



Audio Engineering Society Convention Paper

Presented at the 113th Convention
2002 October 5–8 Los Angeles, CA, USA

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

An Alternative Model for Sound Signals Encountered in Reverberant Environments; Robust Maximum Likelihood Localization and Parameter Estimation Based on a Sub-Gaussian Model

Panayiotis G. Georgiou, Chris Kyriakakis

¹*Immersive Audio Laboratory, Integrated Media Systems Center, University of Southern California, Los Angeles, CA 90089-2564*

Correspondence should be addressed to Panayiotis G. Georgiou (georgiou@sipi.usc.edu)

ABSTRACT

In this paper we investigate an alternative to the Gaussian density for modeling signals encountered in audio environments. The observation that sound signals are impulsive in nature, combined with the reverberation effects commonly encountered in audio, motivates the use of the Sub-Gaussian density.

The new Sub-Gaussian statistical model and the separable solution of its Maximum Likelihood estimator are derived. These are used in an array scenario to demonstrate with both simulations and two different microphone arrays the achievable performance gains.

The simulations exhibit the robustness of the sub-Gaussian based method while the real world experiments reveal a significant performance gain, supporting the claim that the sub-Gaussian model is better suited for sound signals.

1 INTRODUCTION

In this paper we present an alternative model for signals encountered in audio environments. Motivated by the observation that noise in a room environment is mostly due to reverberation rather than independent sources, we derive a model of dependent source and noise. In addition, based on the demonstration of the impulsiveness of sound in previous work from these and other authors [11, 18], we decide to use the α -stable class of distributions, and more specifically their sub-Gaussian subset for our model.

We first derive the probability density function of a sub-Gaussian process with impulsiveness equal to that of a Cauchy distribution. Sub-Gaussian distributions are a special case of α -stable random processes [25], and they are variance mixtures of Gaussian random processes [7]. As such, irrespective of the correlation structure of the underlying Gaussian, the sub-Gaussian elements cannot be independent. The fact that reverberation in acoustic environments is not actually white uncorrelated noise, but rather highly dependent on the source signal, motivated our investigation of the sub-Gaussian model for audio signals.

Secondly, in order to test the validity of our model we express the issue of localization of multiple sources as a parameter estimation problem and we formulate the *Maximum Likelihood* (ML) estimator based on the derived density. The noise and signals are modeled as jointly sub-Gaussian (*i.e.*, they are being produced by the same Lévy sequence). We assume a scenario under which there are multiple sources received by an array of a greater or equal number of sensors. The transfer function each signal undergoes while traveling to the array can be modeled as an attenuation and a delay based on the far-field assumption.

We proceed to derive a separable solution for this estimator both for the statistics of the signals and the *Directions-of-Arrival* (DOA's). The separable solution assumes known statistics in order to recover the DOA's, and known DOA's in order to recover the statistics.

We then test, initially on simulated data, the performance of the sub-Gaussian based and Gaussian based ML estimators, and show the improved performance of the estimator based on the new model in comparison to the one based on the Gaussian. The algorithms are evaluated under a variety of signal conditions, and we demonstrate the robustness of the sub-Gaussian based ML, which performs well even under these other conditions, while the Gaussian based ML degrades in performance significantly when subjected to non-ideal conditions.

Finally, in order to test the localization algorithm with some real data, we constructed synthetic 20- and 41-microphone arrays in our Audio Lab (a room with acoustical characteristics resembling an average living room). The audio channels were played together through our 10.2 channel system at 48kHz and 2 microphones were shifted to form a linear array. The synchronized playback-recording feature of ProTools, confirmed by the addition of chirp synchronization signals at the start of the recording, ensured that the arrays were correctly created.

Results of localization demonstrate that the sub-Gaussian based ML method has a much better localization performance than its Gaussian counterpart. The sub-Gaussian ML is localizing correctly the sound sources in almost every case, while the Gaussian based ML has a very low probability of localization and a high mean square error. We additionally demonstrate the ability of the sub-Gaussian based method to accurately localize strong echos.

The development of the aforementioned work will follow an introduction of α -stable theory in Section 2 for the mathematically inclined reader. In Section 3, we will introduce the new model, derive its density, ML estimator and separable solution. The performance of the ML estimator based on the new sub-Gaussian model will be assessed via simulations in Section 4, and on the real data in Section 5. Finally, Section 6 will give one simple application of direction finding using in ML in audio.

2 BACKGROUND: ALPHA-STABLE DISTRIBUTIONS

The Gaussian distribution has traditionally been the most widely accepted distribution and used, as a rule, as a realistic model for various kinds of noise. 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, Samorodnitsky and Taquu [8] 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, in the field of linear filtering problems, and in the domain of spectral representation. A textbook of comprehensive

coverage of the α -stable theory was written by Samorodnitsky and Taqqu in 1994 [25].

In 1993, Nikias and Shao [27] gave an introductory review of α -stable distributions from a statistical signal processing viewpoint that was followed by a book from the same authors in 1995 [21].

Alpha-stable distributions have been used to model diverse phenomena such as radar clutter [31], random fluctuations of gravitational fields, economic market indices, data file sizes on the Web, and network traffic [1].

2.1 Theory

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.

- First, it satisfies the *stability property*, which states that if X , X_1 , and X_2 are α -stable independent random variables of the same distribution, then there exist μ_1 and μ_2 satisfying:

$$\nu_1 X_1 + \nu_2 X_2 \stackrel{d}{=} \mu_1 X + \mu_2 \tag{1}$$

where ν_1 , ν_2 , μ_1 and μ_2 are constants and $\stackrel{d}{=}$ denotes equality in distribution.

- Second, it satisfies the *Generalized Central Limit Theorem* [21, 25, 29] stating: X is α -stable, if and only if X is the limit in distribution of the sum:

$$S_n = \frac{X_1 + X_2 + \dots + X_n}{a_n} - b_n \tag{2}$$

where X_1, X_2, \dots , are i.i.d. r.v.'s and $n \rightarrow \infty$. Parameter b_n is real and a_n is real and positive.

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

$$\varphi(t) = \exp(i\lambda t - \gamma|t|^\alpha [1 + i\beta \text{sign}(t)\omega(t, \alpha)]) \tag{3}$$

where:

$$\omega(t, \alpha) = \begin{cases} \tan \frac{\alpha\pi}{2}, & \text{if } \alpha \neq 1 \\ \frac{2}{\pi} \log |t|, & \text{if } \alpha = 1 \end{cases} \tag{4}$$

$$\text{sign}(t) = \begin{cases} 1, & \text{if } t > 0 \\ 0, & \text{if } t = 0 \\ -1, & \text{if } t < 0 \end{cases} \tag{5}$$

and:

- α is the *characteristic exponent* satisfying $0 < \alpha \leq 2$. The characteristic exponent controls the

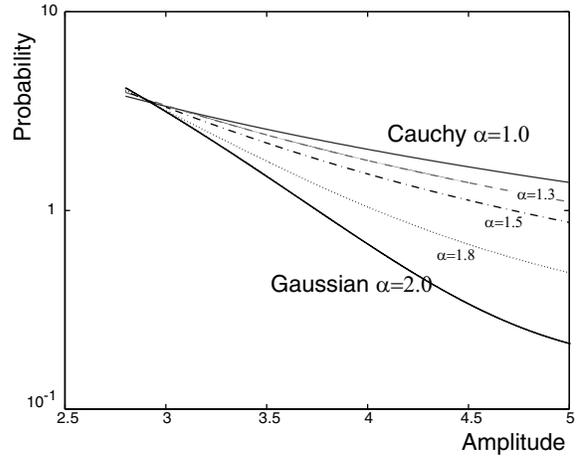


Fig. 1: The tails of the probability density function of a symmetric α -stable distribution for different values of α . The case of $\alpha = 2$ being the less impulsive case of Gaussian noise and $\alpha = 1$ the more impulsive Cauchy case. In each of the above cases the dispersion was kept constant at $\gamma = 1$.

thickness (also referred to as heaviness) of the tails of the density function. The tails are heavier, and thus the noise more impulsive for low values of α while for a larger α the distribution has a less impulsive behavior (Figs. 1 and 2).

- λ is the *location parameter* ($-\infty < \lambda < \infty$). It corresponds to the mean for $1 < \alpha \leq 2$ and the median for $0 < \alpha \leq 1$.

- γ is the *dispersion parameter* ($\gamma > 0$), which determines the spread of the density around its location parameter. The dispersion behaves in a similar way to the variance of the Gaussian density, and it is in fact equal to half the variance when $\alpha = 2$, the Gaussian case.

- β is the index of symmetry ($-1 \leq \beta \leq 1$). When $\beta = 0$, the distribution is symmetric around the location parameter.

