Available online at www.sciencedirect.com
SCIENCE DIRECT
Signal Processing 86 (2006) 3061–3075
SIGNAL PROCESSING
www.elsevier.com/locate/sigpro

Maximum likelihood parameter estimation under impulsive conditions, a sub-Gaussian signal approach

Panayiotis G. Georgiou*, Chris Kyriakakis

University of Southern California, 3740 McClintock Av., EEB400, Los Angeles, CA 90089-2564, USA

Received 5 March 2004; received in revised form 21 June 2005; accepted 7 January 2006
Available online 23 February 2006

Abstract

In this paper we present an alternative to the Gaussian and Cauchy distributions for modeling stochastic signals. The proposed model has the same impulsiveness as the Cauchy density, but it is derived as a sub-Gaussian process, i.e., a variance mixture of Gaussian random variables.

We proceed to use the derived model in the problem of signal parameter estimation through the use of multisensor data. Both the data and noise are assumed to be stochastic. The main problem of interest is the estimation of the DOA and statistics of the signal. A maximum likelihood algorithm is presented for the solution of this problem, and a pseudo-maximum-likelihood separable solution approach is derived. Finally, simulations are presented to demonstrate the robustness of the proposed algorithm.

Keywords: Sub-Gaussian; α-Stable; Maximum likelihood; Array

1. Introduction

Statistical array processing based on second-order moments has been the focus of considerable research over the last decades. The focus on second-order statistics stemmed from the assumption that the Gaussian distribution is the most realistic model for a variety of noise signals. In recent years however, there has been a tremendous interest in the class of α-stable distributions, which are a generalization of the Gaussian distribution, but are able to model a wider range of phenomena and can be of a more impulsive nature. In fact, the Gaussian is the least impulsive α-stable distribution, while other widely known distributions of the α-stable class are the Cauchy and the Lévy.

In 1991, Cambanis et al. [1] gave a review of α-stable processes from a statistical point of view. Several other statisticians have provided valuable work in the theory of α-stable distributions. Cambanis, Weron, Zolotarev, Miller et al. have done extensive work on the properties of the α-stable distributions, and in the fields of linear filtering and spectral representation. A textbook of comprehensive coverage of the α-stable theory was written by Samorodnitsky and Taqqu [2] in 1994.

In 1993, Nikias and Shao [3] gave an introductory review of α-stable distributions from a statistical signal processing viewpoint that was followed by a book from the same authors [4] in 1995.

*Corresponding author. Tel.: +1 213 740 4654; fax: +1 213 740 4651.
E-mail addresses: georgiou@sipi.usc.edu (P.G. Georgiou), ckyriak@sipi.usc.edu (C. Kyriakakis).

0165-1684/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2006.01.007

α-stable distributions have been used to model diverse phenomena such as random fluctuations of gravitational fields, economic market indices [5], climate changes [6], and radar clutter [7]. More recently additional areas of application of the α-stable distributions have become apparent. For example, Crovella et al. have used the stable law to model data file sizes on the Web while Willinger et al. used it to model network traffic [5]. Furthermore there is evidence of the appropriateness of the α-stable distributions for modeling noises encountered in audio environments [8] and in designing time delay estimation techniques for localization of speech sources based on these models.

Tsakalides and Nikias [9,10] gave maximum likelihood (ML) and multiple signal classification (MUSIC) based localization algorithms for uncorrelated, impulsive signals. In this paper we will present a ML algorithm for signals that are dependent and impulsive in nature. We will proceed by first giving a brief introduction on α-stable distributions and sub-Gaussian signals. We will then present an overview of the array problem under investigation in Section 3, and the solution of the ML problem under Gaussian conditions in Section 3.1 as background material. In Section 3.2, we will derive the $\alpha = 1$ sub-Gaussian signal model and the ML algorithm under these signal conditions (Fig. 3), and in the next section we will formulate a separable solution for this ML problem. Finally, in Section 4 we will present simulations comparing the Gaussian based ML and the sub-Gaussian based ML algorithms.

2. Theory: random processes

2.1. α-stable distributions

The α-stable distribution, which can model phenomena of an impulsive nature, is a generalization of the Gaussian distribution and is appealing because of two main reasons: it satisfies the *stability property* and the *generalized central limit theorem* [4,2,11].

Unfortunately, there is no closed form expression for the probability density function of α-stable distributions, but the characteristic function $\varphi(x)$ is given by

$$\varphi(x) = \exp(\mathrm{i}\lambda x - \gamma|x|^{\alpha}[1 + \mathrm{i}\beta\,\mathrm{sign}(x)\omega(x,\alpha)]), \qquad (1)$$

where

$$\omega(x,\alpha) = \begin{cases} \tan\dfrac{\alpha\pi}{2} & \text{if } \alpha \neq 1, \\ \dfrac{2}{\pi}\log|x| & \text{if } \alpha = 1, \end{cases} \qquad (2)$$

$$\mathrm{sign}(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases} \qquad (3)$$

The parameters of the characteristic function are the *location parameter* ($-\infty < \lambda < \infty$), the *dispersion* parameter ($\gamma > 0$), the index of *symmetry* ($-1 \leqslant \beta \leqslant 1$) and the *characteristic exponent* ($0 < \alpha \leqslant 2$). The characteristic exponent controls the thickness of the tails of the density function. The tails are heavier, and thus the noise is more impulsive for lower values of α.

The case of $\alpha = 2$, $\beta = 0$ corresponds to the Gaussian distribution, while $\alpha = 1$, $\beta = 0$ corresponds to the Cauchy distribution. The density functions in these two cases are given by

$$f_{\alpha=2}(\gamma,\lambda;x) = \frac{1}{\sqrt{4\pi\gamma}}\exp\left\{-\frac{(x-\lambda)^2}{4\gamma}\right\}, \qquad (4)$$

$$f_{\alpha=1}(\gamma,\lambda;x) = \frac{\gamma}{\pi[\gamma^2 + (x-\lambda)^2]}. \qquad (5)$$

A closed form expression also exists for the case of the Lévy distribution (also referred to as a Pareto type 5), which has parameters $\beta = 1$ and $\alpha = 0.5$, and therefore it is completely skewed to the positive axis.

$$f(x) = \begin{cases} \dfrac{\gamma}{\sqrt{2\pi}}(x-\lambda)^{-3/2}\exp\left\{-\dfrac{\gamma^2}{2(x-\lambda)}\right\} & \text{if } x > \lambda, \\ 0 & \text{if } x \leqslant \lambda. \end{cases} \qquad (6)$$

The only other closed form expression for a stable distribution is the case obtained by symmetric reflection of the Lévy, i.e., with $\alpha = 0.5$ and $\beta = -1$. The density is given by $f_{\alpha=0.5,\beta=-1}(x) = f_{\mathrm{Levy}}(-x)$.

Refs. [2–5] treat the α-stable theory further.

2.2. Sub-Gaussian random variables

A sub-Gaussian random vector $\underline{\boldsymbol{X}}$ can be defined as a random vector with characteristic function of the form

$$\varphi(\underline{\boldsymbol{u}}) = \exp\left(-\tfrac{1}{2}\left[\underline{\boldsymbol{u}}^{\mathrm{T}}\ \underline{\underline{\mathbf{R}}}\ \underline{\boldsymbol{u}}\right]^{\alpha/2}\right), \qquad (7)$$

where $\underline{\underline{\mathbf{R}}}$ is a positive-definite matrix, and $0 < \alpha \leqslant 2$.

Sub-Gaussian processes are variance mixtures of Gaussian processes [12]. Specifically, $\underline{\boldsymbol{X}}(f)$ is sub-Gaussian with parameter α (denoted by α-SG($\underline{\underline{\mathbf{R}}}$)) if there exist $S(f)$, a positive $\alpha/2$-stable process with dispersion $\cos(\pi\alpha/4)^2$, and $\underline{\boldsymbol{Y}}(f)$, a multivariate Gaussian process independent of $S(f)$, and:

$$\underline{\boldsymbol{X}}(f) = S(f)^{1/2}\,\underline{\boldsymbol{Y}}(f). \tag{8}$$

A Gaussian random vector $\underline{\boldsymbol{Y}}(f)$ of size 4 is shown in Fig. 1, while a sub-Gaussian random vector $\underline{\boldsymbol{X}}(f)$ of impulsiveness $\alpha = 1$ and generated from the signals of Fig. 1 is shown in Fig. 2. The sub-Gaussian signals are very impulsive, and hence their signed $\log_e$ amplitude ($\mathrm{sign}(x)\log_e|x|$) is shown on the graphs. The sub-Gaussian random vector is obtained by (8) using a Lévy random variable (6) for $S(f)$ and the multivariate Gaussian random vector of Fig. 1. As observed from Fig. 2, the sub-Gaussian signals cannot be independent, even in the case of no correlation in the underlying Gaussian vector. For example, the generating top and bottom rows of Fig. 1 are uncorrelated Gaussian realizations, but the corresponding top and bottom rows of Fig. 2 demonstrate a clear dependence. This dependence can be observed clearer in the spikes of the detailed plots on the right-column subplots of Fig. 2. We also note from Fig. 2 that highly correlated signals can present significant differences after the multiplicative factor influences the sources.

Sub-Gaussian processes can be a good model for signals that are spatially uncorrelated but undergo some similar random environmental changes over time. For example, multiple random signal transmissions powered from the same physical power source experience the same impulsive energy fluctuations or echoes in an acoustic environment that are highly dependent on the original source[1] etc.

3. Maximum likelihood estimation

We continue our work by focusing on development of methods relating to the estimation of the parameters of a system where we assume multiple source signals received by a number of sensors (greater than the number of sources).

The transmitted signals in this case are assumed to be stochastic, and as such, the parameters of interest will be their statistics and directions-of-arrival (DOAs). The estimation process combines the measurements to obtain a vector $\underline{\boldsymbol{x}}(f)$, which best describes the observed data. The estimation process is in essence a mathematical algorithm that maximizes a certain cost function with respect to the observation vector $\underline{\boldsymbol{x}}(f)$. The cost function is obtained by assuming a certain statistical model for the signal and a certain optimization criterion. Common optimization criteria are the least-squares (LS), weighted LS, ML, as well as constrained optimization criteria. Despite the wide variety of optimization criteria, the optimal detector is characterized by a single result: the ML ratio test, which was also one of the first methods to be applied in the area of array signal processing [14].

