Maximum Likelihood Localization of Sources in Noise Modeled as a Cauchy Process*

Panagiotis Tsakalides and Chrysostomos L. Nikias
Signal and Image Processing Institute
Department of Electrical Engineering - Systems
University of Southern California
Los Angeles, CA 90089-2564
E-mail: tsakalid@sipi.usc.edu

Abstract

In this paper, a new robust beamformer, based on the Cauchy additive noise assumption, is introduced. The maximum likelihood approach is used for the bearing estimation of multiple sources from a set of snapshots when the interference is impulsive in nature. It is shown that the Cauchy receiver greatly outperforms the Gaussian receiver in a wide variety of non-Gaussian noise environments, and performs comparably to the Gaussian receiver when the additive noise is Gaussian. The Cramér-Rao bound on the estimation error variance is derived, and the robustness of the Cauchy beamformer in a wide range of impulsive interference environments is demonstrated via simulation experiments.

1 Introduction

This paper addresses the signal parameter estimation problem from sensor array data in the presence of impulsive interference. The simplest problem in this context is that of estimating the direction of arrival of incoming signals to an array of receivers. The source signals originate in the far field with arbitrary locations.

In the past, the problem has been studied extensively under the assumption of Gaussian distributed signals and/or noise, and a variety of methods for its solution have been proposed. The Maximum Likelihood (ML) method was one of the first to be investigated [1]. Under certain regularity conditions, the ML estimator is known to be asymptotically efficient, i.e., it achieves the Cramér-Rao bound on the estimation error variance.

Recently, special interest has been shown in relaxing some of the assumptions about the statistical nature of the noise in the bearing estimation problem. One such method is the bispectrum beamformer introduced by Forster and Nikias [2]. It was demonstrated that, for the case of spatially correlated Gaussian additive noise with unknown cross-spectral matrix (CSM), the bispectrum beamformer may provide asymptotically better bearing estimates than the stochastic ML method with known CSM.

Many times in the real world the Gaussian noise assumption can be inadequate, and systems designed under this assumption exhibit a significant performance degradation. There exist physical processes generating interferences containing **noise components that are impulsive in nature**. These processes can be natural, as well as man-made, and include underwater acoustic signals, lightning in the atmosphere, and transients in power lines and car ignitions. In modeling this type of signals the *stable distribution law* provides a very attractive theoretical tool. The stable distribution comes as a generalization of the Gaussian distribution including the Gaussian and Cauchy distributions as special cases.

This paper is devoted to the maximum likelihood estimation of multiple sources in the presence of interference which is modeled as a Cauchy process. The paper is organized as follows: In section 2, the direction of arrival (DOA) problem is formulated. In section 3, the bivariate isotropic stable distribution law is described and the complex Cauchy noise model is introduced. In section 4, the ML estimator is discussed and the Cramér-Rao bound is presented. Finally, simulation results are presented in section 5, and conclusions are drawn in section 6.

2 Problem Formulation

Consider an array of p sensors with arbitrary locations and arbitrary directional characteristics, that receive signals generated by q narrow-band sources with

*The work in this paper was supported by the Office of Naval Research under Contract N00014-92-J-1034

known center frequency ω and locations $\theta_1, \theta_2, \ldots, \theta_q$. Since the signals are narrow-band, the propagation delay across the array is much smaller than the reciprocal of the signal bandwidth, and it follows that, by using a complex envelop representation, the array output can be expressed as

$$\mathbf{x}(t) = \sum_{k=1}^{q} \mathbf{a}(\theta_k) s_k(t) + \mathbf{n}(t) \tag{1}$$

where

- $\mathbf{x}(t) = [x_1(t), \ldots, x_p(t)]^T$ is the vector of the signals received by the array sensors;
- $s_k(t)$ is the signal emitted by the kth source as received at the reference sensor 1 of the array;
- $\mathbf{a}(\theta_k) = [1, e^{-\jmath\omega\tau_2(\theta_k)}, \ldots, e^{-\jmath\omega\tau_p(\theta_k)}]^T$ is the steering vector of the array toward direction θ_k;
- $\tau_i(\theta_k)$ is the propagation delay between the first and the ith sensor for a waveform coming from direction θ_k;
- $\mathbf{n}(t) = [n_1(t), \ldots, n_p(t)]^T$ is the noise vector.