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:

$$f_{\alpha=2}(\gamma, \lambda; x) = \frac{1}{\sqrt{4\pi\gamma}} \exp\left\{-\frac{(x-\lambda)^2}{4\gamma}\right\} \tag{6}$$

$$f_{\alpha=1}(\gamma, \lambda; x) = \frac{\gamma}{\pi[\gamma^2 + (x-\lambda)^2]} \tag{7}$$

A closed form expression also exists for the case of the Lévy distribution, which has parameters $\beta = 1$ and $\alpha =$

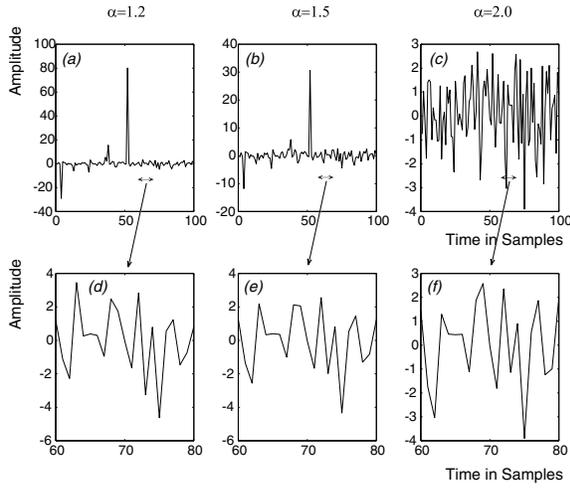


Fig. 2: Sample time series of $S\alpha S$ random variables. The characteristic exponents are $\alpha = 1.2$, $\alpha = 1.5$, and $\alpha = 2.0$ (Gaussian). The second row of figures shows an enlargement of parts of the above row, and demonstrates the similarities between the distributions.

0.5, and therefore is completely skewed to the positive axis.

$$f(x) = \begin{cases} x^{-\frac{3}{2}} \frac{e^{-\frac{1}{4}x}}{2\sqrt{\pi}} & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (8)$$

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 $f_{\alpha=0.5, \beta=-1}(x) = f_{\text{Lévy}}(-x)$.

The impulsiveness of the α -stable distribution can clearly be seen in Fig. 2(a), (b), and (c). However, when we take a closer look at Fig. 2(d), (e), and (f), the time series resulting from the three different distributions do not appear to be very different.¹ This encourages the use of α -stable distributions in situations where the noise has been traditionally modeled as Gaussian, but where sudden “spikes” might occur. For example, in an enclosed room sounds produced by pages turning, pens clicking, or objects falling can give rise to the impulsiveness in the noise.

The class of α -stable distributions does not possess finite second (or higher) moments. In fact, α -stable distributions with $\alpha \neq 2$ have finite moments only for order p

¹It is clear that if the signals were not sampled, self similarity would hold and the similarity between the signals of Fig. 2(d), (e) and (f) would not be apparent.

lower than α :

$$\begin{aligned} \alpha < 2, \mathbb{E}|X_\alpha|^p & \text{ Not Defined } \forall p \geq \alpha \\ \alpha < 2, \mathbb{E}|X_\alpha|^p & < \infty \quad \forall 0 \leq p < \alpha \\ \text{Gaussian: } \alpha = 2, \mathbb{E}|X_\alpha|^p & < \infty \quad \forall p \geq 0 \end{aligned}$$

References [1, 21, 27] and [25] treat the α -stable theory further.

2.2 Properties of α -stable signals

The *covariation* of two signals x and y is defined as:

$$[X, Y]_\alpha \triangleq \int_S xy^{\alpha-1} \mu(ds) = \frac{\mathbb{E}(XY^{\langle p-1 \rangle})}{\mathbb{E}(|Y|^p)} \gamma_y \quad (9)$$

where S is the unit circle, $\mu(\cdot)$ is the spectral measure of the $S\alpha S$ random vector (X, Y) , γ_y is the dispersion parameter of signal Y , p satisfies $1 \leq p < \alpha$, and $y^{\langle k \rangle} = |y|^{k-1} y^*$ is the signed-power non-linearity.

The covariation of complex jointly $S\alpha S$ random variables is not generally symmetric and has the following properties:

P1 If X_1, X_2 , and Y are jointly $S\alpha S$, then

$$[aX_1 + bX_2, Y]_\alpha = a[X_1, Y]_\alpha + b[X_2, Y]_\alpha \quad (10a)$$

for any complex constants a and b .

P2 If Y_1 and Y_2 are independent and Y_1, Y_2 and X are jointly $S\alpha S$, then

$$[aX, bY_1 + cY_2]_\alpha = ab^{\langle \alpha-1 \rangle} [X, Y_1]_\alpha + ac^{\langle \alpha-1 \rangle} [X, Y_2]_\alpha \quad (10b)$$

for any complex constants a, b and c .

P3 If X and Y are independent $S\alpha S$, then

$$[X, Y]_\alpha = 0 \quad (10c)$$

An alternative to the covariation measure is the *Fractional Lower Order Correlation Function* defined as:

$$A_{XY} = \mathbb{E} \{ X^{\langle p \rangle} Y^{*\langle q \rangle} \} \quad (11)$$

2.3 Sub-Gaussian Random Variables

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

$$\varphi(\underline{\mathbf{u}}) = \exp \left(-\frac{1}{2} [\underline{\mathbf{u}}^T \underline{\mathbf{R}} \underline{\mathbf{u}}]^{\alpha/2} \right) \quad (12)$$

where $\underline{\mathbf{R}}$ is a positive-definite matrix.

Sub-Gaussian processes are variance mixtures of Gaussian processes [7]. If $\underline{\mathbf{X}}(t)$ is sub-Gaussian with parameter α (will be denoted by α -SG($\underline{\mathbf{R}}$)) and S is a positive

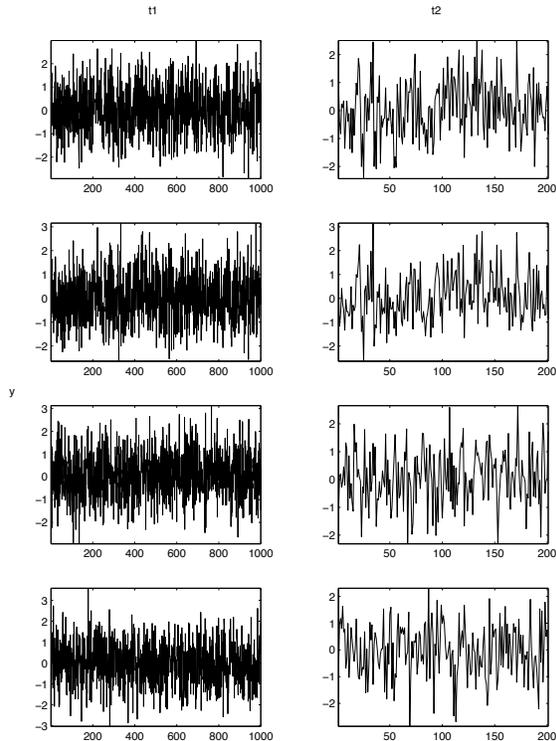


Fig. 3: Multivariate Gaussian random vector of size 4 along time with $\Sigma = [1 \ 0.9 ; 0.1 \ 0]$.

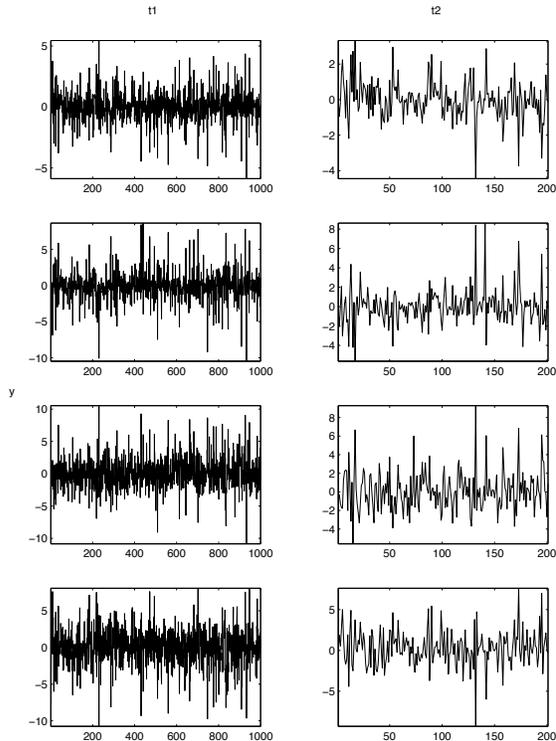


Fig. 4: Multivariate sub-Gaussian random vector of size 4 along time with $\alpha = 1$ and generated with the Gaussian signal of Fig. 3. The signals are very impulsive and hence $\text{sign}(x) \log_e(|x|)$ is shown on the graphs.

stable process with characteristic exponent $\alpha/2$ (i.e., S is $\frac{\alpha}{2}$ -stable random variable completely skewed to the right) and $\underline{Y}(t)$ is a multivariate Gaussian process independent of S , then:

$$\underline{X}(t) = S^{1/2} \underline{Y}(t) \tag{13}$$

Clearly from the above, irrespective of the correlation structure of $\underline{Y}(t)$, the components of $\underline{X}(t)$ can not be independent.

A multivariate Gaussian random vector of size 4 is shown in Fig. 3, while a sub-Gaussian random vector of impulsiveness $\alpha = 1$ is shown in Fig. 4. The sub-Gaussian random vector is obtained by eq. (13) using a Lévy random variable (8) for S and the multivariate Gaussian random of Fig. 3

3 THE NEW MODEL

We considered in previous work [11] a sound source localization method using an array of microphones based on the computationally simple *Time Delay Estimation*

(TDE) *Phase Transform Method* (PHAT). We demonstrated the impulsive nature of sound signals and provided a modified version of the common PHAT method that takes this nature into consideration. The resulting localization algorithm – the *Fractional Lower-Order Statistics* (FLOS) PHAT method – performed significantly better (up to a factor of 4) than the existing PHAT method. The development in [11] focused on a single source, two-sensor scenario.