The ML technique applied to the source localization problem usually makes two different assumptions for the signal waveforms, resulting in two different ML methods. According to the stochastic ML (SML), the signals are modeled as Gaussian random processes motivated by the central limit theorem, and result in closed form mathematical expressions. On the other hand, in the deterministic ML (DML) the signals are considered unknown but deterministic. In the latter case, estimates of the signals as well as the DOAs are desired, while in the former, the only parameters to be estimated are the signal statistics and DOAs. In this paper we deal exclusively with SML estimation.

We assume a scenario under which there are κ sources received by an array of ρ sensors. The transfer function each signal undergoes while traveling to the array can be modeled as an attenuation and a delay. The attenuation will be considered the same at all sensors under the assumption that the sources are in the far field of the array. Additionally we assume frequency independence. These transfer functions are

$$a_{r,k} = \mathrm{e}^{-\mathrm{i}\omega\tau_{r,k}}, \quad r = 1\ldots\rho \quad \text{and} \quad k = 1\ldots\kappa, \tag{9}$$

where $\tau_{r,k}$ is the delay of the signal (of source k) received at sensor r relative to the first sensor, and ω is the angular frequency. We assume the sources to be in the far field and hence, $\tau_{r,k} = \tau_r(\theta_k)$, and it is

[1] In a microphone array scenario, the reverberation can be better modeled as additive noise ($n(f)$ on Fig. 3), which is also dependent on the source signal but has no specific direction-of-arrival to the array. In contrast, the echoes have the same features as the original source but originate from different locations. The usefulness of this method in a reverberant environment is demonstrated in [13].

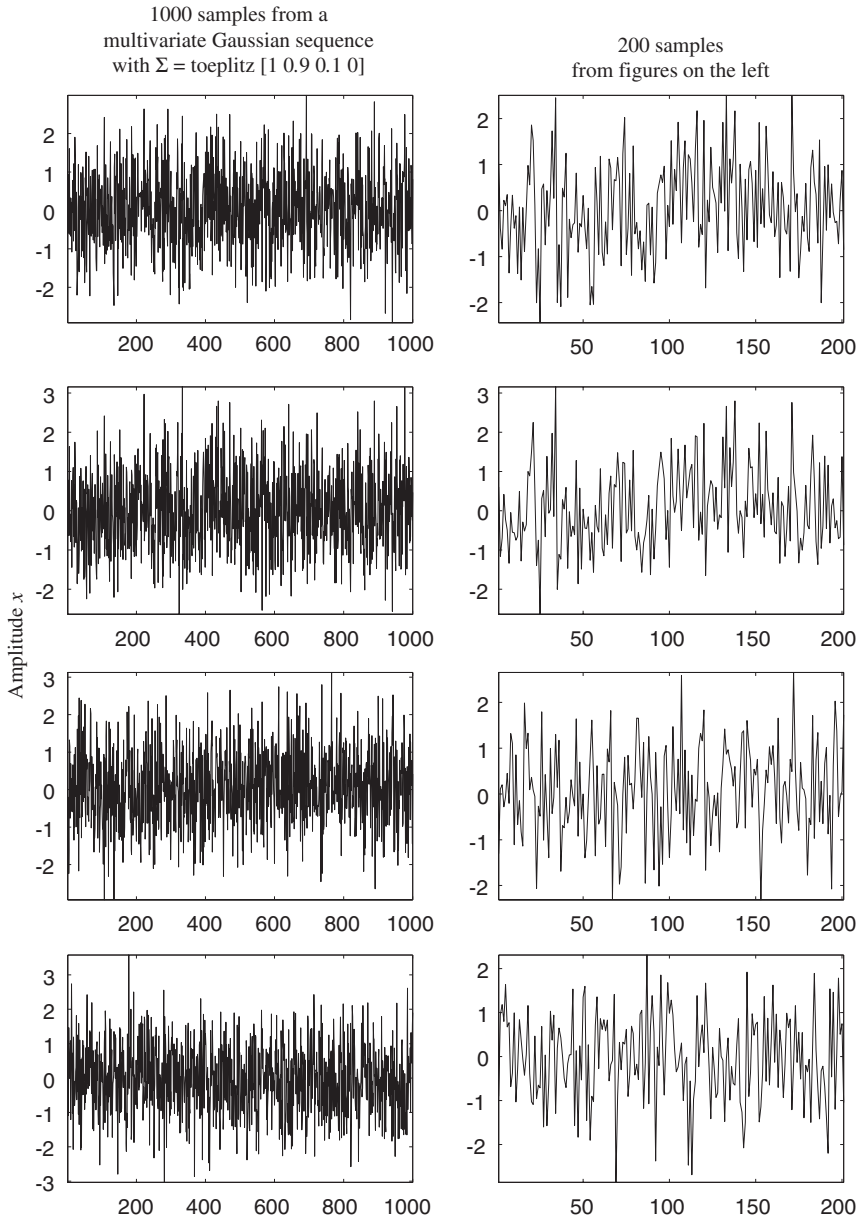

Fig. 1. Multivariate Gaussian random vector of size 4 $\Sigma = [1\ 0.9\ \ 0.1\ 0]$.

also clear that assuming a linear array:

$$\tau_{r,k} = r \cdot \tau_1(\theta_k). \tag{10}$$

We denote the vector of the medium transformations for source k by

$$\underline{\boldsymbol{a}}_k = [a_{1,k} \quad a_{2,k} \quad \ldots \quad a_{\rho,k} \quad]^{\mathrm{T}} = [1 \quad \mathrm{e}^{-\mathrm{i}\omega\tau_{1,k}} \quad \mathrm{e}^{-\mathrm{i}\omega\tau_{2,k}} \quad \ldots \quad \mathrm{e}^{-\mathrm{i}\omega\tau_{\rho-1,k}}]^{\mathrm{T}}. \tag{11}$$

The array's input at a single sensor r, for all frequencies f, is

$$x_r(f) = \sum_{k=1}^{\kappa} a_{r,k} \cdot s_k(f) + n_r(f) \tag{12}$$

and therefore, the array's input vector is

$$\underline{\boldsymbol{x}}(f) = \underline{\underline{\mathbf{A}}} \cdot \underline{\boldsymbol{s}}(f) + \underline{\boldsymbol{n}}(f), \tag{13}$$

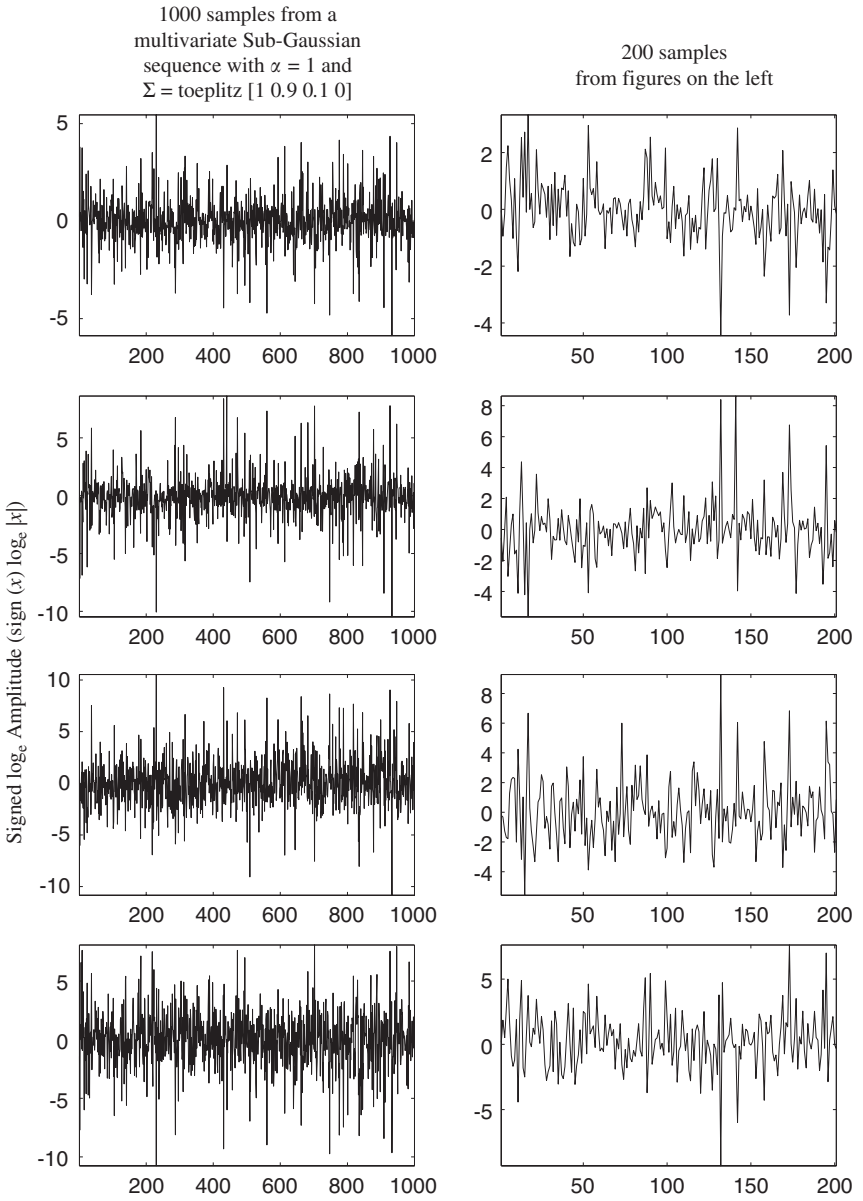

Fig. 2. Multivariate sub-Gaussian random vector of size 4 with $\alpha = 1$. The dependence despite $\sigma_{1,4} = 0$ can be seen from the spikes near samples 130 and 170 on the top right and bottom right subplots.

where

$$\underline{\underline{\mathbf{A}}} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,\kappa} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,\kappa} \\ \vdots & \vdots & \ddots & \\ a_{\rho,1} & a_{\rho,2} & & a_{\rho,\kappa} \end{bmatrix} \quad \text{and} \quad \underline{\boldsymbol{s}}(f) = \begin{bmatrix} s_1(f) \\ s_2(f) \\ \vdots \\ s_\kappa(f) \end{bmatrix}. \tag{14}$$

3.1. Gaussian signals

The most commonly used ML DOA estimator is the Gaussian ML derived either under the assumptions of a deterministic or a stochastic signal. Under the SML DOA estimator for a Gaussian signal, the signals are assumed to be jointly stationary with covariance matrix $\underline{\underline{\boldsymbol{\Sigma}}}_s = \mathbf{E}[\underline{\boldsymbol{s}}(f)\,\underline{\boldsymbol{s}}^\dagger(f)]$, and the noise to be uncorrelated white noise of variance σ^2.