(1) can be expressed in a compact form as

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t) \tag{2}$$

where $\mathbf{A}(\boldsymbol{\theta})$ is the $p \times q$ matrix of the array steering vectors

$$\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_q)] \tag{3}$$

and $\mathbf{s}(t)$ is the $q \times 1$ vector of the signals

$$\mathbf{s}(t) = [s_1(t), \ldots, s_q(t)]^T. \tag{4}$$

Assuming that M snapshots are taken at time instants $t_1, \ldots, t_M$, the data can be expressed as

$$\mathbf{X} = \mathbf{A}(\boldsymbol{\theta})\mathbf{S} + \mathbf{N} \tag{5}$$

where $\mathbf{X}$ and $\mathbf{N}$ are the $p \times M$ matrices

$$\mathbf{X} = [\mathbf{x}(t_1), \ldots, \mathbf{x}(t_M)] \tag{6}$$

$$\mathbf{N} = [\mathbf{n}(t_1), \ldots, \mathbf{n}(t_M)] \tag{7}$$

and $\mathbf{S}$ is the $q \times M$ matrix

$$\mathbf{S} = [\mathbf{s}(t_1), \ldots, \mathbf{s}(t_M)]. \tag{8}$$

Our objective is to estimate the directions of arrival $\theta_1, \ldots, \theta_q$ of the sources from the M snapshots of the array $\mathbf{x}(t_1), \ldots, \mathbf{x}(t_M)$.

Toward this target we are going to make the following assumptions regarding the array, the signals, and the noise.

A.1 The number of signals is known and is smaller than the number of sensors, i.e., $q < p$.

A.2 The steering vectors are linearly independent among themselves.

A.3 The noise samples $n_i(t_j)$; $i = 1, \ldots, p$; $j = 1, \ldots, M$, come from a complex (bivariate) *isotropic stable distribution.*

A.4 The noise samples $n_i(t_j)$ are statistically independent from each other both along the array sensors, namely, along the index i, and along time, namely, along the index j.

Assumptions A.1 and A.2 guarantee the uniqueness of the solution. Assumption A.3 brings the new element in our analysis as we deviate from the conventional assumption that the noise in sensor arrays is a complex valued Gaussian process.

3 Bivariate Isotropic Stable Distributions

Multivariate stable distributions, like the multivariate Gaussian distribution, are characterized by the stability property and the generalized central limit theorem. However, they are much more difficult to describe because they form a nonparametric set [3]. An exception is the family of multidimensional isotropic stable distributions. Here, we concentrate in the two dimensional (bivariate) case which is appropriate for modeling complex interference and noise which are impulsive in nature.

The characteristic function of a bivariate isotropic α-stable distribution has the form

$$\varphi(\omega_1, \omega_2) = \exp(\jmath(\delta_1\omega_1 + \delta_2\omega_2) - \gamma|\boldsymbol{\omega}|^\alpha), \tag{9}$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2)$, and $|\boldsymbol{\omega}| = \sqrt{\omega_1^2 + \omega_2^2}$. Here, α is the *characteristic exponent* restricted to the values $0 < \alpha \leq 2$, and γ ($\gamma > 0$) is the *dispersion* of the distribution. The parameters δ_1, δ_2 are the location parameters. The distribution is isotropic with respect to the point (δ_1, δ_2), and the dispersion parameter γ determines the spread of the distribution around (δ_1, δ_2). The characteristic exponent α is the most important parameter of the *SαS* distribution and it determines its shape. The smaller the characteristic exponent α is, the heavier are the tails of the distribution. In the following we will assume that $(\delta_1, \delta_2) = (0, 0)$. The bivariate *isotropic Cauchy* and *Gaussian* distributions come as special cases for $\alpha = 1$ and $\alpha = 2$, respectively.

