ARRAY PROCESSING IN NON-GAUSSIAN NOISE WITH THE EM ALGORITHM

Richard J. Kozick

Bucknell University
Electrical Eng. Dept.
Lewisburg, PA 17837
kozick@bucknell.edu

Brian M. Sadler

Army Research Laboratory
Inf. Sci. and Tech. Directorate
Adelphi, MD 20783
bsadler@arl.mil

Rick S. Blum

Lehigh University
Elec. Eng. & CS Dept.
Bethlehem, PA 18015
rblum@eecs.lehigh.edu

ABSTRACT

A central problem in sensor array processing is the localization of multiple sources and the reception of the signals emitted by those sources. Many approaches have been studied for this problem when the additive noise in the sensor array data is modeled with a Gaussian distribution. However, the schemes designed for Gaussian noise typically perform very poorly when the noise is non-Gaussian. An algorithm is presented in this paper for array processing in non-Gaussian noise. The algorithm is based on modeling the noise with a Gaussian mixture distribution. The expectation-maximization (EM) algorithm is then used to derive an iterative processing structure that estimates the source locations, estimates the source waveforms, and adapts the processing to match the characteristics of the noise. Simulation examples are presented to illustrate the performance of the algorithm.

1. INTRODUCTION

The general problem in array signal processing is to obtain information about remote sources by processing the data measured from an array of sensors. For example, when applied in a multiple-user communication system, the objective is often to estimate the user locations and the individual user signals. This problem has been studied extensively when the additive noise that corrupts the measured data is modeled with a Gaussian distribution [1]. Many approaches have been considered for this case, including linear beamforming, high-resolution methods based on subspace processing, and maximum likelihood (ML) solutions. However, the techniques designed for Gaussian noise environments typically perform very poorly when the noise is non-Gaussian and "impulsive" in nature [2].

An approach to array processing in non-Gaussian noise is presented in this paper that is based on modeling the probability density function (pdf) of the additive noise with a finite mixture of Gaussian pdfs. An iterative algorithm for estimating the signal and noise parameters is then derived using the expectation-maximization (EM) algorithm.

We have recently applied a similar approach to develop an adaptive spatial diversity receiver for communication channels with fading and impulsive noise [3], [4]. Our experience with the spatial diversity application as well as the array processing application considered in this paper indicates some interesting features of the approach. First, the complexity of each iteration of the EM algorithm is low, being comparable to the complexity of a processor designed for Gaussian noise. Convergence is typically rapid (within 5 iterations), so the increase in computation is modest. Second, the processing adapts to the noise characteristics, with very good performance exhibited for a wide range of non-Gaussian as well as Gaussian pdfs. Third, the method performs well with small sample sizes.

Other approaches to array processing in non-Gaussian noise include references [5]-[8]. Distinctive features of the EM-based algorithm presented in this paper include the automatic adaptation of the processing to the observed noise characteristics, and that the processing structure is similar to that used with Gaussian noise, with modifications to mitigate the effects of impulsive noise.

The paper is organized as follows. Section 2 presents the model for the sensor array data. Section 3 presents the EM algorithm for estimating the source locations, the source waveforms, and the noise characteristics. Simulation examples are presented in Section 4, and Section 5 contains concluding remarks.

2. DATA MODEL

Let us model the complex envelope of narrowband, discrete-time observations at a sensor array with N elements as

$$\begin{aligned} \mathbf{x}(t) &= \sum_{k=1}^{K} \mathbf{a}(\theta_k) s_k(t) + \mathbf{w}(t) \\ &= \mathbf{A}\mathbf{s}(t) + \mathbf{w}(t), \quad t = 1, \ldots, T, \end{aligned} \tag{1}$$