The received signal statistics are

$$\underline{\underline{\mathbf{R}}} = \mathbf{E}[\underline{x}(f)\,\underline{x}^{\dagger}(f)] = \underline{\underline{\mathbf{A}}}\,\underline{\underline{\Sigma}}_s\,\underline{\underline{\mathbf{A}}}^{\dagger} + \sigma^2\,\underline{\underline{\mathbf{I}}}, \tag{15}$$

where $\underline{\underline{\mathbf{A}}}^{\dagger}$ denotes the Hermitian of $\underline{\underline{\mathbf{A}}}$. This leads to the ML estimator over all available frequencies ($f_1 \leqslant f \leqslant f_M$):

$$[\hat{\sigma}^2, \hat{\underline{\underline{\Sigma}}}, \hat{\underline{\theta}}] = \arg\min_{\hat{\sigma}^2, \hat{\underline{\underline{\Sigma}}}, \hat{\underline{\theta}}} \sum_{f=f_1}^{f_M} \{\log_e |\underline{\underline{\mathbf{R}}}| + \underline{x}^{\dagger}(f)\underline{\underline{\mathbf{R}}}^{-1}\,\underline{x}(f)\}. \tag{16}$$

This problem is further investigated in [15,16], and numerical methods are developed for the minimization of the ML function [17].

3.2. Sub-Gaussian signal model and ML estimation

The use of a sub-Gaussian signal of equal impulsiveness provides an alternative to Cauchy distributed signals [9], and allows for dependent sources to be modeled.

Consider the example of a transmission from a device and a reflection from a nearby obstacle. This creates the multipath effect of two highly dependent signals transmitted from two distinct points in space. At the same instance, consider a fast changing channel that affects the attenuation with which the signal will be received, and noise that travels through the same rapidly changing channel. The result is dependent signals and noise.

A second example can be seen in reverberant and echoic environments. Envision a sound source that is reflected off $\kappa - 1$ surfaces thus creating κ dependent signals. Existing studies [8] have demonstrated the impulsive nature of speech signals, and thus the result can be κ impulsive and dependent sources. Furthermore, reverberant noise is also dependent on the source signal. The κ direct and echoic signals will be directional, while the reverberation can be considered as non-directional noise.

To satisfy the above problem formulation, we consider stationary sub-Gaussian signals constructed with a distribution of impulsiveness $\alpha = 0.5$, which is completely skewed to the positive axis, together with a multivariate Gaussian density. There is one distribution with a closed form expression, the Lévy distribution, which satisfies exactly these properties. Fig. 3 gives a top level description of the problem and signals:

- A multivariate Gaussian signal is corrupted by multiplicative Lévy noise to the half power, i.e., $\underline{\boldsymbol{s}}(f) = u(f)^{1/2} \cdot \underline{\boldsymbol{v}}(f) = w(f) \cdot \underline{\boldsymbol{v}}(f)$.
- The resulting signal is transformed through a set of delays $\underline{\underline{\mathbf{A}}}$ to the receiving end of the array $\underline{\boldsymbol{x}}(f) = \underline{\underline{\mathbf{A}}} \cdot \underline{\boldsymbol{s}}(f)$.
- Jointly sub-Gaussian noise $\underline{\boldsymbol{n}}$ where $\underline{\boldsymbol{n}}(f) = w(f) \cdot \underline{\boldsymbol{\eta}}(f)$ can corrupt the signal before it is received by the array.

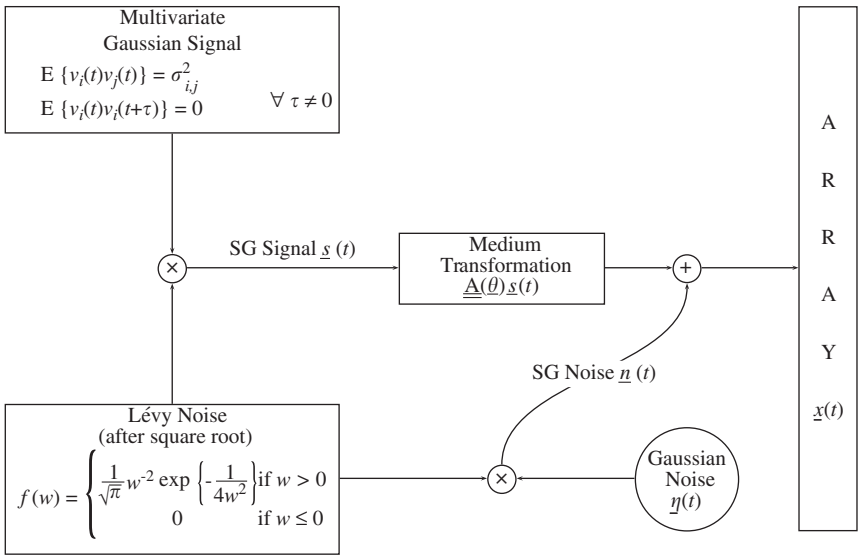

Fig. 3. Overview diagram explaining the various signals of the sub-Gaussian formulation. The multiplicative transformation w can be viewed as part of a rapidly changing multiplicative channel or as a characteristic of the signal source.

To return to the reverberant environment example, the sources of interest would be $s_1 \dots s_\kappa$ while the noise would be $n_1 \dots n_\rho$. However the signal observed by the sensors would be $\underline{\underline{\mathbf{A}}} \cdot \underline{\boldsymbol{s}}(f) + \underline{\boldsymbol{n}}(f)$. In contrast, in the first example, the signal of interest does not include the variable channel attenuation, and as such it is only $\underline{\boldsymbol{v}}$.

The Gaussian density at a single frequency f is

$$f(\underline{\boldsymbol{v}}) = \frac{1}{\pi^\kappa |\underline{\underline{\boldsymbol{\Sigma}}}_v|} \exp(-\underline{\boldsymbol{v}}(f)^\dagger \underline{\underline{\boldsymbol{\Sigma}}}_v^{-1} \underline{\boldsymbol{v}}(f)) \tag{17}$$

and the Lévy distribution [18,2] with $\gamma = 1/\sqrt{2}$ is given by

$$f(u) = \begin{cases} \dfrac{u^{-3/2} \mathrm{e}^{-1/4u}}{2\sqrt{\pi}} & \text{if } u > 0, \\ 0 & \text{if } u < 0. \end{cases} \tag{18}$$

From these it can be shown, as in Appendix A, that

$$f(\underline{\boldsymbol{s}}) = \frac{1}{2\pi^{(2\kappa+1)/2} |\underline{\underline{\boldsymbol{\Sigma}}}|} \frac{\Gamma((2\kappa+1)/2)}{[1/4 + \underline{\boldsymbol{s}}^\dagger \underline{\underline{\boldsymbol{\Sigma}}}^{-1} \underline{\boldsymbol{s}}]^{(2\kappa+1)/2}}, \tag{19}$$

where

$$\Gamma(z) = \int_0^\infty t^{z-1} \mathrm{e}^{-t} \, \mathrm{d}t. \tag{20}$$

Note that if the Gaussian random variable was one-dimensional and real, then

$$f(s) = \frac{\Gamma(1)}{2\pi\sqrt{2}\sigma} \cdot \left[1/4 + \frac{s^2}{2\sigma^2}\right]^{-1} = \frac{1}{2\sqrt{2}\pi\sigma} \cdot \left[1/4 + \frac{s^2}{2\sigma^2}\right]^{-1}. \tag{21}$$

A summary table of the distributions of interest is given in Table 1. Plots are shown in Fig. 4 for the one-dimensional case, and we can see that as expected, at $\sigma = \sqrt{2}$ the sub-Gaussian random process is equal in distribution to the Cauchy of $\gamma = 1$. Note that although this expression specifically addresses the $\alpha = 1$ sub-Gaussian case, as previous work has shown [8,10], algorithms developed under impulsive conditions are expected to be robust for all signals of lower impulsiveness, i.e., $\alpha \geqslant 1$.

Now the received signal $\underline{\boldsymbol{x}} = [x_1 \dots x_\rho]^{\mathrm{T}}$ is again of sub-Gaussian form. The *underlying* Gaussian statistics of $\underline{\boldsymbol{x}}(f)$ after introducing jointly sub-Gaussian noise (the noise is a sub-Gaussian process produced by the same Lévy sequence as shown in Fig. 3) are given by

$$\underline{\underline{\mathbf{R}}} = \underline{\underline{\mathbf{A}}} \, \underline{\underline{\boldsymbol{\Sigma}}}_v \, \underline{\underline{\mathbf{A}}}^\dagger + \sigma_\eta^2 \underline{\underline{\mathbf{I}}}_\rho. \tag{22}$$

Table 1
Distributions of interest: the sub-Gaussian density generated from the Lévy and Gaussian densities, in its one-dimensional and multi-dimensional form

Lévy	$f(u) = \begin{cases} \frac{u^{-3/2}\mathrm{e}^{-1/4u}}{2\sqrt{\pi}} & \text{if } u>0 \\ 0 & \text{if } u<0 \end{cases}$
Gaussian	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-x^2/2\sigma^2}$
1-D 1-sub-Gaussian	$f(x) = \frac{1}{2\sqrt{2\pi}\sigma} \cdot \left[\frac{x^2}{2\sigma^2} + \frac{1}{4}\right]^{-1}$
ρ-D Gaussian	$f(\underline{\mathbf{X}}) = \frac{1}{\pi^\rho \lvert\underline{\underline{\boldsymbol{\Sigma}}}\rvert} \exp(-\underline{\boldsymbol{x}}^\dagger \underline{\underline{\boldsymbol{\Sigma}}}^{-1} \underline{\boldsymbol{x}})$
ρ-D 1-sub-Gaussian	$f(\underline{\boldsymbol{X}}) = \frac{\Gamma((2\rho+1)/2)}{2\pi^{(2\rho+1)/2} \lvert\underline{\underline{\boldsymbol{\Sigma}}}\rvert} \left[\frac{1}{4} + \underline{\boldsymbol{x}}^\dagger \underline{\underline{\boldsymbol{\Sigma}}}^{-1} \underline{\boldsymbol{x}}\right]^{-(2\rho+1)/2}$

To simplify notation, the subscript v will be dropped and symbols $\underline{\underline{\boldsymbol{\Sigma}}}$ will be used subsequently to denote the underlying statistics of the transmitted signal $\underline{\boldsymbol{s}}(f)$ and $\underline{\underline{\mathbf{R}}}$ the underlying statistics of the received signal $\underline{\boldsymbol{x}}(f)$. Hat notations will be used to denote estimates, and the frequency dependence will be dropped.