We continue in this chapter our work on localization by focusing on the development of methods relating to the estimation of the parameters of a system – such as the one shown in Fig. 5 – where we assume multiple sources received by an arbitrary number of sensors (greater than the number of sources). Additionally, we are aiming to provide a more accurate statistical description of the signals encountered in acoustical environments.

This problem, which we initially visit from a completely theoretical perspective and later in the microphone array

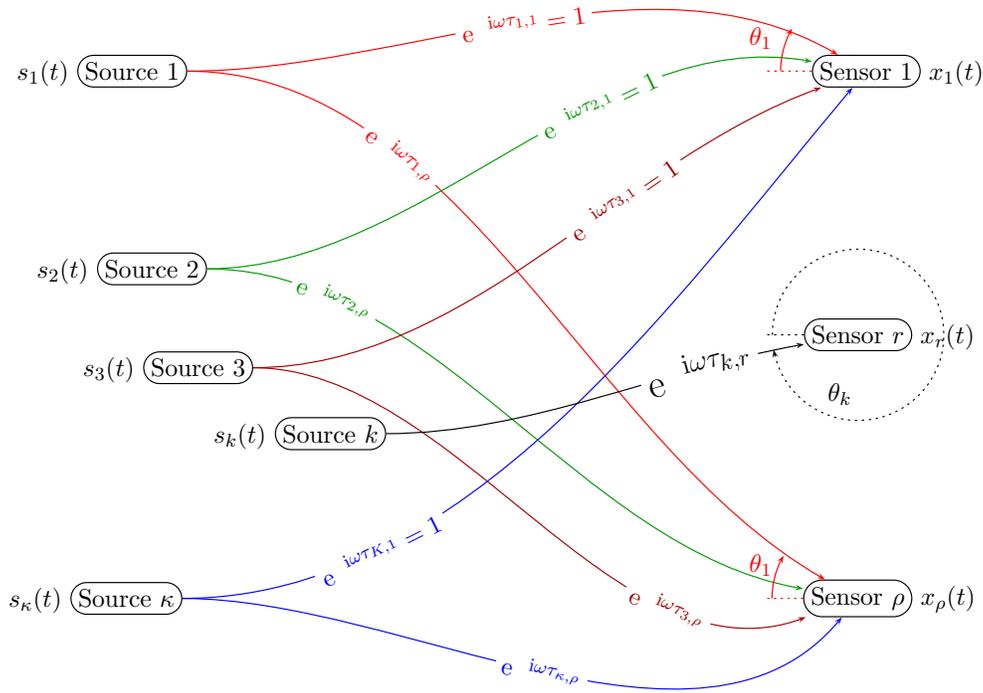


Fig. 5: We assume κ source signals being received by $\rho \geq \kappa$ sensors. Sources are assumed to be in the far-field, and thus there is a single incidence angle for each source on all sensors.

signal processing framework, can have significant applications in a variety of fields. The scenario of impulsive and multiplicative noise is encountered for instance in communications, owing to the presence of local scatterers in the vicinity of the mobile or due to wavefronts that propagate through random inhomogeneous media. Gershman *et al* [12] have, for example, presented a method that assumes a random phase perturbation along all source-sensor paths. Their method has led to a non-Gaussian model, and did not result in a ML estimator. Besson *et al* [3] suggested a similar localization algorithm for a source, which appears as a scatter of sources. Similarly Stoica *et al* [30] have presented a Gaussian based ML method in the presence of multiplicative noise, but constraining the amplitude of the noise to be 1. The model we propose in this chapter is well suited for such cases, even though experiments will be performed for audio signals only.

The transmitted signals for the development of the localization algorithm are assumed to be stochastic, and

as such, the parameters of interest will be their statistics and *Directions-of-Arrival* (DOA's). The estimation process combines the measurements to obtain a vector $\underline{x}(t)$, 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 $\underline{x}(t)$, and 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, *Maximum Likelihood* (ML), as well as constrained optimization criteria. Despite the wide variety of optimization criteria, the optimal detector is characterized by a single result: the Maximum Likelihood ratio test, which was also one of the first methods to be applied in the area of array signal processing [17], and which we will be using in this work.

The Maximum Likelihood technique applied to the source localization problem usually makes two different assumptions for the signal waveforms, resulting in two

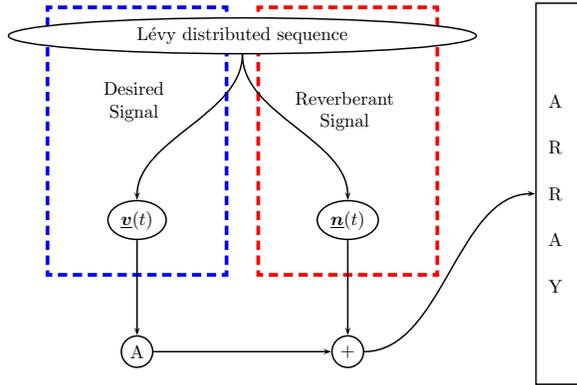


Fig. 6: A multivariate Gaussian signal corrupted by multiplicative Lévy noise is transformed through a set of delays to the receiving end of the array. Similarly, the additive noise is generated by the same Lévy sequence. This may be a good model for a reverberation noise, which is highly dependent on the signal of interest.

different ML methods. According to the *Stochastic ML* (SML), the signals are usually 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 to be unknown but deterministic. In this case, estimates of the signals as well as the DOA's are desired, while in the former case, the only parameters to be estimated are the statistics and DOA's. In this paper we deal exclusively with *Stochastic ML* estimation, and we will deviate from the usual Gaussian assumption to work with the alternative impulsive model.

3.1 Motivation for a sub-Gaussian model

The demonstration of the impulsiveness of sound signals motivates our work in improving the signal model. Additionally, one of the most important sources of noise in any acoustical environment is the reverberation (while similar effects such as multipath can be observed in other environments). As we are interested in a more accurate model for acoustical signals, we attempt to model both these effects.

The sub-Gaussian processes are attractive in this respect for two main reasons. First, sub-Gaussian processes are impulsive, and hence are able to account for the impulsiveness of the signals. Secondly, the components of a multivariate sub-Gaussian process can not be independent. We suspect this process to be a good model for reverberant noise, which is highly related to the signal itself. As Fig. 6 shows, the noise, which mostly consists of unwanted reverberant signals, can be considered as

jointly sub-Gaussian with the signal, as would be the signals produced from the same Lévy process.

We begin with a theoretical analysis for the SML estimator of a Gaussian signal in the presence of Gaussian noise. This analysis is given as a precursor to the derivation of the sub-Gaussian density and the SML estimation of a signal modeled as a sub-Gaussian random process.

3.2 Framework

We assume a scenario as described on Fig. 5, 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. These transfer functions are

$$a_{r,k} = e^{-i\omega\tau_{r,k}}, \quad r = 1 \dots \rho \quad \text{and} \quad k = 1 \dots \kappa \quad (14)$$

where $\tau_{r,k}$ is the delay of the signal (of source k) received at sensor r relative to the first sensor.

We assume the sources to be in the far-field and hence, $\tau_{r,k} = \tau_r(\theta_k)$, and it is also clear that assuming a linear array

$$\tau_{r,k} = (r - 1) \cdot \tau_1(\theta_k) \quad (15)$$

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

$$\begin{aligned} \underline{\mathbf{a}}_k &= [a_{1,k} \quad a_{2,k} \quad \dots \quad a_{\rho,k}]^T \\ &= [1 \quad e^{-i\omega\tau_{1,k}} \quad e^{-i\omega\tau_{2,k}} \quad \dots \quad e^{-i\omega\tau_{\rho,k}}]^T \end{aligned} \quad (16)$$

The array's input at a single sensor r is

$$x_r(f) = \sum_{k=1}^{\kappa} \underline{\mathbf{a}}_{r,k} \cdot s_k(f) + n_r(f) \quad (17)$$

and therefore the array's input vector is

$$\underline{\mathbf{x}}(f) = \underline{\mathbf{A}} \cdot \underline{\mathbf{s}}(f) + \underline{\mathbf{n}}(f) \quad (18)$$

where

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

3.3 Gaussian Signals

The most commonly used Maximum Likelihood DOA estimator is the Gaussian ML derived either under the assumptions of a deterministic or a stochastic signal. We present in this section the *Stochastic ML* (SML) DOA

estimator for a Gaussian signal in additive white Gaussian noise as background material for the SML DOA estimator to be presented in section 3.4, which is based on sub-Gaussian signals.