where $\mathbf{x}(t) = [x_1(t), \ldots, x_N(t)]^T$, $\mathbf{a}(\theta) = [a_1(\theta), \ldots, a_N(\theta)]^T$ is the array response to a unit-amplitude source in direction θ, K is the number of sources, $\mathbf{\Theta} = [\theta_1, \ldots, \theta_K]$ are the source directions, $\mathbf{A}(\mathbf{\Theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K)]$ is an $N \times K$ matrix, $\mathbf{s}(t) = [s_1(t), \ldots, s_K(t)]^T$ is the vector of source complex amplitudes at time t, and $\mathbf{w}(t)$ is the additive noise vector. (The superscripts T, $*$, and H denote the transpose, complex conjugate, and conjugate-transpose operations, respectively.) We shall model the noise $\mathbf{w}(t)$ as independent and identically distributed (iid) in space and time, with marginal pdf given by the L-term Gaussian mixture

$$f_W(w) = \sum_{l=1}^{L} \frac{\lambda_l}{2\pi\sigma_l^2} \exp\left(-\frac{|w|^2}{2\sigma_l^2}\right). \qquad (2)$$

The objective is to estimate the source directions $\theta_1, \ldots, \theta_K$ and the deterministic source waveforms $\mathrm{s}(1), \ldots, \mathrm{s}(T)$. The number of sources K is assumed to be known.

In the pdf (2), λ_l represents the probability that W is chosen from the l^{th} term in the mixture pdf, with $\sum_{l=1}^{L} \lambda_l = 1$. For the case of $L = 2$ terms, a typical model for impulsive noise has $\sigma_2^2 \gg \sigma_1^2$ with $\lambda_2 < \lambda_1$, so that large noise samples with variance σ_2^2 occur with frequency λ_2 in a background of Gaussian noise with variance σ_1^2. Three justifications for this noise model are (1) the set of Gaussian mixture distributions includes an approximation to Middleton's canonical class A model [9], (2) Fan's theorem [10] indicates that Gaussian mixture distributions can approximate a large class of pdfs, and (3) the Gaussian mixture distribution naturally includes the Gaussian thermal noise that is present in essentially all electronic systems. We will assume that the number of terms L in the mixture pdf (2) has been determined prior to applying the methods developed in this paper. Our experience indicates that choosing $L \leq 4$ provides an accurate model in a variety of noise environments.

3. EM ALGORITHM

This section begins with a brief summary of the EM algorithm. General references for the EM algorithm include [11]-[15]. Then the EM algorithm for array processing in non-Gaussian noise is presented in Section 3.2.

3.1. General EM algorithm

The first step in applying the EM algorithm is the specification of a set of "complete data" $\mathbf{X}_c$ and "incomplete data" $\mathbf{X}$ for the problem. The pdfs for $\mathbf{X}_c$ and $\mathbf{X}$ are characterized by a common set of parameters $\mathbf{\Phi}$. The complete data $\mathbf{X}_c$ is not available, but it is chosen in such a way so that if it were available, then the maximum-likelihood (ML) estimate of $\mathbf{\Phi}$ would be easy to find. That is, if $h_c(\mathbf{X}_c|\mathbf{\Phi})$ is the pdf of the complete data, then it is straightforward to find $\mathbf{\Phi}$ that maximizes the likelihood $h_c(\mathbf{x}_c|\mathbf{\Phi})$ for a given set of complete data $\mathbf{X}_c = \mathbf{x}_c$. Conversely, the incomplete data $\mathbf{X}$ *is* available, but the ML estimate of $\mathbf{\Phi}$ based on $\mathbf{X}$ is difficult to find directly. The EM algorithm addresses this situation and provides an *iterative procedure* for ML estimation of $\mathbf{\Phi}$ based on the incomplete data $\mathbf{X}$.

For the application of array processing based on the model (1), the set of parameters is $\mathbf{\Phi} = \{\mathbf{\Theta}, \mathrm{s}(1), \ldots, \mathrm{s}(T), \lambda_1, \ldots, \lambda_L, \sigma_1^2, \ldots, \sigma_L^2\}$. The complete data $\mathbf{X}_c$ is defined to include a "label" for each observation that identifies which term $\{1, \ldots, L\}$ in the mixture pdf (2) produced the additive noise sample in the observation. Thus the complete data $\mathbf{X}_c$ is corrupted by additive *Gaussian* noise with variance that changes from sample to sample. Of course, the noise samples are not labeled in the incomplete data $\mathbf{X}$ that is available, hence the role of the EM algorithm.