Therefore, the ML estimator over all available frequencies is

$$[\hat{\underline{\underline{\boldsymbol{\Sigma}}}}, \hat{\underline{\boldsymbol{\theta}}}, \sigma_\eta] = \arg \max_{\hat{\underline{\underline{\boldsymbol{\Sigma}}}}, \hat{\underline{\boldsymbol{\theta}}}, \sigma_\eta} \prod_{f=f_1}^{f_M} \frac{1}{2\sqrt{\pi}\pi^\rho |\underline{\underline{\mathbf{R}}}|} \times [\underline{\boldsymbol{x}}^\dagger \underline{\underline{\mathbf{R}}}^{-1} \underline{\boldsymbol{x}} + 1/4]^{-(2\rho+1)/2}. \tag{23}$$

To simplify, take the $\log_\mathrm{e}$ and remove known terms:

$$[\hat{\underline{\underline{\boldsymbol{\Sigma}}}}, \hat{\underline{\boldsymbol{\theta}}}, \sigma_\eta] = \arg \min_{\hat{\underline{\underline{\boldsymbol{\Sigma}}}}, \hat{\underline{\boldsymbol{\theta}}}, \sigma_\eta} \sum_{f=f_1}^{f_M} \left\{ \log_\mathrm{e} |\underline{\underline{\mathbf{R}}}| + \left(\frac{2\rho+1}{2}\right) \times \log_\mathrm{e}[\underline{\boldsymbol{x}}^\dagger \underline{\underline{\mathbf{R}}}^{-1} \underline{\boldsymbol{x}} + 1/4] \right\}. \tag{24}$$

Comparing (24) with (16) we can see the diminished effect of outliers in the sub-Gaussian based ML estimator. The logarithmic operation limits the effect of large values, while the addition of $1/4$ ensures numerical stability of the algorithm in (24). In contrast the second order operation appearing in (16) amplifies the importance of any large amplitude outliers.

3.3. Maximum likelihood—a separable solution

Solution of the above optimization function requires minimizing for $\underline{\underline{\boldsymbol{\Sigma}}}$ (κ^2 parameters) and $\underline{\boldsymbol{\theta}}$ (κ parameters). Reducing the problem to a separable

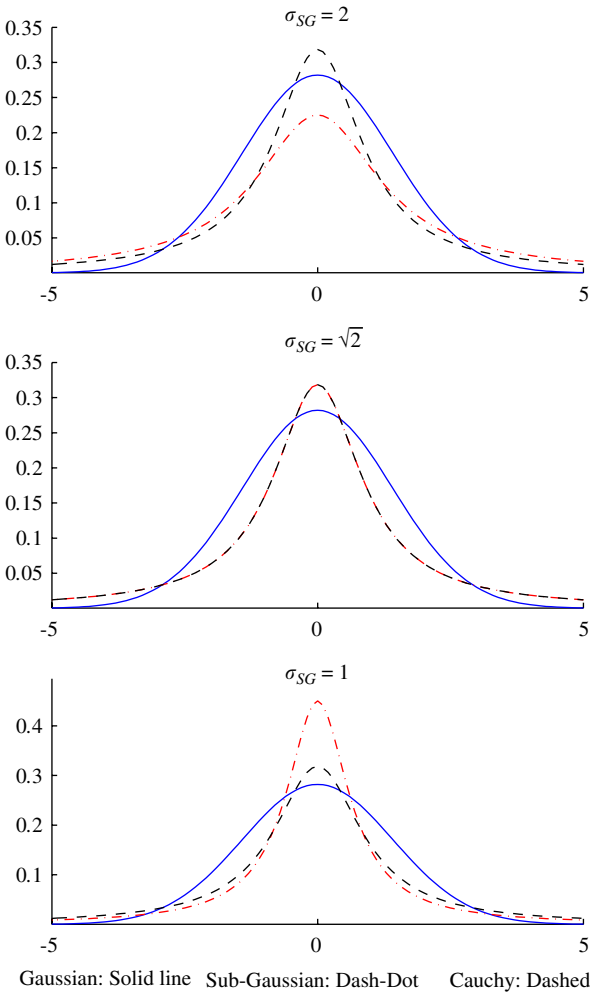

Fig. 4. Sub-Gaussian versus Cauchy and Gaussian distributions. When the dispersion of the underlying Gaussian of the sub-Gaussian process is equal to 1 (i.e., $\gamma_{SG} = 1 \Rightarrow \sigma_{SG} = \sqrt{2}$), the sub-Gaussian is equal in distribution to the normalized Cauchy.

solution, where the statistics and angle are estimated separately, is desirable to decrease the dimensionality of the problem. We will show here that the solution to the statistics estimator can be converted to a closed form expression (solvable by the numeric iteration method) and the DOA estimator is reduced to the search of a κ-dimensional space.

3.3.1. Estimating the statistics

We proceed to reach an alternative minimization function with regards to the statistics, assuming known DOAs. A procedure similar to the one of the Gaussian ML estimator as presented by Jaffer [16] is followed. Detailed derivations are presented in Appendix B.

$$\underline{\underline{\mathbf{\Sigma}}}_{\mathrm{ML}} = \frac{1}{M}\sum_{f=f_1}^{f_M}\left[\underline{\underline{\mathbf{A}}}^{-}\left(\frac{(\rho+0.5)\,\underline{\boldsymbol{x}}\,\underline{\boldsymbol{x}}^{\dagger}}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1}\,\underline{\boldsymbol{x}}\,\underline{\boldsymbol{x}}^{\dagger}\right]+1/4}-\sigma_{\eta}^{2}\right)\underline{\underline{\mathbf{A}}}^{-\dagger}\right], \tag{25}$$

where $\underline{\underline{\mathbf{A}}}^{-} = (\underline{\underline{\mathbf{A}}}^{\dagger}\,\underline{\underline{\mathbf{A}}})^{-1}\underline{\underline{\mathbf{A}}}^{\dagger}$, Tr is the trace operator, and

$$\underline{\underline{\mathbf{R}}}^{-1} = \frac{1}{\sigma_{\eta}^{2}}\{\underline{\mathbf{I}} - \underline{\mathbf{A}}(\underline{\mathbf{\Sigma}}\,\underline{\mathbf{A}}^{\dagger}\,\underline{\mathbf{A}} + \sigma_{\eta}^{2}\,\underline{\mathbf{I}})^{-1}\,\underline{\mathbf{\Sigma}}\,\underline{\mathbf{A}}^{\dagger}\}. \tag{26}$$

We can observe from (25) and (26) that the unknown $\underline{\underline{\mathbf{\Sigma}}}_{\mathrm{ML}}$ appears on both sides of the equation. However, the above equation can be easily and rapidly solved using the numerical iteration method.

In our experience, the iteration algorithm converges in very few cycles, and thus random initial conditions are used in the above equation in the following experiments. However, it is possible to start from a rough estimate of the underlying Gaussian statistics of a sub-Gaussian signal using fractional lower-order statistics as will be demonstrated later.

3.3.2. DOA estimation

Clearly, the above assumes that the DOA vector is known, and here we approach the localization part of the problem. Using a pseudo-ML approach, we can express the modified ML function as

$$\underline{\hat{\boldsymbol{\theta}}} = \arg\min_{\underline{\hat{\boldsymbol{\theta}}}}\sum_{f=f_1}^{f_M}\left\{\log_{\mathrm{e}}|\underline{\underline{\mathbf{R}}}| + \left(\frac{2\rho+1}{2}\right)\times\log_{\mathrm{e}}[\underline{\boldsymbol{x}}^{\dagger}\underline{\underline{\mathbf{R}}}^{-1}\,\underline{\boldsymbol{x}}+1/4]\right\}. \tag{27}$$

But since the statistics $|\underline{\underline{\mathbf{R}}}|$ are not a function of the data, the ML function is be reduced to

$$\underline{\hat{\boldsymbol{\theta}}} = \arg\min_{\underline{\hat{\boldsymbol{\theta}}}}\sum_{f=f_1}^{f_M}\{\log_{\mathrm{e}}[\underline{\boldsymbol{x}}^{\dagger}\underline{\underline{\mathbf{R}}}^{-1}\,\underline{\boldsymbol{x}}+1/4]\}, \tag{28}$$

where $\underline{\underline{\mathbf{R}}}$ can be substituted with any valid statistics (identity matrix for instance). A search algorithm such as the one described in [19] can be used to find the solution to the above equation. Note that iterating between (28) and (25) can achieve even higher accuracy results.

4. Simulations

4.1. DOA estimation

Several sets of simulations need to be performed to test the validity of the algorithm. In the following tests, $\underline{\underline{\Sigma}} = \underline{\underline{I}}$ is assumed to hold although the test matrix had a random correlation structure, but always with diagonal elements of unit dispersion.

In all cases, the impulsiveness was kept constant ($\alpha = 1$ for cases 1 and 4, and $\alpha = 2$ for 2 and 3 as described below). The generalized signal-to-noise ratio used below is defined as

$$\mathrm{GSNR} = 10\log_{10}\left(\frac{\gamma_s}{\gamma_n}\right) = -10\log_{10}(\gamma_n) \tag{29}$$

where dispersion of signal is γ_s and of noise γ_n with s and n defined in Fig. 3. The assumption that $\gamma_s = 1$ does not reduce generality since the GSNR was varied by changing the dispersion of the noise.