Assuming the signals to be jointly stationary Gaussian stochastic processes with covariance matrix $\underline{\underline{\Sigma}} = \text{E}[\underline{\mathbf{s}}(t)\underline{\mathbf{s}}^\dagger(t)] = \text{E}[\underline{\mathbf{s}}(f)\underline{\mathbf{s}}^\dagger(f)]$, and the noise to be uncorrelated white noise of variance σ^2 , we can express the covariance matrix of the received signal as

$$\begin{aligned} \underline{\underline{\mathbf{R}}} &= \text{E}[\underline{\mathbf{x}}(f)\underline{\mathbf{x}}^\dagger(f)] \\ &= \text{E}\left[\left[\underline{\underline{\mathbf{A}}}\underline{\mathbf{s}}(f) + \underline{\mathbf{n}}(f)\right]\left[\underline{\underline{\mathbf{A}}}\underline{\mathbf{s}}(f) + \underline{\mathbf{n}}(f)\right]^\dagger\right] \\ &= \underline{\underline{\mathbf{A}}}\underline{\underline{\Sigma}}\underline{\underline{\mathbf{A}}}^\dagger + \sigma^2\underline{\underline{\mathbf{I}}} \end{aligned} \quad (19)$$

From the assumption that the snapshots are independent and identically distributed, the density function of the complete data set of size M is

$$f(\underline{\underline{\mathbf{X}}}) = \prod_{f=f_1}^{f_M} \frac{1}{\pi^\rho |\underline{\underline{\mathbf{R}}}|} \exp\left(-\underline{\mathbf{x}}^\dagger(f)\underline{\underline{\mathbf{R}}}^{-1}\underline{\mathbf{x}}(f)\right) \quad (20)$$

where

$$\underline{\underline{\mathbf{X}}} = \underline{\mathbf{x}}(f_1), \underline{\mathbf{x}}(f_2), \dots, \underline{\mathbf{x}}(f_M) \quad (21)$$

In order to solve the SML problem, we need to estimate $\hat{\sigma}^2$, $\hat{\underline{\underline{\Sigma}}}$, and $\hat{\underline{\underline{\theta}}}$ by maximizing eq. (20) with respect to these parameters

$$\left[\hat{\sigma}^2, \hat{\underline{\underline{\Sigma}}}, \hat{\underline{\underline{\theta}}}\right] = \arg \max_{\hat{\sigma}^2, \hat{\underline{\underline{\Sigma}}}, \hat{\underline{\underline{\theta}}}} \sum_{f=f_1}^{f_M} \left\{ -\rho \log_e(\pi) - \log_e |\underline{\underline{\mathbf{R}}}| - \underline{\mathbf{x}}^\dagger(f)\underline{\underline{\mathbf{R}}}^{-1}\underline{\mathbf{x}}(f) \right\} \quad (22)$$

Removing constant terms and terms independent of the parameters $\hat{\sigma}^2$, $\hat{\underline{\underline{\Sigma}}}$ and $\hat{\underline{\underline{\theta}}}$, we reach:

$$\left[\hat{\sigma}^2, \hat{\underline{\underline{\Sigma}}}, \hat{\underline{\underline{\theta}}}\right] = \arg \min_{\hat{\sigma}^2, \hat{\underline{\underline{\Sigma}}}, \hat{\underline{\underline{\theta}}}} \sum_{f=f_1}^{f_M} \left\{ \log_e |\underline{\underline{\mathbf{R}}}| + \underline{\mathbf{x}}^\dagger(f)\underline{\underline{\mathbf{R}}}^{-1}\underline{\mathbf{x}}(f) \right\} \quad (23)$$

or

$$\left[\hat{\sigma}^2, \hat{\underline{\underline{\Sigma}}}, \hat{\underline{\underline{\theta}}}\right] = \arg \min_{\hat{\sigma}^2, \hat{\underline{\underline{\Sigma}}}, \hat{\underline{\underline{\theta}}}} \left\{ \log_e |\underline{\underline{\mathbf{R}}}| + \text{Tr} \left[\underline{\underline{\mathbf{R}}}^{-1} \hat{\underline{\underline{\mathbf{R}}}} \right] \right\} \quad (24)$$

where

$$\hat{\underline{\underline{\mathbf{R}}}} = \frac{1}{M} \sum_{f=f_1}^{f_M} \left[\underline{\mathbf{x}}(f)\underline{\mathbf{x}}^\dagger(f) \right] \quad (25)$$

The problem is further investigated in [4] and [16], and numerical methods are developed for the minimization

of the ML function, an introduction of which is given here for completeness.

From eq. (19) we can deduce that:

$$\hat{\underline{\underline{\Sigma}}}(\underline{\underline{\theta}}) = \underline{\underline{\mathbf{A}}}^{-} \left[\hat{\underline{\underline{\mathbf{R}}}} - \hat{\sigma}^2(\underline{\underline{\theta}})\underline{\underline{\mathbf{I}}}\right] \underline{\underline{\mathbf{A}}}^{-\dagger} \quad (26)$$

and [4, 16]

$$\hat{\sigma}^2 = \frac{1}{\rho - \kappa} \text{Tr} \left[\underline{\underline{\mathbf{P}}}_{\underline{\underline{\mathbf{A}}}}^\perp \hat{\underline{\underline{\mathbf{R}}}} \right] \quad (27)$$

where in the above

$$\underline{\underline{\mathbf{A}}}^{-} = \left(\underline{\underline{\mathbf{A}}}^\dagger \underline{\underline{\mathbf{A}}} \right)^{-1} \underline{\underline{\mathbf{A}}}^\dagger \quad (28)$$

is the pseudo-inverse of $\underline{\underline{\mathbf{A}}}$ and $\underline{\underline{\mathbf{P}}}_{\underline{\underline{\mathbf{A}}}}^\perp$ projects into the null space of this pseudo-inverse, *i.e.*,

$$\underline{\underline{\mathbf{P}}}_{\underline{\underline{\mathbf{A}}}}^\perp = \underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{A}}}\underline{\underline{\mathbf{A}}}^{-} \quad (29)$$

From eqs. (24), (26), and (27), we can deduce that:

$$\hat{\underline{\underline{\theta}}}_{\text{ML}} = \arg \min_{\underline{\underline{\theta}}} \log_e \left| \underline{\underline{\mathbf{A}}}\hat{\underline{\underline{\Sigma}}}\underline{\underline{\mathbf{A}}}^\dagger + \hat{\sigma}^2\underline{\underline{\mathbf{I}}}\right| \quad (30)$$

We should reiterate here that $\underline{\underline{\mathbf{A}}}$ above is a function of $\underline{\underline{\theta}}$, although the dependence has been dropped for notational convenience.

Numerical methods have to be employed to solve this optimization problem[28].

3.4 Sub-Gaussian Signals

An alternative to modeling the signal as Gaussian distributed described in the previous section is by employing a Sub-Gaussian random process. This model allows both for the impulsiveness and dependence appearing in audio signals. For this purpose, we can use a distribution of impulsiveness $\alpha = 0.5$, which is completely skewed to the positive axis together with a multivariate Gaussian density. The Lévy (Fig. 8) distribution satisfies exactly these properties (also referred to as a Pareto type 5 distribution with an index of symmetry $\beta = 1$ and characteristic exponent $\alpha = 0.5$). Fig. 7 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.*, $s_k(f) = u_k(f)^{\frac{1}{2}} \cdot \underline{\mathbf{v}}_k(f) = w_k(f) \cdot \underline{\mathbf{v}}_k(f)$
- The resulting signal $s_k(f)$ is transformed through a set of delays $\underline{\mathbf{x}}(f) = \underline{\underline{\mathbf{A}}}\underline{\mathbf{s}}(f)$ to the receiving end of the array
- Any added noise can be modeled as jointly sub-Gaussian

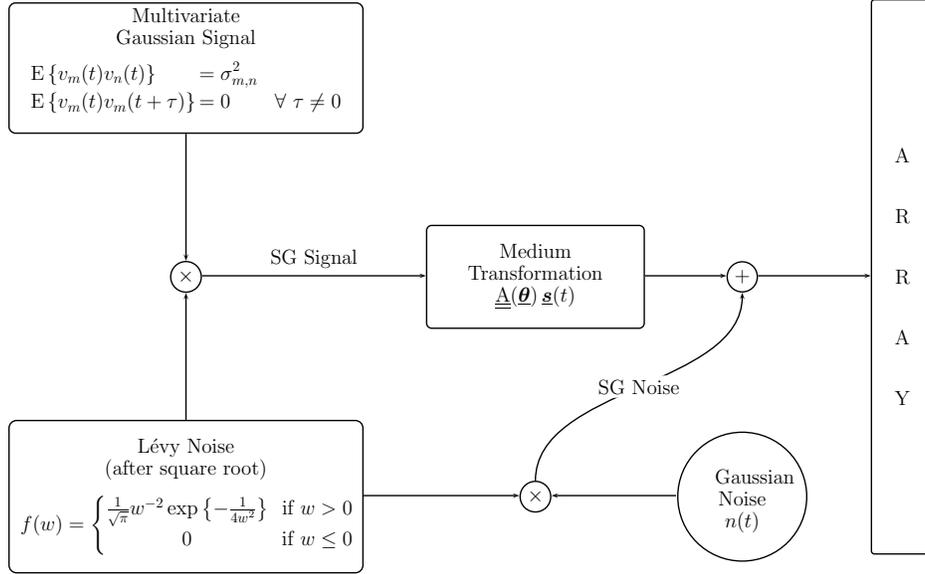


Fig. 7: A multivariate Gaussian signal, corrupted by multiplicative Lévy noise, is then transformed through a set of delays to the receiving end of the array. The addition of white Gaussian noise is also desirable. The noise can be generated from the same Lévy process in order to be jointly sub-Gaussian with the signal.

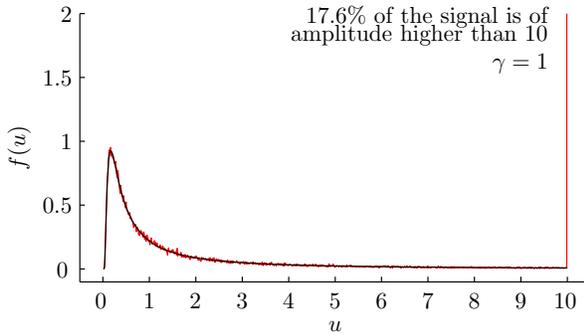


Fig. 8: Lévy distribution along with the histogram of the data generator.

In order to find the distribution of $\underline{s}_k(t)$, we first need the distribution of $w_k(t)$.