When $\alpha \neq 1$ or $\alpha \neq 2$, no closed form expressions exist for the density function of the bivariate stable random variable. By using the polar coordinate $r = |\boldsymbol{x}| = \sqrt{x_1^2 + x_2^2}$, the density function can be written as $f_{\alpha,\gamma}(x_1, x_2) = \chi_{\alpha,\gamma}(r)$, and is given in a power series expansion form [4]. The bivariate isotropic Cauchy can be written as

$$\chi_{1,\gamma}(r) = \frac{\gamma}{2\pi(r^2+\gamma^2)^{3/2}}. \qquad (10)$$

4 The Maximum Likelihood Estimator

In this section we develop the Maximum Likelihood (ML) estimator of the source locations in the presence of noise modeled as a **complex isotropic Cauchy process** with dispersion γ. In a similar approach as in [5] we do not regard the source signals as sample functions of random processes but rather we regard them as *unknown deterministic sequences*.

Under assumption A.4, it follows from (1) and (10) that the joint density function of the sampled data is given by

$$f(\mathbf{X}) = \prod_{t=1}^{M}\prod_{i=1}^{p}\chi_{1,\gamma}\left(\left|x_i(t) - \sum_{k=1}^{q} a_i(\theta_k)s_k(t)\right|\right) \qquad (11)$$

or

$$f(\mathbf{X}) = \prod_{t=1}^{M}\prod_{i=1}^{p}\frac{1}{2\pi}\frac{\gamma}{\left(\gamma^2 + |x_i(t) - \sum_{k=1}^{q} a_i(\theta_k)s_k(t)|^2\right)^{3/2}},$$

where $a_1(\theta_k) = 1$ and $a_i(\theta_k) = e^{-j\omega\tau_i(\theta_k)}$; $i = 2,\ldots,p$. Hence the log likelihood function $L(\mathbf{X};\gamma,\mathbf{S},\boldsymbol{\theta})$, ignoring constant terms, is expressed as:

$$L(\mathbf{X};\gamma,\mathbf{S},\boldsymbol{\theta}) = Mp\log(\gamma) - \frac{3}{2}\sum_{t=1}^{M}\sum_{i=1}^{p}\log\left(\gamma^2 + \left|x_i(t) - \sum_{k=1}^{q} a_i(\theta_k)s_k(t)\right|^2\right).$$

The ML estimator is obtained by maximizing $L(\mathbf{X};\gamma,\mathbf{S},\boldsymbol{\theta})$ with respect to γ, $\mathbf{S}$, and $\boldsymbol{\theta}$, i.e.,

$$\max_{\gamma,\mathbf{S},\boldsymbol{\theta}} L(\mathbf{X};\gamma,\mathbf{S},\boldsymbol{\theta}). \qquad (12)$$

To reduce the dimension of this optimization problem, we first fix γ and $\boldsymbol{\theta}$, and minimize $L(\mathbf{X};\gamma,\mathbf{S},\boldsymbol{\theta})$ with respect to the signal $\mathbf{S}$. For fixed t we take the derivative of $L(\mathbf{X};\gamma,\mathbf{S},\boldsymbol{\theta})$ with respect to $s_k(t)$:

$$\frac{\partial\mathcal{L}}{\partial s_k(t)} = -3\sum_{i=1}^{p}\frac{a_i(\theta)\left[x_i(t) - a_i(\theta_k)s_k(t)\right]^*}{\gamma^2 + |x_i(t) - a_i(\theta_k)s_k(t)|^2}. \qquad (13)$$

Unfortunately, no explicit solution of (13) is possible. In order to be able to obtain closed form expressions for the signals, we resort to the application of the pseudo-maximum likelihood (PML) estimation. PML estimation is an important method in applications where probability models abound for which the analytical derivation of the maximum likelihood estimate for all the parameters is virtually impossible. The problem formulated in [6] can be stated as follows:

Let $X_1,\ldots,X_n$ be i.i.d. random variables with probability distribution $f(X;\boldsymbol{\theta},\mathbf{S})$ indexed by two sets of parameters. Let $\hat{\mathbf{S}} = \hat{\mathbf{S}}(X_1,\ldots,X_n)$ be an estimate of $\mathbf{S}$ other than the maximum likelihood estimate, and let $\hat{\boldsymbol{\theta}}$ be the solution of the likelihood equation $\partial/\partial\boldsymbol{\theta}\ \log L(\mathbf{X};\boldsymbol{\theta},\hat{\mathbf{S}}) = 0$ which maximizes the likelihood. Then, $\hat{\boldsymbol{\theta}}$ is called a pseudo maximum likelihood estimate of $\boldsymbol{\theta}$, and under certain conditions it is consistent and asymptotically normal.