Fig. 5 shows the mean squared error (average of the square of the errors in the angle estimation,

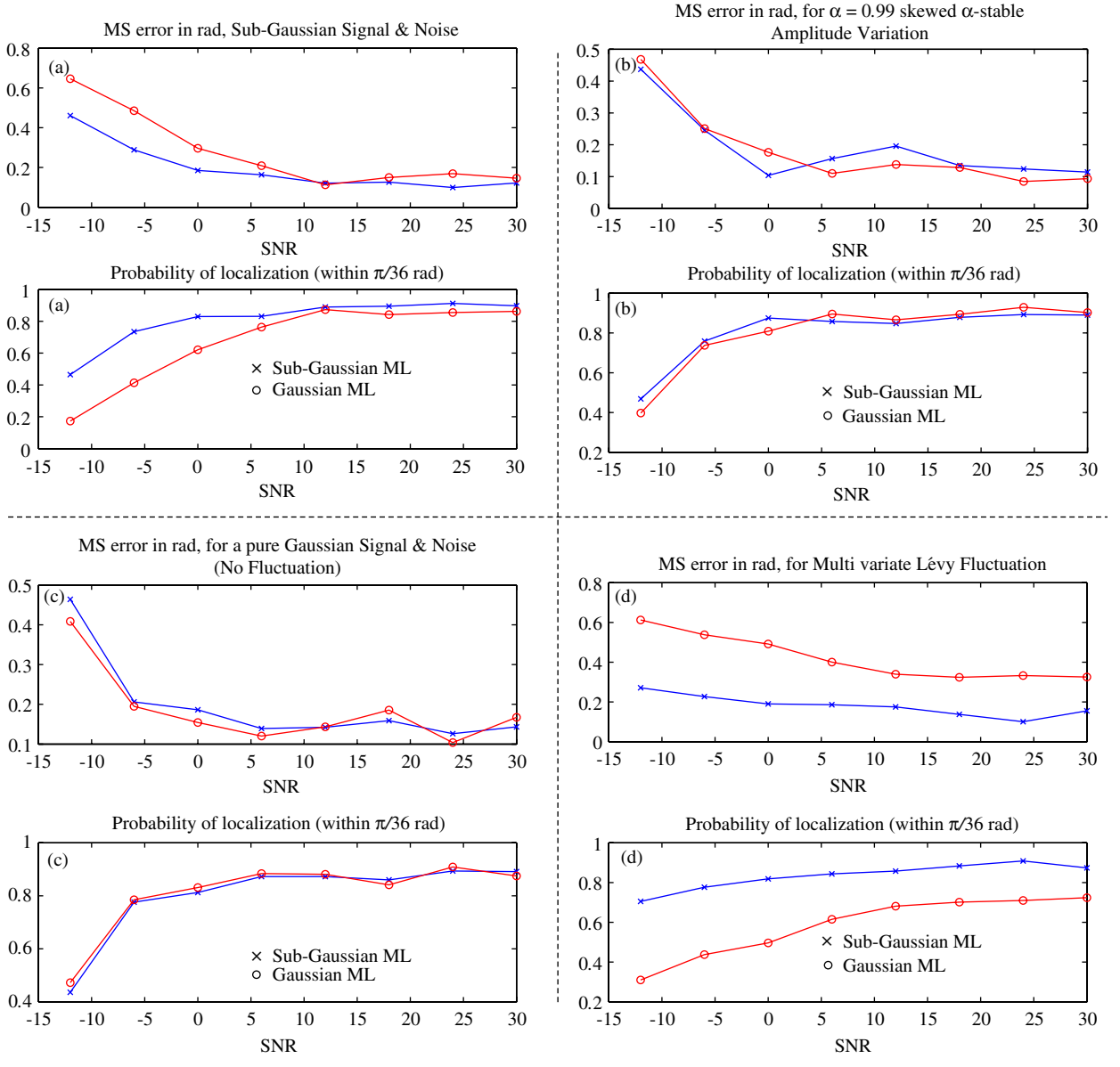

Fig. 5. Localization simulations for different noise conditions as described in text show that the sub-Gaussian based ML estimator is more robust than the Gaussian.

where the error is measured in rad) and the probability of localization for the conditions described in Fig. 3. A source is considered correctly localized if the estimate is within $\pm 5^\circ$ of the real azimuth. In the experiments, two sources were randomly (uniform in $0 \leqslant \theta \leqslant \pi$) located and the signal was received by 8 sensors. The block size was 32 samples and 1000 Monte Carlo runs were performed for each data point. Four cases were simulated, and in all cases the noise follows the same assumptions as the signal:

(a) Exactly as per the derivation assumptions (Fig. 5a): received signal is sub-Gaussian, created from a multivariate Gaussian and a univariate Lévy (can be viewed as Lévy energy fluctuation). Received signal impulsiveness is $\alpha = 1$.
(b) The impulsiveness assumption is relaxed (Fig. 5b), and the signal is created from a multivariate Gaussian ($\underline{\boldsymbol{v}}$) and a univariate skewed ($\beta = 1$) α-stable with $\alpha = 0.99$ (w). Received signal impulsiveness is $\alpha = 1.98$, which is almost Gaussian for all practical considerations.
(c) The signal is a multivariate Gaussian (Fig. 5c), and it undergoes *no* energy fluctuation ($w = 1$, $\underline{\boldsymbol{v}} = \underline{\boldsymbol{s}}$). This conforms to the assumptions of the well known Gaussian based ML. Clearly, the received signal impulsiveness is $\alpha = 2$.
(d) Finally, the dependence assumption is removed (Fig. 5d). The signal is created from a multivariate Gaussian ($\underline{\boldsymbol{v}}$) and a *multivariate* Lévy ($\underline{\boldsymbol{w}}$). In this case, the signals can be viewed as simply Cauchy, or as Gaussian with a different Lévy energy fluctuation for each source. Received signal impulsiveness is $\alpha = 1$.

As expected, the sub-Gaussian ML method performs better when the derivation assumptions hold (Fig. 5a). Likewise, when the signal is a multivariate Gaussian, the Gaussian ML algorithm performs better (Fig. 5c).

In the cases that neither assumption holds however, we can see how more robust the sub-Gaussian ML method is. When the signal follows (2) (Fig. 5b), the sub-Gaussian ML performs slightly better than the Gaussian ML under low SNR conditions. However, the real benefit of the proposed ML method can be observed when the signals are impulsive due to random multiplicative noise, independent from one source to the next (Fig. 5d). We expect that in any multiplicative noise environment the conditions encountered will be likely reflected in a mixture of the conditions in (a), (b), and (d) where the impulsiveness will vary, as will the strength of dependence of the various sources.

4.2. Estimating signal statistics

4.2.1. Estimating Gaussian statistics from sub-Gaussian signal

For initial estimates of $\underline{\underline{\boldsymbol{\Sigma}}}_{\mathrm{ML}}$ for Eq. (25), fractional lower order statistics can be employed. This would reduce the computation for converging to the correct estimates in (25). Consider the Gaussian signals v_1 and v_2, with covariance σ_{12}^2, i.e., $\Sigma = [1 \;\; \sigma_{12};\;\; \sigma_{12}^* \;\; 1]$ and a Lévy sequence u, used to create a sub-Gaussian signal. The requirement is to extract Σ from the signal $s = u \;\; \underline{\boldsymbol{v}}$.

If second order statistics were defined, then one could proceed in the usual way of:

$$\begin{aligned}\mathbf{E}[(u\,v_1)(u\,v_2)^*] &= \mathbf{E}[u\,v_1\,u^*\,v_2^*] \\ &= \mathbf{E}[u\,u^*]\,\mathbf{E}[v_1\,v_2^*] \\ &\propto \sigma_{12}.\end{aligned}$$

However, $\mathbf{E}[u\,u^*]$ does not exist, so one is required to use lower order statistics. We use the fractional order correlation function (FOCF) commonly employed as a robust alternative to the covariation function [4,8] as for example in the development of the FLOS-PHAT algorithm [8] (Fig. 6). The FOCF A_{xy} is defined as

$$A_{xy} = \mathbf{E}\{u^{\langle p\rangle} v^{*\langle q\rangle}\}. \tag{30}$$

Clearly, in this case:

$$\begin{aligned}&\mathbf{E}[(u\;v_1)^{\langle p\rangle}(u\;v_2)^{*\langle p\rangle}] \\ &\quad = \mathbf{E}[u^{\langle p\rangle}\;u^{*\langle p\rangle}]\mathbf{E}[v_1^{\langle p\rangle}\;v_2^{*\langle p\rangle}] \\ &\quad \propto \mathbf{E}[v_1^{\langle p\rangle}\;v_2^{*\langle p\rangle}] \quad \forall\, p<0.25.\end{aligned}$$

The following simulations demonstrate the connection between the FLOS statistics and the second order statistics of the Gaussian part of the signal. We see a deterministic correspondence when $2p$ is lower than $0.5 = \alpha_u$. Clearly, as we increase the value of p, we approach the region where statistics are not defined, and for any value of $p > 0.25$ (i.e., $2p > 0.5$, the characteristic exponent of the Lévy density), we have undefined statistics as predicted by theory [4,2,5]. This property can be used to estimate the underlying Gaussian statistics of a sub-Gaussian signal. Note however that these plots assume that both underlying Gaussians are of

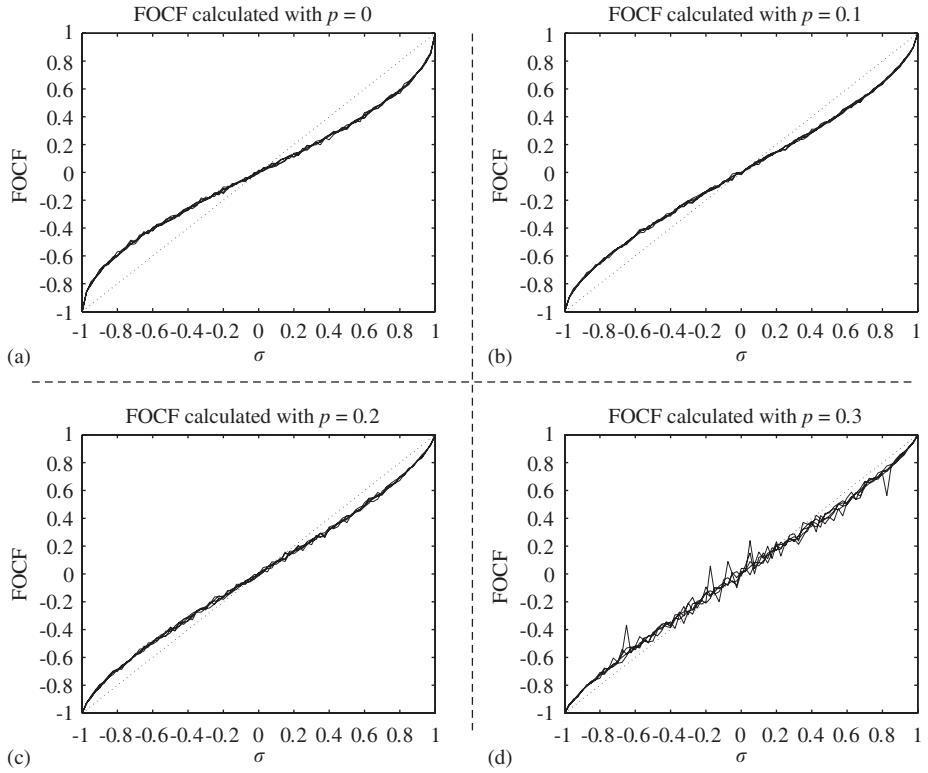

Fig. 6. Estimates of the FOCF measure can be deterministically connected to the covariance of the underlying Gaussian densities. Plot denotes true covariance versus the FOCF estimate.

the same dispersion (unity), and thus, in order to be useful, mapping tables need to be created for different dispersion ratios.