From [23]

$$w = g(u) = \sqrt{u} \quad (33)$$

where the dependence has been dropped for convenience. Using the root of this equation, $u_1 = w^2$

$$g'(u_1) = \frac{u_1^{-\frac{1}{2}}}{2} = \frac{w^{-1}}{2} \quad (34)$$

and therefore

$$\begin{aligned} f_w(w) &= \frac{f_u(w^2)}{|\frac{1}{2}w^{-1}|} = 2|w|f_u(w^2) \\ &= |w| \frac{w^{-3} e^{-\frac{1}{4w^2}}}{\sqrt{\pi}} \end{aligned} \quad (35)$$

From eq. (31):

$$f(w) = \begin{cases} \frac{w^{-2} e^{-\frac{1}{4w^2}}}{\sqrt{\pi}} & \text{if } w > 0 \\ 0 & \text{if } w < 0 \end{cases} \quad (36)$$

The distribution of the transmitted signal can now be

3.5 The Sub-Gaussian Density Function

The Gaussian density is similar in form to the one of eq. (20), and the Lévy distribution [32] is given by:

$$f(u) = \begin{cases} \frac{u^{-\frac{3}{2}} e^{-\frac{1}{4u}}}{2\sqrt{\pi}} & \text{if } u > 0 \\ 0 & \text{if } u < 0 \end{cases} \quad (31)$$

So from eq. (13), the signal $\underline{s} = [s_1 \dots s_\kappa]^T$ is of the form

$$s_k(t) = u_k(t)^{\frac{1}{2}} \cdot \underline{\mathbf{v}}_k(t) = w_k(t) \cdot \underline{\mathbf{v}}_k(t) \quad (32)$$

given by the multivariate distribution function:

$$\begin{aligned}
 F(\underline{\mathbf{s}}) &= \int_{w=-\infty}^{+\infty} \int_{\underline{\mathbf{v}}=-\infty}^{\underline{\mathbf{s}}/w} f(w) f(\underline{\mathbf{v}}) d\underline{\mathbf{v}} dw \\
 &= \int_{w=0}^{+\infty} \int_{\underline{\mathbf{v}}=-\infty}^{\underline{\mathbf{s}}/w} \frac{w^{-2} e^{-\frac{1}{4w^2}}}{\sqrt{\pi}} \frac{1}{\pi^\kappa |\underline{\underline{\Sigma}}|} \\
 &\quad \cdot \exp\left(-\underline{\mathbf{v}}^\dagger(t) \underline{\underline{\Sigma}}^{-1} \underline{\mathbf{v}}(t)\right) d\underline{\mathbf{v}} dw \\
 &= \int_{w=0}^{+\infty} \frac{w^{-2} e^{-\frac{1}{4w^2}}}{\sqrt{\pi}} \operatorname{erf}\left[\frac{\underline{\mathbf{s}}}{w}\right] dw \quad (37)
 \end{aligned}$$

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

$$\begin{aligned}
 f(\underline{\mathbf{s}}) &= \frac{d}{ds} \int_{w=0}^{+\infty} \int_{\underline{\mathbf{v}}=-\infty}^{\underline{\mathbf{s}}/w} \frac{w^{-2} e^{-\frac{1}{4w^2}}}{\sqrt{\pi}} \frac{1}{\pi^\kappa |\underline{\underline{\Sigma}}|} \\
 &\quad \cdot \exp\left(-\underline{\mathbf{v}}^\dagger(t) \underline{\underline{\Sigma}}^{-1} \underline{\mathbf{v}}(t)\right) d\underline{\mathbf{v}} dw \\
 &= \int_{w=0}^{+\infty} \frac{d}{dw} \int_{\underline{\mathbf{v}}=-\infty}^{\underline{\mathbf{s}}/w} \frac{w^{-2} e^{-\frac{1}{4w^2}}}{\sqrt{\pi}} \frac{1}{\pi^\kappa |\underline{\underline{\Sigma}}|} \\
 &\quad \cdot \exp\left(-\underline{\mathbf{v}}^\dagger(t) \underline{\underline{\Sigma}}^{-1} \underline{\mathbf{v}}(t)\right) d\underline{\mathbf{v}} \left(\frac{1}{\frac{ds}{d\underline{\mathbf{v}}}}\right) dw \\
 &= \int_{w=0}^{+\infty} \frac{w^{-2} e^{-\frac{1}{4w^2}}}{\sqrt{\pi}} \frac{1}{\pi^\kappa |\underline{\underline{\Sigma}}|} \\
 &\quad \cdot \exp\left(-\underline{\mathbf{s}}^\dagger(t) \underline{\underline{\Sigma}}^{-1} \underline{\mathbf{s}}(t)/w^2\right) \cdot w^{-1} dw \\
 &= \int_{w=0}^{+\infty} C \cdot w^{-3} \cdot \exp\left\{-\frac{1}{w^2} G\right\} dw \\
 &= \frac{C}{2G} \quad (38)
 \end{aligned}$$

where

$$C = \frac{1}{\sqrt{\pi} \pi^\kappa |\underline{\underline{\Sigma}}|} \text{ and } G = \left[1/4 + \underline{\mathbf{s}}^\dagger(t) \underline{\underline{\Sigma}}^{-1} \underline{\mathbf{s}}(t)\right] \quad (39)$$

Therefore

$$f(\underline{\mathbf{s}}) = \frac{1}{2\sqrt{\pi} \pi^\kappa |\underline{\underline{\Sigma}}|} \cdot \left[1/4 + \underline{\mathbf{s}}^\dagger(t) \underline{\underline{\Sigma}}^{-1} \underline{\mathbf{s}}(t)\right]^{-1} \quad (40)$$

Note that if the Gaussian random variable was one di-

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

Table 1: Distributions of interest

mensional and real, then

$$\begin{aligned}
 f(\underline{\mathbf{s}}) &= \frac{1}{2\sqrt{\pi}\sqrt{2\pi}\sigma} \cdot \left[1/4 + \frac{s^2}{2\sigma^2}\right]^{-1} \\
 &= \frac{1}{2\sqrt{2\pi}\sigma} \cdot \left[1/4 + \frac{s^2}{2\sigma^2}\right]^{-1} \quad (41)
 \end{aligned}$$

Plots are shown on Fig. 9 for the one dimensional case.

3.6 Sub-Gaussian based ML

Using the derived density function of the previous section, we can now proceed to derive the ML solution of the array problem as described in Fig. 7.

The received signal $\underline{\mathbf{x}} = [x_1 \dots x_\rho]^\top$ is now of the form:

$$\underline{\mathbf{x}}_r(t) = y(t)^{1/2} \cdot \underline{\mathbf{z}}_r(t) \quad (42)$$

where again, as the transmitted signal, the received signal is sub-Gaussian.

$$\begin{aligned}
 \underline{\mathbf{x}}_r(f) &= y(f)^{1/2} \underline{\mathbf{z}}_r(f) \\
 &= A_1 v(f)^{1/2} \underline{\underline{\mathbf{A}}}_2 \underline{\mathbf{u}}_k(f) \\
 &= v(f)^{1/2} \underline{\underline{\mathbf{A}}}_2 \underline{\mathbf{u}}_k(f) \quad (43)
 \end{aligned}$$

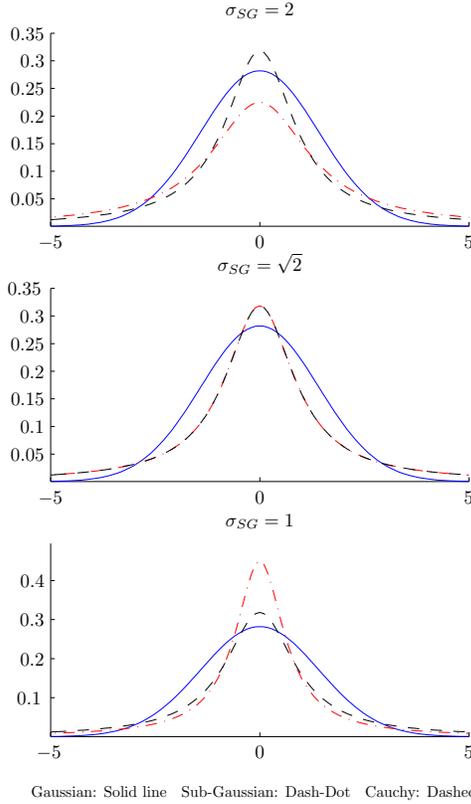


Fig. 9: Sub-Gaussian versus Cauchy and Gaussian distributions. When the dispersion of the underlying Gaussian of the sub-Gaussian process is equal to 1 (i.e., $\gamma_{SG} = 1 \Rightarrow \sigma_{SG} = \sqrt{2}$), the sub-Gaussian is equal in distribution to the normalized Cauchy.

Clearly, there exists a scalar A_1 such that $y = v$.