In addition, the complete data $\mathbf{X}_c$ is also defined to facilitate the estimation of parameters due to multiple sources. The approach is similar to that used in [16] and [17] in the derivation of EM algorithms based on a Gaussian noise model. The complete data $\mathbf{X}_c$ is defined as observations of the K individual source waves in noise, i.e., $\mathbf{x}^{(k)}(t) = \mathbf{a}(\theta_k)s_k(t) + \mathbf{w}^{(k)}(t)$, $k = 1, \ldots, K$. The incomplete data $\mathbf{X}$ are the observations in (1), which are related to $\mathbf{X}_c$ by $\mathbf{x}(t) = \sum_{k=1}^{K} \mathbf{x}^{(k)}(t)$. This formulation of complete data $\mathbf{X}_c$ is motivated by the fact that the ML estimates of source location θ_k and waveform $s_k(t)$ are well-known for a *single* source.

The EM algorithm is an iterative procedure that is defined as follows. The EM algorithm computes an updated estimate for $\mathbf{\Phi}$ based on a previous estimate $\mathbf{\Phi}'$ and the incomplete data $\mathbf{X}$ by iterating the following two steps:

E-step: $Q(\mathbf{\Phi}|\mathbf{\Phi}') = E\{\log h_c(\mathbf{X}_c|\mathbf{\Phi}) \mid \mathbf{X},\ \mathbf{\Phi} = \mathbf{\Phi}'\}$

M-step: $\mathbf{\Phi} = \operatorname{argmax}\ Q(\mathbf{\Phi}|\mathbf{\Phi}')$

The E-step averages over the unavailable parts of the complete data $\mathbf{X}_c$, given the available data $\mathbf{X}$ and the previous parameter estimates $\mathbf{\Phi}'$. The sequence of estimates produced by iterating the EM algorithm have increasing likelihood [11], so they converge to a (local) maximum of the incomplete data likelihood function.

We have omitted derivations of the EM algorithms presented in this paper. The derivations involve performing the E- and M-steps listed above, with modifications similar to the SAGE algorithm [15] for simplified processing and improved convergence.

3.2. EM algorithm for array processing

The EM algorithm to update from previous estimates $\mathbf{\Phi}'$ to new estimates $\mathbf{\Phi}$ based on $\mathbf{x}(1), \ldots, \mathbf{x}(T)$ begins with the definitions

$$p_l'(n,t) = \frac{\lambda_l'}{\sigma_l'^2} \exp\left(-\frac{\left|x_n(t) - \sum_{k=1}^{K} a_n(\theta_k')s_k(t)'\right|^2}{2\sigma_l'^2}\right)$$

$$g_l'(n,t) = \frac{p_l'(n,t)}{\sum_{q=1}^{L} p_q'(n,t)}, \qquad \begin{array}{l} l = 1, \ldots, L \\ n = 1, \ldots, N \\ t = 1, \ldots, T. \end{array} \qquad (3)$$

Diagonal matrices are defined as

$$\mathbf{G}_l'(\mathbf{x}(t)) = \operatorname{diag}\{g_l'(1,t), \ldots, g_l'(N,t)\} \qquad (4)$$

$$\mathbf{G}'(\mathbf{x}(t)) = \sum_{l=1}^{L} \frac{1}{\sigma_l'^2} \mathbf{G}_l'(\mathbf{x}(t)). \qquad (5)$$

Source location estimates $\theta_1, \ldots, \theta_K$ are updated as

$$\widehat{\mathbf{x}^{(k)}}(t)' = \mathbf{a}(\theta_k')s_k(t)' + \frac{1}{K}\left[\mathbf{x}(t) - \mathbf{A}(\mathbf{\Theta}')\mathrm{s}(t)'\right] \qquad (6)$$

$$\theta_k = \underset{\theta}{\operatorname{argmax}} \sum_{t=1}^{T} \frac{\left|\mathbf{a}(\theta)^H \mathbf{G}'(\mathbf{x}(t))\widehat{\mathbf{x}^{(k)}}(t)'\right|^2}{\mathbf{a}(\theta)^H \mathbf{G}'(\mathbf{x}(t))\mathbf{a}(\theta)}. \qquad (7)$$