4.2.2. Simulations

As mentioned earlier, random initial conditions were sufficient for the solution of (25). Fig. 7 shows the estimates for a 3-source problem with

$$\underline{\underline{\Sigma}} = \begin{bmatrix} 2 & -1-0.4\mathrm{i} & 1.0-1.6\mathrm{i} \\ -1+0.4\mathrm{i} & 4 & -0.3-0.8\mathrm{i} \\ 1+1.6\mathrm{i} & -0.3+0.8\mathrm{i} & 3 \end{bmatrix}.$$

The sample statistics are slightly different from the above and are plotted on Fig. 7 as well. The histogram plots show the sample statistics on the positive side, and the estimates of the diagonal elements of $\hat{\underline{\underline{\Sigma}}}$ obtained by (25) on the negative side. The small complex components of the diagonal of $\hat{\underline{\underline{\Sigma}}}$ are ignored due to prior knowledge. The scatter plots present the off-diagonal elements of the statistics matrix on an Argand diagram. The smaller circles denote the 95% confidence interval for the actual sample statistics while the estimates are shown with the dots, and their 95% confidence curve with the larger circles. Individual realizations of the actual statistics are not shown since the confidence ellipse is sufficient.

The number of sensors is far more important than the total number of samples, as can be observed from the top row of Fig. 7. We note that cases (a) and (b) have the same overall number of samples, but the performance is far superior in case (b) where the number of sensors is 4 times the ones in (a). In fact, we can see from (c) that a significant decrease of SNR can be compensated by an increase in the number of sensors.

Likewise, on the second row of Fig. 7 we observe that even a significant increase in SNR from (d) to (e) provides little improvement in the accuracy of the estimates, while an increase in the number of sensors dramatically improves the accuracy.

5. Conclusions

We have shown that an alternative to the Cauchy density can be found with the help of

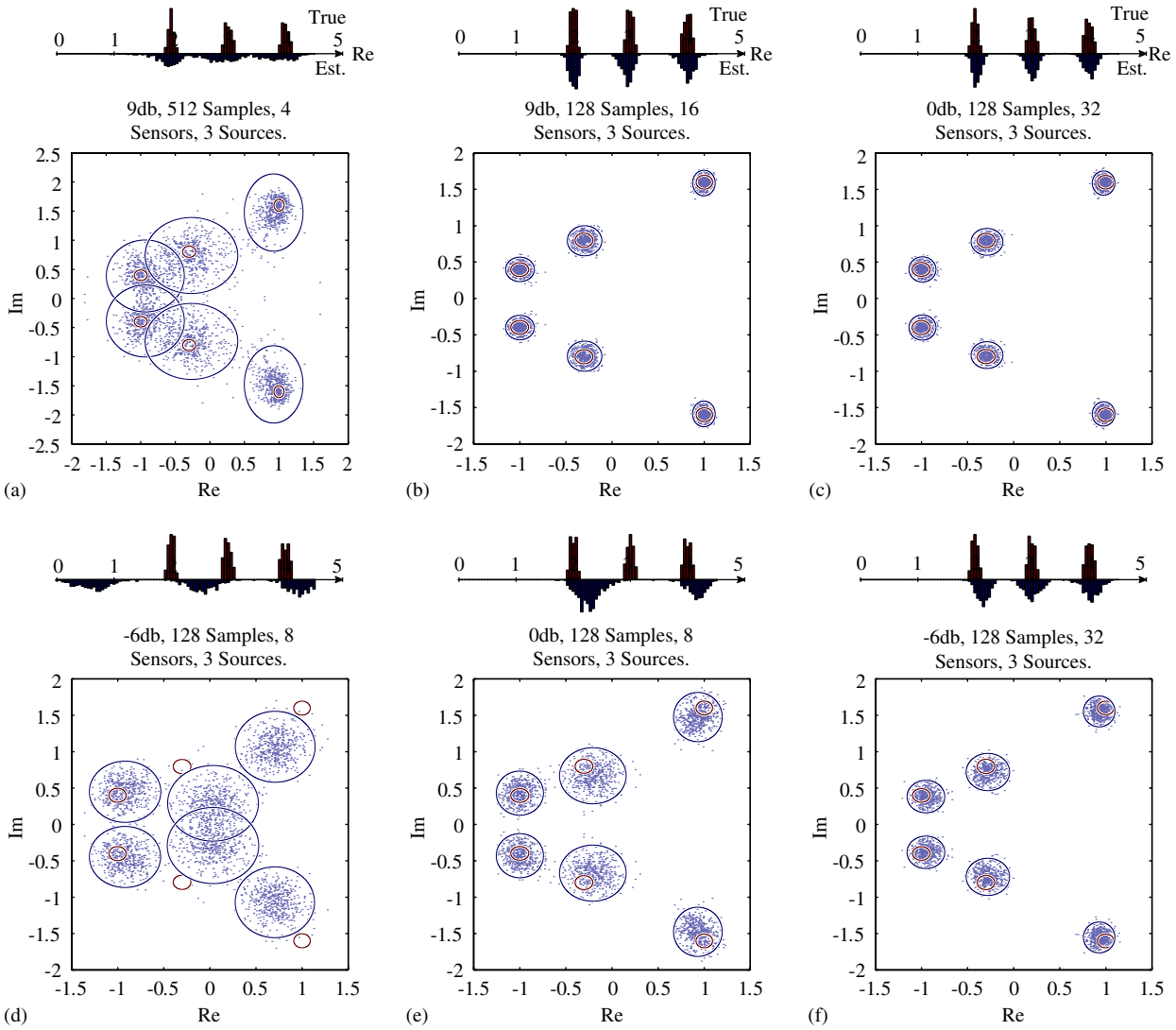

Fig. 7. Simulations show the effectiveness of the separable ML estimation of statistics for a 3-source problem under various noise conditions and array arrangements. The GSNR, number of frequency bins, M, number of sensors and number of samples shown above.

sub-Gaussian distributions. The advantage of the sub-Gaussian model is the simple connection between the data and the underlying signal statistics, and the ability to accommodate signals of a much more impulsive nature than the Gaussian model. Additionally, the model does not suffer when the dependence assumption is weakened as observed from the simulations.

The derived ML algorithm is robust in that it can operate under both impulsive and Gaussian signal conditions, and under both dependent and independent signals of interest. The suggested model can be used in several different scenarios: Gaussian sources undergoing the same energy fluctuations, dependent and impulsive sources, independent and impulsive sources, etc. Possible applications among others are modeling multipath signals, which can be highly dependent and impulsive, and modeling echoic and reverberant signals in an acoustic environment.

Acknowledgments

The authors would like to thank Prof. C.L. Nikias who initiated the study on the subject and whose knowledge and support are always available.

Appendix A. Derivation of 1-sub-Gaussian density

The derivations in this appendix follow work in [20]. From (8) the signal $\underline{s} = [s_1 \dots s_\kappa]^{\mathrm{T}}$ is of the form (dropping time or frequency dependencies for convenience):

$$s_k = u_k^{1/2} \cdot \underline{v}_k = w_k \cdot \underline{v}_k. \tag{A.1}$$

In order to find the distribution of $\underline{s}_k$, we need the distribution of w_k. From [21] it can be shown that:

$$f(w) = \begin{cases} \dfrac{w^{-2}\mathrm{e}^{-1/4w^2}}{\sqrt{\pi}} & \text{if } w>0, \\ 0 & \text{if } w<0, \end{cases} \tag{A.2}$$

The distribution of the transmitted signal can now be given by the multivariate distribution function:

$$\begin{aligned} F(\underline{\underline{\mathbf{S}}}) &= \int_{w=-\infty}^{+\infty} \int_{\underline{v}=-\infty}^{\underline{s}/w} f(w) f(\underline{v})\, \mathrm{d}\,\underline{v}\, \mathrm{dw} \\ &= \int_{w=0}^{+\infty} \int_{\underline{v}=-\infty}^{\underline{s}/w} \frac{w^{-2}\mathrm{e}^{-1/4w^2}}{\sqrt{\pi}} \frac{1}{\pi^\kappa |\underline{\underline{\boldsymbol{\Sigma}}}|} \\ &\quad \times \exp(-\underline{v}^\dagger \underline{\underline{\boldsymbol{\Sigma}}}^{-1}\, \underline{v})\, \mathrm{d}\,\underline{v}\, \mathrm{dw}. \end{aligned} \tag{A.3}$$

Differentiating with respect to $\underline{s}$, and then integrating with respect to w

$$\begin{aligned} f(\underline{s}) &= \frac{\mathrm{d}}{\mathrm{d}\,\underline{s}} \int_{w=0}^{+\infty} \int_{\underline{v}=-\infty}^{\underline{s}/w} \frac{w^{-2}\mathrm{e}^{-1/4w^2}}{\sqrt{\pi}} \frac{1}{\pi^\kappa |\underline{\underline{\boldsymbol{\Sigma}}}|} \\ &\quad \times \exp(-\underline{v}^\dagger \underline{\underline{\boldsymbol{\Sigma}}}^{-1}\, \underline{v})\, \mathrm{d}\,\underline{v}\, \mathrm{dw} \\ &= \int_{w=0}^{+\infty} \frac{\mathrm{d}}{\mathrm{d}\,\underline{s}} \left(\int_{\underline{v}=-\infty}^{\underline{s}/w} \frac{w^{-2}\mathrm{e}^{-1/4w^2}}{\sqrt{\pi}\pi^\kappa |\underline{\underline{\boldsymbol{\Sigma}}}|} \exp(-\underline{v}^\dagger \underline{\underline{\boldsymbol{\Sigma}}}^{-1}\, \underline{v})\, \mathrm{d}\,\underline{v} \right) \mathrm{dw} \\ &= \int_{w=0}^{+\infty} \frac{\mathrm{d}}{\mathrm{d}\,\underline{v}} \left(\int_{\underline{v}=-\infty}^{\underline{s}/w} \frac{w^{-2}\mathrm{e}^{-1/4w^2}}{\sqrt{\pi}\pi^\kappa |\underline{\underline{\boldsymbol{\Sigma}}}|} \right. \\ &\quad \left. \times \exp(-\underline{v}^\dagger \underline{\underline{\boldsymbol{\Sigma}}}^{-1}\, \underline{v})\, \mathrm{d}\,\underline{v} \right) w^{-2\kappa}\, \mathrm{dw} \\ &= \int_{w=0}^{+\infty} \frac{w^{-2}\mathrm{e}^{-1/4w^2}}{\sqrt{\pi}\pi^\kappa |\underline{\underline{\boldsymbol{\Sigma}}}|} \exp(-\underline{s}^\dagger \underline{\underline{\boldsymbol{\Sigma}}}^{-1}\, \underline{s}/w^2) w^{-2\kappa}\, \mathrm{dw} \\ &= \int_{w=0}^{+\infty} \frac{1}{\pi^{(2\kappa+1)/2} |\underline{\underline{\boldsymbol{\Sigma}}}|}\, w^{-2-2\kappa} \\ &\quad \times \exp\left(-\tfrac{1}{4} w^{-2} - \underline{s}^\dagger \underline{\underline{\boldsymbol{\Sigma}}}^{-1}\, \underline{s}/w^2 \right) \mathrm{dw} \\ &= \int_{w=0}^{+\infty} P w^{-2-2\kappa} \exp(-G w^{-2})\, \mathrm{dw}, \end{aligned} \tag{A.4}$$

where

$$P = \frac{1}{\pi^{(2\kappa+1)/2} |\underline{\underline{\boldsymbol{\Sigma}}}|} \tag{A.5}$$

and

$$G = \tfrac{1}{4} + \underline{s}^\dagger \underline{\underline{\boldsymbol{\Sigma}}}^{-1}\, \underline{s}. \tag{A.6}$$