Without any loss of generality, we can now assume that the linear transformation A_1 on the one-dimensional Lévy distribution can be incorporated in the matrix transformation $\underline{\underline{A}} = A_1 \underline{\underline{A}}_2$. It is therefore straightforward to show that the received signal's correlation matrix will follow eq. (19), but in this case, the characteristics of $\underline{\underline{z}}$ will be relating to those of $\underline{\underline{v}}$ (assuming a noise-free scenario):

$$\begin{aligned}
 \underline{\underline{R}} &= \text{E} \left[\underline{\underline{z}}(f) \underline{\underline{z}}^\dagger(f) \right] \\
 &= \text{E} \left[\left[\underline{\underline{A}} \underline{\underline{v}}(f) \right] \left[\underline{\underline{A}} \underline{\underline{v}}(f) \right]^\dagger \right] \\
 &= \underline{\underline{A}} \underline{\underline{\Sigma}}_v \underline{\underline{A}}^\dagger
 \end{aligned} \tag{44}$$

or in a noisy environment:

$$\underline{\underline{R}} = \underline{\underline{A}} \underline{\underline{\Sigma}}_v \underline{\underline{A}}^\dagger + \sigma_n^2 \underline{\underline{I}}_\rho \tag{45}$$

Therefore, the maximum likelihood estimator is

$$\left[\hat{\underline{\underline{\Sigma}}}, \hat{\underline{\underline{\theta}}} \right] = \arg \max_{\underline{\underline{\Sigma}}, \underline{\underline{\theta}}} \prod_{f=f_1}^{f_M} \frac{2}{\sqrt{\pi} \pi^\rho |\underline{\underline{R}}|} \left[\underline{\underline{x}}^\dagger(f) \underline{\underline{R}}^{-1} \underline{\underline{x}}(f) + 1/4 \right]^{-1} \tag{46}$$

To simplify, take the \log_e

$$\begin{aligned}
 \left[\hat{\underline{\underline{\Sigma}}}, \hat{\underline{\underline{\theta}}} \right] &= \arg \min_{\underline{\underline{\Sigma}}, \underline{\underline{\theta}}} \sum_{f=f_1}^{f_M} \left\{ \log_e |\underline{\underline{R}}| \right. \\
 &\quad \left. + \log_e \left[\underline{\underline{x}}^\dagger(f) \underline{\underline{R}}^{-1} \underline{\underline{x}}(f) + 1/4 \right] \right\}
 \end{aligned} \tag{47}$$

Introducing the original signal statistics:

$$\begin{aligned}
 \left[\hat{\underline{\underline{\Sigma}}}, \hat{\underline{\underline{\theta}}} \right] &= \arg \min_{\underline{\underline{\Sigma}}, \underline{\underline{\theta}}} \sum_{f=f_1}^{f_M} \left\{ \log_e \left| \underline{\underline{A}} \underline{\underline{\Sigma}}_v \underline{\underline{A}}^\dagger + \sigma_n^2 \underline{\underline{I}}_\rho \right| \right. \\
 &\quad \left. + \log_e \left[\underline{\underline{x}}^\dagger(f) \left\{ \underline{\underline{A}} \underline{\underline{\Sigma}}_v \underline{\underline{A}}^\dagger + \sigma_n^2 \underline{\underline{I}}_\rho \right\}^{-1} \underline{\underline{x}}(f) + 1/4 \right] \right\}
 \end{aligned} \tag{48}$$

3.7 Separable Solution

We proceed to reach an alternative minimization function to reduce the search space. To do so we follow the derivations of [16] in which the ML function is first minimized w.r.t. the signal statistics, assuming known DOA: Summation is omitted for the derivations, and thus we define the function to be minimized as:

$$\begin{aligned}
 \mathcal{L} &= \left\{ \log_e \left| \underline{\underline{A}} \underline{\underline{\Sigma}}_v \underline{\underline{A}}^\dagger + \sigma_n^2 \underline{\underline{I}}_\rho \right| \right. \\
 &\quad \left. + \log_e \left[\underline{\underline{x}}^\dagger(f) \left\{ \underline{\underline{A}} \underline{\underline{\Sigma}}_v \underline{\underline{A}}^\dagger + \sigma_n^2 \underline{\underline{I}}_\rho \right\}^{-1} \underline{\underline{x}}(f) + 1/4 \right] \right\} \\
 &= \underbrace{\log_e |\underline{\underline{R}}|}_{\mathcal{L}_1} + \underbrace{\log_e \left[\underline{\underline{x}}^\dagger(f) \underline{\underline{R}}^{-1} \underline{\underline{x}}(f) + 1/4 \right]}_{\mathcal{L}_2}
 \end{aligned} \tag{49}$$

Differentiating:

$$\begin{aligned}
 \frac{\partial \mathcal{L}}{\partial \sigma_{ij}} &= \frac{\partial}{\partial \sigma_{ij}} \left\{ \log_e |\underline{\underline{R}}| + \log_e \left[\underline{\underline{x}}^\dagger(f) \underline{\underline{R}}^{-1} \underline{\underline{x}}(f) + 1/4 \right] \right\} \\
 &= \frac{\partial \log_e |\underline{\underline{R}}|}{\partial \sigma_{ij}} + \frac{\partial \log_e \left[\underline{\underline{x}}^\dagger(f) \underline{\underline{R}}^{-1} \underline{\underline{x}}(f) + 1/4 \right]}{\partial \sigma_{ij}}
 \end{aligned} \tag{50}$$

Working separately on the two terms:

$$\begin{aligned} \frac{\partial \mathcal{L}_1}{\partial \sigma_{ij}} &= \frac{\partial \log_e |\underline{\mathbf{R}}|}{\partial \sigma_{ij}} \\ &= \text{Tr} \left[\left\{ \frac{\partial \log_e |\underline{\mathbf{R}}|}{\partial \underline{\mathbf{R}}} \right\}^T \frac{\partial \underline{\mathbf{R}}}{\partial \sigma_{ij}} \right] \end{aligned}$$

but

$$\frac{\partial \log_e |\underline{\mathbf{R}}|}{\partial \underline{\mathbf{R}}} = [\underline{\mathbf{R}}^{-1}]^T \quad \frac{\partial \underline{\mathbf{R}}}{\partial \sigma_{ij}} = \underline{\mathbf{a}}_i \underline{\mathbf{a}}_j^\dagger \quad (51)$$

Hence

$$\frac{\partial \mathcal{L}_1}{\partial \sigma_{ij}} = \frac{\partial \log_e |\underline{\mathbf{R}}|}{\partial \sigma_{ij}} = \text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{a}}_i \underline{\mathbf{a}}_j^\dagger] = \underline{\mathbf{a}}_j^\dagger \underline{\mathbf{R}}^{-1} \underline{\mathbf{a}}_i \quad (52)$$

Similarly, for the second term:

$$\begin{aligned} \frac{\partial \mathcal{L}_2}{\partial \sigma_{ij}} &= \frac{\partial \log_e [\underline{\mathbf{x}}^\dagger(f) \underline{\mathbf{R}}^{-1} \underline{\mathbf{x}}(f) + 1/4]}{\partial \sigma_{ij}} \\ &= \frac{\partial \log_e [\text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}}] + 1/4]}{\partial \sigma_{ij}} \\ &= \frac{1}{\text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}}] + 1/4} \cdot \frac{\partial}{\partial \sigma_{ij}} \left\{ \text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}}] + 1/4 \right\} \\ &= -\frac{1}{\text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}}] + 1/4} \cdot \text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}} \underline{\mathbf{R}}^{-1} \underline{\mathbf{a}}_i \underline{\mathbf{a}}_j^\dagger] \\ &= -\frac{1}{\text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}}] + 1/4} \cdot \underline{\mathbf{a}}_j^\dagger \underline{\mathbf{R}}^{-1} \underline{\mathbf{C}} \underline{\mathbf{R}}^{-1} \underline{\mathbf{a}}_i \quad (53) \end{aligned}$$

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

Therefore

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \sigma_{ij}} &= \frac{\partial \mathcal{L}_1}{\partial \sigma_{ij}} + \frac{\partial \mathcal{L}_2}{\partial \sigma_{ij}} \\ &= \underline{\mathbf{a}}_j^\dagger \underline{\mathbf{R}}^{-1} \underline{\mathbf{a}}_i - \frac{\underline{\mathbf{a}}_j^\dagger \underline{\mathbf{R}}^{-1} \underline{\mathbf{C}} \underline{\mathbf{R}}^{-1} \underline{\mathbf{a}}_i}{\text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}}] + 1/4} \\ &= \underline{\mathbf{a}}_j^\dagger \left[\underline{\mathbf{R}}^{-1} - \frac{\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}} \underline{\mathbf{R}}^{-1}}{\text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}}] + 1/4} \right] \underline{\mathbf{a}}_i \quad (54) \end{aligned}$$

or in matrix notation, and at the ML value of $\underline{\Sigma}$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \underline{\Sigma}} &= \underline{\mathbf{A}}^\dagger \left[\underline{\mathbf{R}}^{-1} - \frac{\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}} \underline{\mathbf{R}}^{-1}}{\text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}}] + 1/4} \right] \underline{\mathbf{A}}_i \\ &= \underline{\mathbf{A}}^\dagger \underline{\mathbf{R}}^{-1} \left[\underline{\mathbf{R}} - \frac{\underline{\mathbf{C}}}{\text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}}] + 1/4} \right] \underline{\mathbf{R}}^{-1} \underline{\mathbf{A}}_i \\ &= 0 \quad (55) \end{aligned}$$

Using Sherman-Morrison-Woodbury identity:

$$\left(\underline{\mathbf{A}} + \underline{\mathbf{U}} \underline{\mathbf{V}}^\dagger \right)^{-1} = \underline{\mathbf{A}}^{-1} - \underline{\mathbf{A}}^{-1} \underline{\mathbf{U}} \left(\underline{\mathbf{I}} + \underline{\mathbf{V}}^\dagger \underline{\mathbf{A}}^{-1} \underline{\mathbf{U}} \right)^{-1} \underline{\mathbf{V}}^\dagger \underline{\mathbf{A}}^{-1} \quad (56)$$