The PML estimator $\hat{\boldsymbol{\theta}}$ has good large sample properties when $\hat{\mathbf{S}}$ does. In general, the asymptotic analysis for $\hat{\boldsymbol{\theta}}$ will depend on the asymptotic characteristics of $\hat{\mathbf{S}}$.

Returning to the optimization problem described in (12), we observe that maximizing $L(\mathbf{X};\gamma,\mathbf{S},\boldsymbol{\theta})$ with respect to the signal $\mathbf{S}$ is equivalent to the following minimization problem:

$$\min_{\mathbf{S}}\mathcal{L}(\mathbf{X};\gamma,\mathbf{S},\boldsymbol{\theta}) = \min_{\mathbf{S}}\left\{\sum_{t=1}^{M}\sum_{i=1}^{p}\log(1 + \frac{|x_i(t) - \sum_{k=1}^{q} a_i(\theta_k)s_k(t)|^2}{\gamma^2})\right\}.$$

As we can see, (14) involves minimizing a double sum expression of logarithmic functions of the form $\log(1+z)$. In the unit disc $B_1(0) = \{z \in \mathcal{C} : |z| < 1\}$, the function $\log(1+z)$ can be expressed as an infinite series:

$$\log(1+z) = \sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}z^n = z - \frac{z^2}{2} + - \cdots; \quad |z| < 1 \qquad (14)$$

Hence for $|x_i(t) - \sum_{k=1}^{q} a_i(\theta_k)s_k(t)| < \gamma$ the functional $\mathcal{L}(\mathbf{X};\boldsymbol{\theta},\mathbf{S})$ can be written in the form

$$\mathcal{L}(\mathbf{X};\boldsymbol{\theta},\mathbf{S}) = \sum_{t=1}^{M}\sum_{i=1}^{p}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n\gamma^{2n}}|x_i(t) - \sum_{k=1}^{q} a_i(\theta_k)s_k(t)|^{2n}. \qquad (15)$$

A first order approximation of the above expression results in the following $\mathcal{L}^{(1)}(\mathbf{X};\boldsymbol{\theta},\mathbf{S})$ functional:

$$\mathcal{L}^{(1)}(\mathbf{X};\boldsymbol{\theta},\mathbf{S}) = \frac{1}{\gamma^2}\sum_{t=1}^{M}\sum_{i=1}^{p}|x_i(t) - \sum_{k=1}^{q} a_i(\theta_k)s_k(t)|^2 \qquad (16)$$

which, by using (2), can be written in a more compact form as:

$$\mathcal{L}^{(1)}(\mathbf{X};\boldsymbol{\theta},\mathbf{S}) = \frac{1}{\gamma^2}\sum_{t=1}^{M}|\mathbf{x}(t) - \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t)|^2. \qquad (17)$$

Hence the minimization of $\mathcal{L}^{(1)}(\mathbf{X};\boldsymbol{\theta},\mathbf{S})$ with respect to $\mathbf{S}$ is equivalent to the Least-Squares (LS) estimation of $\mathbf{S}$. This problem has a well known solution:

$$\hat{\mathbf{s}}(t) = (\mathbf{A}^H(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}))^{-1}\mathbf{A}^H(\boldsymbol{\theta})\mathbf{x}(t). \tag{18}$$

The dispersion γ can be estimated by using the measurements of the first sensor and the expression for the fractional lower order moments (FLOM) of the noise, where $E|X|^p$ is approximated by an average sum:

$$\hat{\gamma} = \frac{\left[\frac{1}{M}\sum_{t=1}^{M}|x_1(t) - \sum_{k=1}^{q}\hat{s}_k(t)|^p\right]^{\frac{1}{p}}}{[C_2(p,1)]^{\frac{1}{p}}}, \tag{19}$$

where $p < 1$, and $C_2(p,1)$ is given by:

$$C_2(p,1) = p2^p\frac{\Gamma(\frac{p}{2})\Gamma(-p)}{\Gamma(-\frac{p}{2})}. \tag{20}$$