The remaining parameter estimates are updated as

$$
\begin{aligned}
\mathrm{s}(t) &= \left(\mathbf{A}(\boldsymbol{\Theta})^H \mathbf{G}'(\mathbf{x}(t))\mathbf{A}(\boldsymbol{\Theta})\right)^{-1} \\
&\quad \cdot \left(\mathbf{A}(\boldsymbol{\Theta})^H \mathbf{G}'(\mathbf{x}(t))\right)\mathbf{x}(t) \qquad (8) \\
\lambda_l &= \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T} g_l{}'(n,t) \qquad (9) \\
\hat{\mathbf{w}}(t) &= \mathbf{x}(t) - \mathbf{A}(\boldsymbol{\Theta})\mathrm{s}(t) \qquad (10) \\
\sigma_l^2 &= \frac{\sum_{t=1}^{T} \hat{\mathbf{w}}(t)^H \mathbf{G}_l'(\mathbf{x}(t))\hat{\mathbf{w}}(t)}{2NT\lambda_l}. \qquad (11)
\end{aligned}
$$

The EM algorithm (3)-(11) is valid for $K \geq 1$ sources. Interpretations and simulation examples are provided for the case of a single source in the following section.

4. DISCUSSION AND SIMULATION

Let us consider the case of a single source $K = 1$ and compare with the ML processor for Gaussian noise. With $K = 1$, step (6) is trivial with $\widehat{\mathbf{x}^{(k)}}(t)' = \mathbf{x}(t)$. The ML solution for Gaussian noise chooses θ_1 to maximize the spatial power spectrum

$$P_{\text{Gaussian}}(\theta) = \frac{\mathbf{a}(\theta)^H \hat{\mathbf{R}}_{xx}\mathbf{a}(\theta)}{\|\mathbf{a}(\theta)\|^2} \qquad (12)$$

where $\hat{\mathbf{R}}_{xx} = \sum_{t=1}^{T} \mathbf{x}(t)\mathbf{x}(t)^H$ is the estimated array correlation matrix. The signal waveform is then estimated by the linear beamformer

$$s_1(t) = \frac{\mathbf{a}(\theta_1)^H \mathbf{x}(t)}{\|\mathbf{a}(\theta_1)\|_2^2}, \quad t = 1, \ldots, T. \qquad (13)$$

Comparing (12) and (13) for Gaussian noise processing with (7) and (8) in the EM algorithm, we can identify the EM estimate of the "spatial spectrum" in (7) as

$$P_{\text{EM}}(\theta) = \mathbf{a}(\theta)^H \left[\sum_{t=1}^{T} \frac{\mathbf{y}(t)\mathbf{y}(t)^H}{\mathbf{a}(\theta)^H \mathbf{G}'(\mathbf{x}(t))\,\mathbf{a}(\theta)}\right] \mathbf{a}(\theta) \qquad (14)$$

where $\mathbf{y}(t) = \mathbf{G}'(\mathbf{x}(t))\,\mathbf{x}(t)$. Similarly, the EM signal estimate (8) is a "nonlinear beamformer"

$$s_1(t) = \frac{\mathbf{a}(\theta_1)^H \mathbf{y}(t)}{\mathbf{a}(\theta)^H \mathbf{G}'(\mathbf{x}(t))\,\mathbf{a}(\theta)}, \quad t = 1, \ldots, T \qquad (15)$$

that suppresses the effects of impulsive noise.