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

$$\underline{\mathbf{R}} = \underline{\mathbf{A}} \underline{\Sigma} \underline{\mathbf{A}}^\dagger + \sigma_n^2 \underline{\mathbf{I}} \quad (57)$$

Therefore

$$\underline{\mathbf{R}}^{-1} = \frac{1}{\sigma_n^2} \left\{ \underline{\mathbf{I}} - \underline{\mathbf{A}} \left(\underline{\Sigma} \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} + \sigma_n^2 \underline{\mathbf{I}} \right)^{-1} \underline{\Sigma} \underline{\mathbf{A}}^\dagger \right\} \quad (58)$$

and in order to substitute back in eq. (55), we calculate

$$\begin{aligned} \underline{\mathbf{R}}^{-1} \underline{\mathbf{A}} &= \frac{1}{\sigma_n^2} \left\{ \underline{\mathbf{I}} - \underline{\mathbf{A}} \left(\underline{\Sigma} \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} + \sigma_n^2 \underline{\mathbf{I}} \right)^{-1} \underline{\Sigma} \underline{\mathbf{A}}^\dagger \right\} \underline{\mathbf{A}} \\ &= \frac{1}{\sigma_n^2} \left\{ \underline{\mathbf{A}} - \underline{\mathbf{A}} \left(\underline{\Sigma} \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} + \sigma_n^2 \underline{\mathbf{I}} \right)^{-1} \underline{\Sigma} \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} \right\} \\ &= \frac{1}{\sigma_n^2} \underline{\mathbf{A}} \left\{ \underline{\mathbf{I}} - \left(\underline{\Sigma} \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} + \sigma_n^2 \underline{\mathbf{I}} \right)^{-1} \underline{\Sigma} \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} \right\} \\ &= \frac{1}{\sigma_n^2} \underline{\mathbf{A}} \left\{ \underline{\Sigma} \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} + \sigma_n^2 \underline{\mathbf{I}} \right\}^{-1} \\ &\quad \cdot \left\{ \underline{\Sigma} \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} + \sigma_n^2 \underline{\mathbf{I}} - \underline{\Sigma} \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} \right\} \\ &= \underline{\mathbf{A}} \left\{ \underline{\Sigma} \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} + \sigma_n^2 \underline{\mathbf{I}} \right\}^{-1} \quad (59) \end{aligned}$$

Substituting back in eq. (55)

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \underline{\Sigma}} &= \left\{ \underline{\Sigma} \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} + \sigma_n^2 \underline{\mathbf{I}} \right\}^{-1} \underline{\mathbf{A}}^\dagger \left[\underline{\mathbf{R}} \right. \\ &\quad \left. - \frac{\underline{\mathbf{C}}}{\text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}}] + 1/4} \right] \underline{\mathbf{A}} \left\{ \underline{\Sigma} \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} + \sigma_n^2 \underline{\mathbf{I}} \right\}^{-1} \quad (60) \end{aligned}$$

But $\frac{\partial \mathcal{L}}{\partial \underline{\Sigma}} = 0$, therefore

$$\begin{aligned} \underline{\mathbf{A}}^\dagger \left[\underline{\mathbf{A}} \underline{\Sigma}_v \underline{\mathbf{A}}^\dagger + \sigma_n^2 \underline{\mathbf{I}} - \frac{\underline{\mathbf{C}}}{\text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}}] + 1/4} \right] \underline{\mathbf{A}} = \\ \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} \underline{\Sigma}_v \underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} + \underline{\mathbf{A}}^\dagger \sigma_n^2 \underline{\mathbf{A}} - \frac{\underline{\mathbf{A}}^\dagger \underline{\mathbf{C}} \underline{\mathbf{A}}}{\text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{C}}] + 1/4} = 0 \end{aligned} \quad (61)$$

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

$$\begin{aligned} \underline{\Sigma}_{\text{ML}} = \frac{1}{M} \sum_{t=t_1}^{t_M} \left[\left(\underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} \right)^{-1} \underline{\mathbf{A}}^\dagger \left(\frac{\underline{\mathbf{x}} \underline{\mathbf{x}}^\dagger}{\text{Tr} [\underline{\mathbf{R}}^{-1} \underline{\mathbf{x}} \underline{\mathbf{x}}^\dagger] + 1/4} \right. \right. \\ \left. \left. - \sigma_n^2 \right) \underline{\mathbf{A}} \left(\underline{\mathbf{A}}^\dagger \underline{\mathbf{A}} \right)^{-1} \right] \end{aligned} \quad (62)$$

To solve the above, an initial estimate of $\underline{\mathbf{R}}$ can be found from the data using a covariation measure, as will be demonstrated in the next section. However, experience has shown this step not to be necessary, since the recursion converges rapidly from an identity matrix.

The noise variance σ_n^2 can also be found from the same covariation measure assuming the number of sources and sensors are known, similarly to eq. (27).

3.8 DOA Estimation

Assuming that eq. (45) holds, *i.e.*, that signal and noise are jointly sub-Gaussian, we can proceed to estimate the DOA. Using the pseudo-ML approach, the modified ML function can now be expressed as:

$$\begin{aligned} \hat{\underline{\theta}} = \arg \min_{\underline{\theta}} \sum_{f=f_1}^{f_M} \left\{ \log_e |\underline{\mathbf{R}}| + \right. \\ \left. + \log_e \left[\underline{\mathbf{x}}^\dagger(f) \underline{\mathbf{R}}^{-1} \underline{\mathbf{x}}(f) + 1/4 \right] \right\} \end{aligned} \quad (63)$$

where the inverse of $\hat{\underline{\mathbf{R}}}$ can be estimated using the Woodbury identity. It is also significant to note here that the first term $\log_e |\underline{\mathbf{R}}|$ is not a function of the data, and hence can be left out of the minimization process in the case that we are only searching for the angle parameters.

Therefore:

$$[\hat{\underline{\theta}}] = \arg \min_{\underline{\theta}} \sum_{f=f_1}^{f_M} \left\{ \log_e \left[\underline{\mathbf{x}}^\dagger(f) \underline{\mathbf{R}}^{-1} \underline{\mathbf{x}}(f) + 1/4 \right] \right\} \quad (64)$$

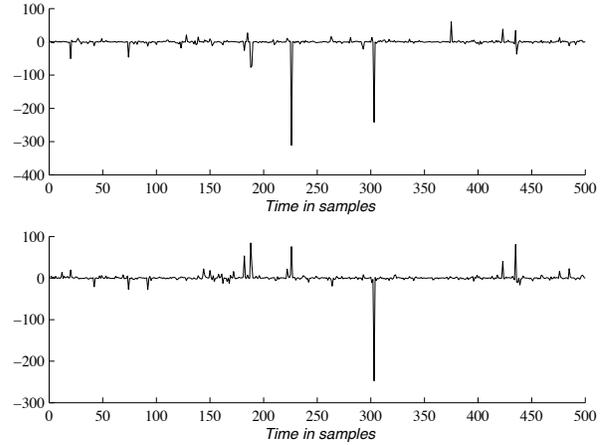


Fig. 10: Sample transmitted (real) signal for a two-source problem.

4 SIMULATIONS

A sample realization based on the derivation assumptions of eq. (40) for a two source problem is shown on Fig. 10. By observing the sample signal the reason why second order statistics fail under this conditions becomes clear. Considering that the dominating data will be the few realizations of high amplitude, the overall second order statistics will fluctuate significantly depending on the statistics of these spikes. For example, while a data block between 250 and 350 will give a normalized auto-correlation of near 1, it will be closer to -1 between 200 and 300.

4.1 Array Spacing

Simulations in this section are performed using a narrowband signal, thus we briefly discuss the intersensor spacing of the array.

We assume that d , the intersensor distance, is equal to $\xi \lambda / 2$, where $\xi \leq 1$. Therefore, from (15)

$$-i\omega\tau = -i\xi\pi \sin \theta$$

and hence

$$\underline{\mathbf{A}} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{-i\xi\pi \sin(\theta_1)} & e^{-i\xi\pi \sin(\theta_2)} & \dots & e^{-i\xi\pi \sin(\theta_\kappa)} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-i\xi\pi \rho \sin(\theta_1)} & e^{-i\xi\pi \rho \sin(\theta_2)} & \dots & e^{-i\xi\pi \rho \sin(\theta_\kappa)} \end{bmatrix} \quad (65)$$

4.2 DOA Estimation

Several sets of simulations need to be performed to test the validity of the algorithm. In each of the following

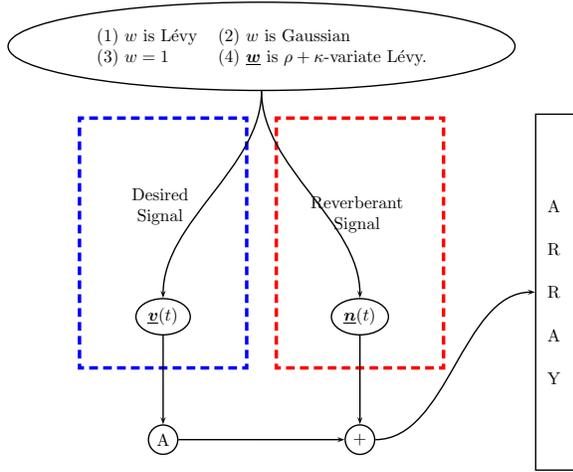


Fig. 11: Explanation of the conditions of the four testing cases.