By using the above estimates for the signal $\mathbf{S}$ and the noise dispersion γ, we obtain the following reduced optimization problem:

$$\max_{\boldsymbol{\theta}} L(\mathbf{X};\hat{\gamma},\hat{\mathbf{S}},\boldsymbol{\theta}) = \max_{\boldsymbol{\theta}}\{Mp\log(\hat{\gamma})$$
$$-\frac{3}{2}\sum_{t=1}^{M}\sum_{i=1}^{p}\log(\hat{\gamma}^2 + |x_i(t) - \sum_{k=1}^{q}a_i(\theta_k)\hat{s}_k(t)|^2)\}.$$

Now we can apply an iterative procedure based on the gradient descent principle in order to solve for $\boldsymbol{\theta}$. Concluding this section we point out, for one more time, the two main assumptions made in order to obtain a closed form expression for the signal estimate $\hat{\mathbf{S}}$:

B.1 Assumption $|x_i(t) - \sum_{k=1}^{q}a_i(\theta_k)s_k(t)| < \gamma$ (with probability $(\sqrt{2}-1)/\sqrt{2}$) enabled us to express the logarithmic function as an infinite series;

B.2 Assumption $|x_i(t) - \sum_{k=1}^{q}a_i(\theta_k)s_k(t)| \ll \gamma$ enabled us to take a first order approximation of the infinite series, and thus to obtain a least-squares closed form estimate for the signal.

Obviously, if the above assumptions are not satisfied the performance of the estimator will not be "optimal".

Finally, under the assumptions stated the CRB for $\boldsymbol{\theta}$ is given by [7]

$$CRB(\boldsymbol{\theta}) = \frac{5\gamma^2}{3}\left\{\sum_{t=1}^{M}\Re\{\mathbf{S}^*(t)\mathbf{D}^*\left[I - \mathbf{A}\left(\mathbf{A}^*\mathbf{A}\right)^{-1}\mathbf{A}^*\right]\mathbf{D}\mathbf{S}(t)\}\right\}^{-1},$$

where $\mathbf{S}(t) = diag\{s_1(t),\ldots,s_q(t)\}$, $\mathbf{D} = [\mathbf{d}(\theta_1),\ldots,\mathbf{d}(\theta_q)]$, and $\mathbf{d}(\theta_i) = \partial\mathbf{a}(\theta_i)/\partial\theta_i$.

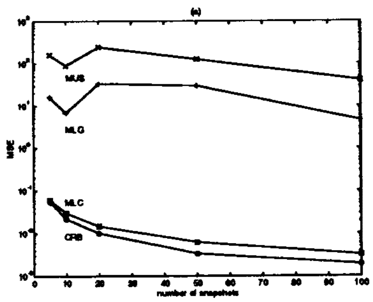

Figure 1: MSE of the estimates of DOA and CRB as a function of the number of snapshots M. Experiment #1. Exact signal knowledge.

5 Simulation Results

To demonstrate the performance of the ML estimator for the direction of arrival (DOA) problem we conducted two simulation experiments. We compare the ML estimator based on the Cauchy noise assumption (MLC) with the ML estimator based on the Gaussian noise assumption (MLG), and with the MUSIC estimator (MUS).

The array is linear with four sensors spaced half wavelength apart. The signal is a 16-level QAM impinging from 5^o. At first, the signal is assumed to be known at the receiver and the ML method is applied to estimate the direction of arrival. Also, the pseudo ML method is applied using a least squares estimate of the signal. The noise is assumed to follow the bivariate isotropic stable distribution. In every experiment we performed 100 Monte-Carlo runs and computed the mean and the mean square error (MSE) of the direction of arrival estimates.

Experiment #1

In the first experiment we study the influence of the number of snapshots, M, available to the performance of the algorithm. The noise follows the Cauchy distribution with dispersion γ (c.f. Eq. 10). Figure 1 shows the resulted MSE of the estimated DOA as a function of the number of snapshots when the signal is known. The improved performance of the MLC estimator even for a small number of snapshots is evident.