Substitute:

$$t = w^{-2} \therefore w = t^{-1/2} \therefore \mathrm{dw} = -\tfrac{1}{2} \mathrm{t}^{-3/2}\, \mathrm{dt} \tag{A.7}$$

$$\begin{aligned} f(\underline{s}) &= \int_{w=0}^{+\infty} P w^{-2-2\kappa} \exp(-G w^{-2})\, \mathrm{dw} \\ &= \int_{t=+\infty}^{0} -\frac{1}{2}\, P\, t^{(2+2\kappa)/2} \exp(-Gt) t^{-3/2}\, \mathrm{dt} \\ &= \int_{t=+\infty}^{0} -\frac{1}{2}\, P\, t^{(-1+2\kappa)/2} \exp(-Gt)\, \mathrm{dt}. \end{aligned} \tag{A.8}$$

From tables [22]:

$$\int_{t=0}^{\infty} t^{n-1} \mathrm{e}^{-(a+1)t}\, \mathrm{d}t = \frac{\Gamma(n)}{(a+1)^n} \quad \forall n>0,\ a>-1 \tag{A.9}$$

$$\begin{aligned} f(\underline{s}) &= \int_{t=+\infty}^{0} -\frac{1}{2}\, P\, t^{(-1+2\kappa)/2} \exp(-Gt)\, \mathrm{dt} \\ &= \frac{1}{2}\, P\, \frac{\Gamma((1+2\kappa)/2)}{G^{(1+2\kappa)/2}} \\ &= \frac{1}{2\pi^{(2\kappa+1)/2} |\underline{\underline{\boldsymbol{\Sigma}}}|} \frac{\Gamma((1+2\kappa)/2)}{[1/4 + \underline{s}^\dagger \underline{\underline{\boldsymbol{\Sigma}}}^{-1}\, \underline{s}]^{(1+2\kappa)/2}}. \end{aligned} \tag{A.10}$$

Therefore

$$f(\underline{s}) = \frac{1}{2\pi^{(2\kappa+1)/2} |\underline{\underline{\boldsymbol{\Sigma}}}|} \frac{\Gamma((2\kappa+1)/2)}{[1/4 + \underline{s}^\dagger \underline{\underline{\boldsymbol{\Sigma}}}^{-1}\, \underline{s}]^{(2\kappa+1)/2}}. \tag{A.11}$$

where the Gamma function is given by

$$\Gamma(z) = \int_0^\infty t^{z-1} \mathrm{e}^{-t}\, \mathrm{dt}.$$

Appendix B. Derivation of ML statistics estimator

Summation is omitted for the moment, and thus we define the function to be minimized as

$$\begin{aligned} \mathscr{L} &= \left\{ \log_{\mathrm{e}} |\underline{\underline{\mathbf{A}}}\, \underline{\underline{\boldsymbol{\Sigma}}}_v\, \underline{\underline{\mathbf{A}}}^\dagger + \sigma_\eta^2 \underline{\underline{\mathbf{I}}}_\rho| \right. \\ &\quad \left. + \left(\frac{2\rho+1}{2} \right) \log_{\mathrm{e}} [\underline{x}^\dagger \{ \underline{\underline{\mathbf{A}}}\, \underline{\underline{\boldsymbol{\Sigma}}}_v \underline{\underline{\mathbf{A}}}^\dagger + \sigma_\eta^2 \underline{\underline{\mathbf{I}}}_\rho \}^{-1}\, \underline{x} + 1/4] \right\} \\ &= \underbrace{\log_{\mathrm{e}} |\underline{\underline{\mathbf{R}}}|}_{\mathscr{L}_1} + \left(\frac{2\rho+1}{2} \right) \underbrace{\log_{\mathrm{e}} \left[\underline{x}^\dagger \underline{\underline{\mathbf{R}}}^{-1}\, \underline{x} + 1/4 \right]}_{\mathscr{L}_2}. \end{aligned} \tag{B.1}$$

Operating on the first term:

$$\frac{\partial \mathscr{L}_1}{\partial \sigma_{ij}} = \frac{\partial \log_e |\underline{\underline{\mathbf{R}}}|}{\partial \sigma_{ij}} = \mathrm{Tr}\left[\left\{\frac{\partial \log_e |\underline{\underline{\mathbf{R}}}|}{\partial \underline{\underline{\mathbf{R}}}}\right\}^{\mathrm{T}} \frac{\partial \underline{\underline{\mathbf{R}}}}{\partial \sigma_{ij}}\right] \tag{B.2}$$

but

$$\frac{\partial \log_e |\underline{\underline{\mathbf{R}}}|}{\partial \underline{\underline{\mathbf{R}}}} = [\underline{\underline{\mathbf{R}}}^{-1}]^{\mathrm{T}} \quad \text{and} \quad \frac{\partial \underline{\underline{\mathbf{R}}}}{\partial \sigma_{ij}} = \underline{\boldsymbol{a}}_i \underline{\boldsymbol{a}}_j^{\dagger}, \tag{B.3}$$

where $\underline{\boldsymbol{a}}_j$ is the jth column of $\underline{\underline{\mathbf{A}}}$. Hence

$$\frac{\partial \mathscr{L}_1}{\partial \sigma_{ij}} = \frac{\partial \log_e |\underline{\underline{\mathbf{R}}}|}{\partial \sigma_{ij}} = \mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\boldsymbol{a}}_i \underline{\boldsymbol{a}}_j^{\dagger}\right] = \underline{\boldsymbol{a}}_j^{\dagger} \underline{\underline{\mathbf{R}}}^{-1} \underline{\boldsymbol{a}}_i. \tag{B.4}$$

Similarly, for the second term:

$$\begin{aligned}
\frac{\partial \mathscr{L}_2}{\partial \sigma_{ij}} &= \frac{\partial \log_e\left[\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{C}}}\right] + 1/4\right]}{\partial \sigma_{ij}} \\
&= \frac{1}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{C}}}\right] + 1/4} \cdot \frac{\partial}{\partial \sigma_{ij}} \{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{C}}}\right] + 1/4\} \\
&= \frac{1}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{C}}}\right] + 1/4} \cdot \mathrm{Tr}\left[\left[\frac{\partial}{\partial \underline{\underline{\mathbf{R}}}}\left\{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{C}}}\right]\right\}\right]^{\mathrm{T}} \frac{\partial \underline{\underline{\mathbf{R}}}}{\partial \sigma_{ij}}\right] \\
&= -\frac{1}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{C}}}\right] + 1/4} \cdot \mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \ \underline{\underline{\mathbf{C}}} \ \underline{\underline{\mathbf{R}}}^{-1} \ \underline{\boldsymbol{a}}_i \ \underline{\boldsymbol{a}}_j^{\dagger}\right] \\
&= -\frac{1}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{C}}}\right] + 1/4} \cdot \underline{\boldsymbol{a}}_j^{\dagger} \underline{\underline{\mathbf{R}}}^{-1} \ \underline{\underline{\mathbf{C}}} \ \underline{\underline{\mathbf{R}}}^{-1} \ \underline{\boldsymbol{a}}_i,
\end{aligned} \tag{B.5}$$

where we define $\underline{\underline{\mathbf{C}}} = \underline{\boldsymbol{x}}\,\underline{\boldsymbol{x}}^{\dagger}$.

Therefore

$$\begin{aligned}
\frac{\partial \mathscr{L}}{\partial \sigma_{ij}} &= \frac{\partial \mathscr{L}_1}{\partial \sigma_{ij}} + \left(\frac{2\rho+1}{2}\right) \frac{\partial \mathscr{L}_2}{\partial \sigma_{ij}} \\
&= \underline{\boldsymbol{a}}_j^{\dagger} \underline{\underline{\mathbf{R}}}^{-1} \underline{\boldsymbol{a}}_i - \frac{((2\rho+1)/2)}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{C}}}\right] + 1/4} \cdot \underline{\boldsymbol{a}}_j^{\dagger} \underline{\underline{\mathbf{R}}}^{-1} \ \underline{\underline{\mathbf{C}}} \ \underline{\underline{\mathbf{R}}}^{-1} \ \underline{\boldsymbol{a}}_i \\
&= \underline{\boldsymbol{a}}_j^{\dagger} \left[\underline{\underline{\mathbf{R}}}^{-1} - \left(\frac{2\rho+1}{2}\right) \frac{\underline{\underline{\mathbf{R}}}^{-1} \ \underline{\underline{\mathbf{C}}} \ \underline{\underline{\mathbf{R}}}^{-1}}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{C}}}\right] + 1/4}\right] \underline{\boldsymbol{a}}_i
\end{aligned} \tag{B.6}$$

or in matrix notation

$$\begin{aligned}
\frac{\partial \mathscr{L}}{\partial \Sigma} &= \underline{\underline{\mathbf{A}}}^{\dagger} \left[\underline{\underline{\mathbf{R}}}^{-1} - \left(\frac{2\rho+1}{2}\right) \frac{\underline{\underline{\mathbf{R}}}^{-1} \ \underline{\underline{\mathbf{C}}} \ \underline{\underline{\mathbf{R}}}^{-1}}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{C}}}\right] + 1/4}\right] \underline{\underline{\mathbf{A}}}_i \\
&= \underline{\underline{\mathbf{A}}}^{\dagger} \underline{\underline{\mathbf{R}}}^{-1} \left[\underline{\underline{\mathbf{R}}} - \frac{(2\rho+1/2)\,\underline{\underline{\mathbf{C}}}}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{C}}}\right] + 1/4}\right] \underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{A}}}_i.
\end{aligned} \tag{B.7}$$