Let us consider a numerical example of a uniform line array with $N = 10$ elements and half-wavelength spacing. Additive noise is generated with the Gaussian mixture pdf (2) with $L = 2$ terms and parameters $\lambda_1 = 0.95, \lambda_2 = 0.05, \sigma_1^2 = 1$, and $\sigma_2^2 = 1000$. The EM algorithm (3)-(11) is used with $L = 2$ terms in the noise model, thus allowing an exact match between the noise model and the actual noise pdf. The signal variance is 10, and one source ($K = 1$) is located at angle $\theta_1 = 10^o$ or $u_1 = \sin\theta_1 = 0.1736$. The EM algorithm is iterated 5 times. Figure 1a contains plots of the spatial power spectra with Gaussian processing (12) and EM processing (14) for a typical run using $T = 100$ time samples. Note that the impulsive noise distorts the Gaussian power spectrum and yields an inaccurate estimate of the source direction $\hat{u}_1 = 0.153$. In contrast, the EM power spectrum provides an accurate estimate of the source direction $\hat{u}_1 = 0.173$. Figure 1b shows the amplitude of the estimated signal waveform for the first 35 time samples based on (13) for Gaussian processing and (8) for EM processing. The effects of impulsive noise on the Gaussian estimate are evident, while the EM estimate follows the actual signal amplitude closely.

In order to test the algorithm with additive noise that is not exactly modeled by a Gaussian mixture pdf of the form (2), the simulation was repeated with complex-valued noise samples whose real and imaginary parts are described by a Cauchy pdf with scale parameter 1. The results were very similar to those presented above for Gaussian mixture noise. However, $L = 4$ terms were needed in the noise model (2) for good performance.

5. CONCLUDING REMARKS

An approach to array signal processing in non-Gaussian noise has been presented based on modeling the additive noise with a Gaussian mixture pdf and using the EM algorithm to estimate the signal and noise parameters. The approach is general and can be applied to other signal processing applications, such as digital communication over channels with intersymbol interference and non-Gaussian noise [18].

6. REFERENCES

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach", *IEEE Sig. Proc. Mag.*, vol. 13, no. 4, pp. 67-94, July 1996.

[2] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: a survey," *Proceedings IEEE*, vol. 73, pp. 433-481, March 1985.

[3] R.S. Blum, R.J. Kozick, and B.M. Sadler, "An adaptive spatial diversity receiver for non-Gaussian interference and noise," *First IEEE Workshop on Signal Processing Advances in Wireless Communication*, pp. 385-388, Paris, France, April 16-18, 1997.

[4] R.J. Kozick, R.S. Blum, and B.M. Sadler, "EM-Based Approaches to Adaptive Signal Detection on Fading Channels with Impulsive Noise," *Proc. of the 1997 Conf. on Inf. Sci. and Sys.*, Johns Hopkins University, pp. 112-117, March 19-21, 1997.

[5] Y. Yardimci, J.A. Cadzow, A.E. Cetin, "Robust direction of arrival estimation in non-Gaussian noise", to appear in *IEEE Trans. Signal Processing.*

[6] B. M. Sadler, G. B. Giannakis, and S. Shamsunder, "Noise subspace techniques in non-Gaussian noise using cumulants," *IEEE Trans. on Aero. and Elec. Sys.*, Vol. 31, No. 3, pp. 1009-1018, July 1995.

[7] D.D. Lee and R.L. Kashyap, "Robust maximum likelihood bearing estimation in contaminated Gaussian noise", *IEEE Trans. Sig. Proc.*, vol. 40, no. 8, pp. 1983-1986, Aug. 1992.