Figure 2 shows similar plots for the case of the pseudo ML estimators with a LS estimate for the signal. The MLC estimator gives again the least MSE. Comparing these curves with the analogous curves obtained assuming exact signal knowledge, we observe a larger MSE for the pseudo ML estimates, as expected.

Experiment #2

In the second experiment we test the robustness of the estimators when the characteristic exponent,

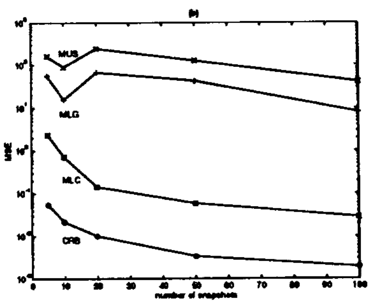

Figure 2: MSE of the estimates of DOA. Experiment #1. Least Squares estimate of the signal.

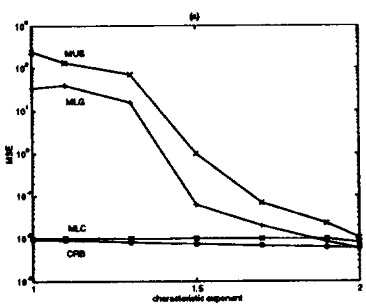

Figure 3: MSE of the estimates of DOA and CRB as a function of the characteristic exponent α. Experiment #2. Exact signal knowledge.

α, of the noise stable law is changing. In other words we test the receiver performance in different impulsive noise environments. Of course, by design, the MLG estimator is optimal for additive Gaussian noise, and the introduced MLC estimator is optimal for additive Cauchy noise. An important property of any receiver is to be able to perform reasonably well in a wide range of noise environments.

Figure 4 shows the resulting MSE of the estimated DOA as a function of the characteristic exponent α. As we can clearly see, for exact signal knowledge (Figure 3), the Cauchy receiver is practically insensitive to the changes of α. Note that, when $\alpha = 2$, i.e., for the Gaussian noise case, the *MLG* receiver has the least MSE as expected.

6 Concluding Remarks

We have presented a novel approach to the DOA estimation problem in the presence of impulsive interference. The method is based on the maximum likelihood estimation technique where the noise is modeled as a complex isotropic Cauchy process. The Cauchy beamformer has been shown to give better DOA estimates than the Gaussian beamformer in a wide range of impulsive noise environments. The technique inherits the computational complexity of the ML family of methods but avoids the eigendecomposition operations of the eigenvector-based methods, and appears to have superior performance for low SNR values and when the number of observation samples is small.

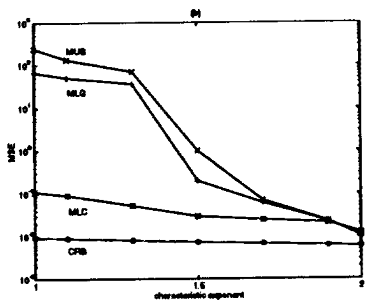

Figure 4: MSE of the estimates of DOA. Experiment #2. Least Squares estimate of the signal.

References

[1] R. O. Schweppe, "Sensor array data processing for multiple signal sources," *IEEE Trans. Inform. Theory*, vol. 14, pp. 294–305, 1968.

[2] P. Forster and C. L. Nikias, "Bearing estimation in the bispectrum domain," *IEEE Trans. Signal Processing*, vol. 39, pp. 1994–2006, 1991.

[3] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proc. IEEE*, vol. 81, pp. 986–1010, 1993.

[4] V. M. Zolotarev, "Integral transformations of distributions and estimates of parameters of multidimensional spherically symmetric stable laws," in *Contributions to Probability* (J. Gani and V. K. Rohatgi, eds.), pp. 283–305, New York: Academic Press, 1981.

[5] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 36, pp. 1553–1560, 1988.

[6] G. Gong and F. J. Samaniego, "Pseudo maximum likelihood estimation:Theory and applications," *Ann. Statist.*, vol. 9, pp. 861–869, 1981.

[7] P. Tsakalides and C. L. Nikias, "Maximum likelihood localization of sources in noise modeled as a stable process," *IEEE Trans. Signal Processing.* submitted on April 4, 1994, pp. 30.