Use Sherman–Morrison–Woodbury identity:

$$(\underline{\underline{\mathbf{A}}} + \underline{\underline{\mathbf{U}}}\,\underline{\underline{\mathbf{V}}}^{\dagger})^{-1} = \underline{\underline{\mathbf{A}}}^{-1} - \underline{\underline{\mathbf{A}}}^{-1} \underline{\underline{\mathbf{U}}} (\underline{\underline{\mathbf{I}}} + \underline{\underline{\mathbf{V}}}^{\dagger} \underline{\underline{\mathbf{A}}}^{-1} \ \underline{\underline{\mathbf{U}}})^{-1} \underline{\underline{\mathbf{V}}}^{\dagger} \underline{\underline{\mathbf{A}}}^{-1} \tag{B.8}$$

with (22), the relation connecting $\underline{\underline{\mathbf{R}}}$ to the original signal statistics $\underline{\underline{\mathbf{\Sigma}}}$,

$$\Rightarrow \quad \underline{\underline{\mathbf{R}}}^{-1} = \frac{1}{\sigma_\eta^2} \{\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{A}}} (\underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger} \ \underline{\underline{\mathbf{A}}} + \sigma_\eta^2 \underline{\underline{\mathbf{I}}})^{-1} \ \underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger}\} \tag{B.9}$$

and in order to substitute back in (B.7) we calculate

$$\begin{aligned}
\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{A}}} &= \frac{1}{\sigma_\eta^2} \{\underline{\underline{\mathbf{I}}}_\rho - \underline{\underline{\mathbf{A}}} (\underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger} \ \underline{\underline{\mathbf{A}}} + \sigma_\eta^2 \underline{\underline{\mathbf{I}}})^{-1} \ \underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger}\} \underline{\underline{\mathbf{A}}} \\
&= \frac{1}{\sigma_\eta^2} \{\underline{\underline{\mathbf{A}}} - \underline{\underline{\mathbf{A}}} (\underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger} \ \underline{\underline{\mathbf{A}}} + \sigma_\eta^2 \underline{\underline{\mathbf{I}}})^{-1} \ \underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger} \underline{\underline{\mathbf{A}}}\} \\
&= \frac{1}{\sigma_\eta^2} \underline{\underline{\mathbf{A}}} \{\underline{\underline{\mathbf{I}}}_\kappa - (\underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger} \ \underline{\underline{\mathbf{A}}} + \sigma_\eta^2 \underline{\underline{\mathbf{I}}})^{-1} \ \underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger} \underline{\underline{\mathbf{A}}}\} \\
&= \frac{1}{\sigma_\eta^2} \underline{\underline{\mathbf{A}}} \{\underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger} \ \underline{\underline{\mathbf{A}}} + \sigma_\eta^2 \underline{\underline{\mathbf{I}}}\}^{-1} \{\underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger} \ \underline{\underline{\mathbf{A}}} + \sigma_\eta^2 \underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger} \ \underline{\underline{\mathbf{A}}}\} \\
&= \underline{\underline{\mathbf{A}}} \{\underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger} \ \underline{\underline{\mathbf{A}}} + \sigma_\eta^2 \underline{\underline{\mathbf{I}}}\}^{-1}.
\end{aligned} \tag{B.10}$$

Therefore, from (B.7)

$$\begin{aligned}
\frac{\partial \mathscr{L}}{\partial \Sigma} = \{\underline{\underline{\mathbf{A}}}^{\dagger} \ \underline{\underline{\mathbf{A}}} \ \underline{\underline{\mathbf{\Sigma}}} + \sigma_\eta^2 \underline{\underline{\mathbf{I}}}\}^{-1} \underline{\underline{\mathbf{A}}}^{\dagger} \Bigg[\underline{\underline{\mathbf{R}}} - \left(\frac{2\rho+1}{2}\right) \\
\times \frac{\underline{\underline{\mathbf{C}}}}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\underline{\mathbf{C}}}\right] + 1/4}\Bigg] \underline{\underline{\mathbf{A}}} \{\underline{\underline{\mathbf{\Sigma}}} \ \underline{\underline{\mathbf{A}}}^{\dagger} \ \underline{\underline{\mathbf{A}}} + \sigma_\eta^2 \underline{\underline{\mathbf{I}}}\}^{-1}
\end{aligned} \tag{B.11}$$

But at ML conditions, $\partial\mathscr{L}/\partial\Sigma = 0$

$$\therefore \underline{\underline{\mathbf{A}}}^{\dagger}\left[\underline{\underline{\mathbf{A}}}\,\underline{\underline{\boldsymbol{\Sigma}}}_{v}\,\underline{\underline{\mathbf{A}}}^{\dagger}+\sigma_{\eta}^{2}\,\underline{\underline{\mathbf{I}}}-\left(\frac{2\rho+1}{2}\right)\frac{\underline{\underline{\mathbf{C}}}}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1}\,\underline{\underline{\mathbf{C}}}\right]+1/4}\right]\underline{\underline{\mathbf{A}}}$$
$$=\underline{\underline{\mathbf{A}}}^{\dagger}\,\underline{\underline{\mathbf{A}}}\,\underline{\underline{\boldsymbol{\Sigma}}}_{v}\,\underline{\underline{\mathbf{A}}}^{\dagger}\,\underline{\underline{\mathbf{A}}}+\underline{\underline{\mathbf{A}}}^{\dagger}\,\sigma_{\eta}^{2}\,\underline{\underline{\mathbf{A}}}$$
$$-\frac{(\rho+0.5)\underline{\underline{\mathbf{A}}}^{\dagger}\,\underline{\underline{\mathbf{C}}}\,\underline{\underline{\mathbf{A}}}}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1}\,\underline{\underline{\mathbf{C}}}\right]+1/4}=0 \qquad \text{(B.12)}$$

Solving for $\underline{\underline{\boldsymbol{\Sigma}}}$, we can find the $\underline{\underline{\boldsymbol{\Sigma}}}_{\mathrm{ML}}$ (over all available data)

$$\underline{\underline{\boldsymbol{\Sigma}}}_{\mathrm{ML}}=\frac{1}{M}\sum_{f=f_1}^{f_M}\left[\underline{\underline{\mathbf{A}}}^{-}\left(\frac{(\rho+0.5)\,\underline{\boldsymbol{x}}\,\underline{\boldsymbol{x}}^{\dagger}}{\mathrm{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1}\,\underline{\boldsymbol{x}}\,\underline{\boldsymbol{x}}^{\dagger}\right]+1/4}-\sigma_{\eta}^{2}\right)\underline{\underline{\mathbf{A}}}^{-\dagger}\right], \qquad \text{(B.13)}$$

where $\underline{\underline{\mathbf{A}}}^{-}=(\underline{\underline{\mathbf{A}}}^{\dagger}\,\underline{\underline{\mathbf{A}}})^{-1}\underline{\underline{\mathbf{A}}}^{\dagger}$, and $\underline{\underline{\mathbf{R}}}$, as defined in (B.9).

References

[1] S. Cambanis, G. Samorodnitsky, M.S. Taqqu (Eds.), Stable Processes and Related Topics, Progress in Probability, Birkhäuser, Boston, 1991 (Selection of Papers from the Workshop on Stable Processes and Related Topics held at Cornell University).

[2] G. Samorodnitsky, M.S. Taqqu, Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance, Chapman & Hall, New York, London, 1994.

[3] M. Shao, C.L. Nikias, Signal processing with fractional lower order moments: stable processes and their applications, Proc. IEEE 81 (7) (1993) 986–1010.

[4] C.L. Nikias, M. Shao, Signal Processing with Alpha-Stable Distributions and Applications, Wiley, New York, 1995.

[5] R. Adler, R.E. Feldman, M.S. Taqqu (Eds.), A Practical Guide to Heavy Tails: Statistical Techniques and Applications, Birkhäuser Editions, 1998.

[6] P.D. Ditlevsen, Observation of α-stable noise induced millennial climate changes from an ice-core record, Geophys. Res. Lett. 26 (1999) 1441–1444.

[7] P. Tsakalides, R. Raspanti, C.L. Nikias, Angle/doppler estimation in heavy-tailed clutter backgrounds, IEEE Trans. Aerospace Electronic Systems 35 (2) (1999) 419–436.

[8] P.G. Georgiou, P. Tsakalides, C. Kyriakakis, Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise, IEEE Trans. Multimedia 1 (3) (1999) 291–301.

[9] P. Tsakalides, C.L. Nikias, Maximum likelihood localization of sources in noise modeled as a stable process, IEEE Trans. Signal Process. 43 (11) (1995) 2700–2713.

[10] P. Tsakalides, C.L. Nikias, The robust covariation based music (ROC-MUSIC) algorithm for bearing in impulsive noise environments, IEEE Trans. Signal Process. 44 (7) (1996) 1623–1633.

[11] H. Stark, J.W. Woods, Probability, Random Processes and Estimation Theory for Engineers, second ed., Prentice-Hall, Englewood Cliffs, NJ, 1994.

[12] S. Cambanis, G. Miller, Linear problems in pth order and stable processes, SIAM J. Appl. Math. 41 (1) (1981) 43–69.

[13] P.G. Georgiou, Robust maximum likelihood source localization; the case for Sub-Gaussian vs. Gaussian, IEEE Trans. Speech, Audio, and Language Processing 14, in press.

[14] D.H. Johnson, D.E. Dudgeon, Array Signal Processing: Concepts and Techniques, Signal Processing Series, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[15] J.F. Bohme, Separated estimation of wave parameters and spectral parameters by maximum likelihood, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, 1986, pp. 2819–2822.

[16] A.G. Jaffer, Maximum likelihood direction finding of stochastic sources: a separable solution, in: Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, vol. 5, 1988, pp. 2893–2896.

[17] D. Starer, A. Nehorai, Newton algorithms for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise, IEEE Trans. Signal Process. 40 (6) (1992) 1528–1534.

[18] V.M. Zolotarev, One-dimensional stable distributions, Translations of Mathematical Monographs, vol. 65, American Mathematical Society, Providence, RI, 1986.

[19] W. Huyer, A. Neumaier, Global optimization by multilevel coordinate search, J. Global Optim. 14 (1999) 331–355.

[20] P. Georgiou, Robust signal processing techniques for source localization and multisource spatial sound rendering for immersive environments, Ph.D. dissertation, University of Southern California, 2002.

[21] A. Papoulis, Probability, Random Variables, and Stochastic Processes, third ed., McGraw-Hill, New York, 1991.

[22] D. Zwillinger, CRC Standard Mathematical Tables and Formulae, 30th ed., CRC-Press, 1995.