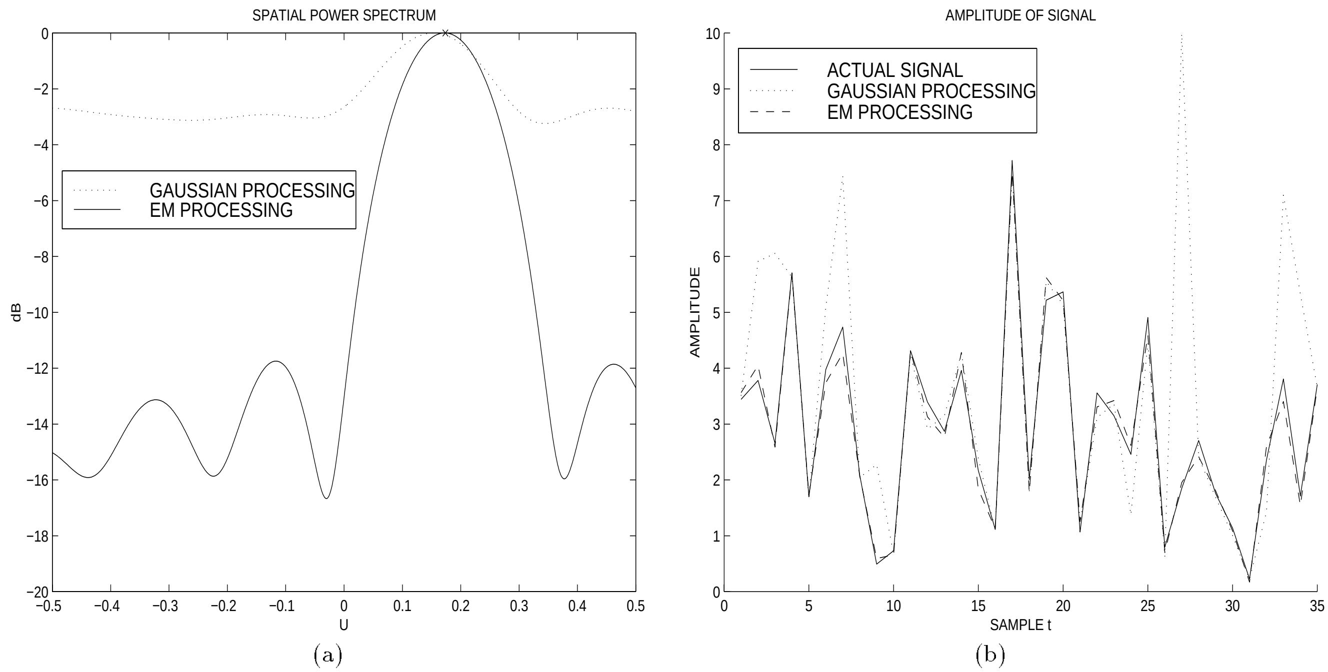

Figure 1: (a) Spatial power spectra versus source location $u = \sin\theta$ and (b) the amplitude of the complex signal waveform estimate using Gaussian processing and EM processing in impulsive noise. The actual source direction is $u_1 = 0.1736$.

[8] P. Tsakalides and C.L. Nikias, "Maximum likelihood localization of sources in noise modeled as a stable process", *IEEE Trans. Sig. Proc.*, pp. 2700-2713, Nov. 1995.

[9] A. Spaulding and D. Middleton, "Optimum reception in an Impulsive Interference Environment-Part I: Coherent detection," *IEEE Trans. on Comm.*, vol. COM-25, pp. 910-923, Sept. 1977.

[10] L. A. Liporace, "Maximum likelihood estimation for multivariate observations of markov sources," *IEEE Trans. Inf. Th.*, vol. IT-28, no. 5, pp. 729-734, Sept. 1982.

[11] A.P. Dempster, N.M. Laird, D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. of the Royal Statistical Soc.*, vol. 39, no. 1, pp. 1-38, 1977.

[12] R.A. Redner, H.F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26. pp. 195-239, April 1984.

[13] G.J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley, 1997.

[14] T.K. Moon, "The expectation-maximization algorithm", *IEEE Signal Processing Magazine*, pp. 47-60, Nov. 1996.

[15] J.A. Fessler and A.O. Hero, "Space-alternating generalized EM algorithm," *IEEE Trans. Signal Processing*, vol. 42, no. 10, pp. 2664-2677, Oct. 1994.

[16] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm", *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. 36, no. 4, pp. 477-489, April 1988.

[17] M.I. Miller and D.R. Fuhrmann, "Maximum-likelihood narrow-band direction finding and the EM algorithm", *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. 38, no. 9, pp. 1560-1577, Sept. 1990.

[18] R.J. Kozick, R.S. Blum, and B.M. Sadler, "Signal processing in non-Gaussian noise using mixture distributions and the EM algorithm", *31st Asilomar Conf. Sig., Sys., and Comp.*, Pacific Grove, CA, Nov. 2-5, 1997.