances", *Proc of the IEEE*, Vol. 75, No. 7, July 1987, pp 892-907.

[8] A.B. Baggeroer, W.A. Kuperman, H. Schmidt, "Matched Field Processing; Source localization in correlation noise as an optimum parameter estimation problem", *JASA*, vol. 83, no. 2, February 1988, pp 571-587.

[9] J. Burg, D. Luenberger, D. Wegner, "Estimation of Structured Covariance Matricies" *Proc of the IEEE*, Vol. 70, No. 9, Sept 1982, pp 963-972.

[10] W.H. Press, "*Numerical Recipies in C, The Art of Scientific Computing*", Cambridge University Press, Cambridge, 1988.

[11] J. Capon, "High Resolution Frequency-Wavenumber Spectrum Analysis", *Proc of the IEEE*, vol. 57, 1969, pp 1408-1418.

[12] D. Snyder, personal communications.

[13] M. Taner, F. Koehler, "Velocity Spectra - Digital computer derivation and applications of velocity functions", *Geophysics*, Vol. 10, 1969, pp 859-881.

[14] Jean Pierre Coidier, "*Velocities and Reflection Seismology*" D. Reidel Publishing, Dordrecht, Holland, 1983.

[15] K. Hsu, A. Baggeroer, "Applicaiton of the maximum-likelihood method (MLM) for sonic velocity logging", *Geophysics*, Vol. 51, No. 3, March 1986, pp 780-787.

[16] T. Parks, C. Morris, J. Ingram, "Velocity Estimation from Short-Time Temporal and Spatial Frequency Estimation", *Proc of ICASSP*, Paris, France, 1982, pp 399-402.