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 dispersion equal to the dispersion of the Lévy sequence ($\gamma_s = \gamma_u = \gamma_v = 1$). In all cases the impulsiveness was kept constant ($\alpha = 1$ for cases 1 & 4, and $\alpha = 2$ for 2 & 3 as described below). The Generalized Signal-to-Noise Ratio used below is defined as:

$$\text{GSNR} = 10 \log_{10} \left(\frac{\gamma_s}{\gamma_n} \right) = -10 \log_{10} (\gamma_n) \quad (66)$$

In the following four cases, we had random DOA's for 2 sources, 8 sensors, and blocks of 32 samples. The four simulation scenarios are described below:

1. Exactly as per the derivation assumptions (Fig. 12a): received signal is sub-Gaussian, created from a Multivariate Gaussian and a univariate Lévy. Received signal impulsiveness is $\alpha = 1$ (impulsiveness – dependence)
2. The signal is a Multivariate Gaussian (Fig. 12b), and is created from a Multivariate Gaussian ($\underline{\mathbf{v}}$) and a univariate Gaussian (w). Received signal impulsiveness is $\alpha = 2$ (no impulsiveness – dependence)
3. The signal is a Multivariate Gaussian (Fig. 12c) and it undergoes *no* energy fluctuation ($w = 1$, $\underline{\mathbf{v}} = \underline{\mathbf{s}}$). This conforms to the assumptions of the well known Gaussian based ML. Clearly, the received signal

impulsiveness is $\alpha = 2$
(no impulsiveness – no dependence)

4. Finally, the received signal is sub-Gaussian (Fig. 12d), created from a Multivariate Gaussian ($\underline{\mathbf{v}}$) and a *Multivariate Lévy* ($\underline{\mathbf{w}}$). In this case, the signals can be viewed as simply Cauchy. Received signal impulsiveness is $\alpha = 1$ (impulsiveness – no dependence)

Fig. 12a shows the mean squared error for the derivation assumption conditions, where signal and noise are jointly sub-Gaussian. The impulsiveness of the noise variation degrades significantly the performance of the Gaussian based ML, especially at low GSNR's.

We evaluate the performance of the sub-Gaussian based ML by testing the robustness when the process ceases to be impulsive. The sub-Gaussian algorithm performs better than the Gaussian ML (Fig. 12b), even under these conditions.

As expected however, when there is no envelope applied to the signals, *i.e.*, the signals are pure Gaussian, the performance of the Gaussian ML method is slightly better than that of the sub-Gaussian-based ML (Fig. 12c).

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. 12d).

4.3 Estimating Statistics using Covariation

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. We seek to extract Σ from the signal $s = u \underline{\mathbf{v}}$.

With the hypothesis that second order statistics are defined for the above signals, then one could proceed in the usual way of:

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

However, as $\text{E}[u u^*]$ does not exist, one is required to use lower order statistics. We investigate the *Fractional Order Correlation Function* encountered previously in the FLOS-PHAT algorithm and defined as:

$$A_{xy} = \text{E} \{ x^{<p>} y^{*<q>} \} \quad (67)$$

Clearly, in this case:

$$\begin{aligned} \text{E}[(u v_1)^{<p>} (u v_2)^{*<q>}] &= \\ &= \text{E}[u^{<p>} u^{*<q>}] \text{E}[v_1^{<p>} v_2^{*<q>}] \\ &= \text{Const. } \text{E}[v_1^{<p>} v_2^{*<q>}] \quad \forall p < 0.5 \end{aligned}$$

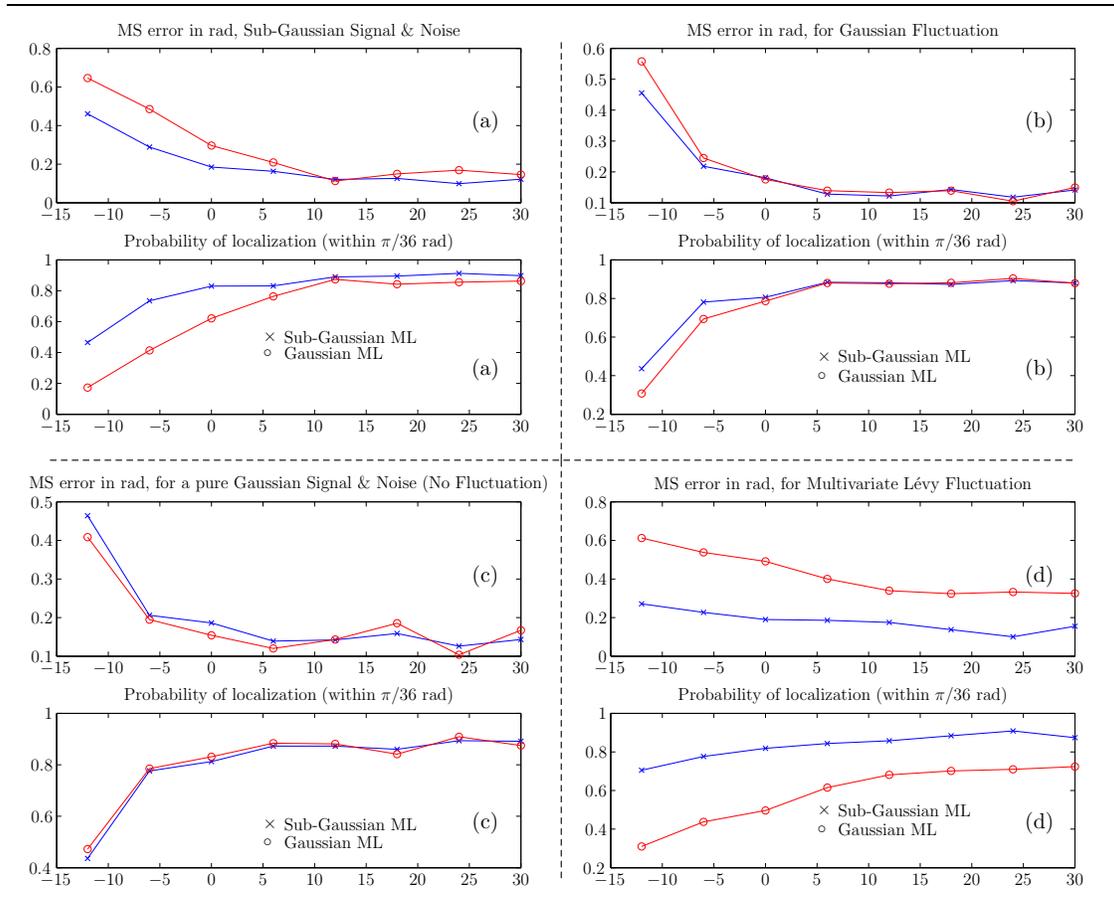


Fig. 12: Simulations demonstrate the obtained benefit in localization by using the Stochastic ML method based on the Lévy Sub-Gaussian processes versus the Gaussian ML method for the conditions described in the text. Robustness of the Sub-Gaussian method is apparent especially in case (d).

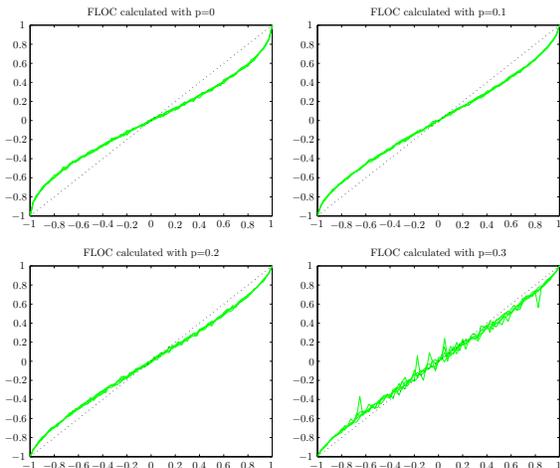


Fig. 13: Estimates of the fractional correlation measure can be deterministically connected to the covariance of the underlying Gaussian densities. The horizontal axis shows the true covariance for the two underlying signals of the sub-Gaussian density, while the vertical axis shows the FLOS estimate. 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 expected.

The plots of Fig. 13 demonstrate the connection between the FLOS statistics and the second order statistics of the Gaussian part of the signal. There exists a deterministic correspondence when $2p$ is lower than $\alpha_u = 0.5$. Similar results can be obtained for complex signals. A generated lookup table can, if required, provide an estimate of the underlying Gaussian statistics.

4.4 Estimating Statistics using ML

As mentioned earlier, random initial conditions are sufficient for the solution of eq. (62). Fig. 14 shows the estimates for a 3-source problem with

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

when the initialization vectors are the identity matrix.

The sample statistics are slightly different from the above depending on the length of the realization, and are plotted on Fig. 14 as well. The histogram plots show on the positive side the sample statistics, and on the negative side the estimates of the diagonal elements of $\underline{\underline{\Sigma}}$ as estimated by eq. (62). The insignificantly small complex components of the diagonal of $\underline{\underline{\Sigma}}$ are ignored due to prior knowledge. The scatter plots present the off-diagonal elements of the statistics matrix on an Argand diagram.

The dots denote the actual sample statistics, while the estimates are shown with ‘x’.

As can be observed from Fig. 14, the number of sensors is far more important than the total number of samples. As an example, we can see 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, a significant decrease of SNR in (c) can be compensated by an increase in the number of sensors. Likewise, we can 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 sources dramatically improves the accuracy in (f).

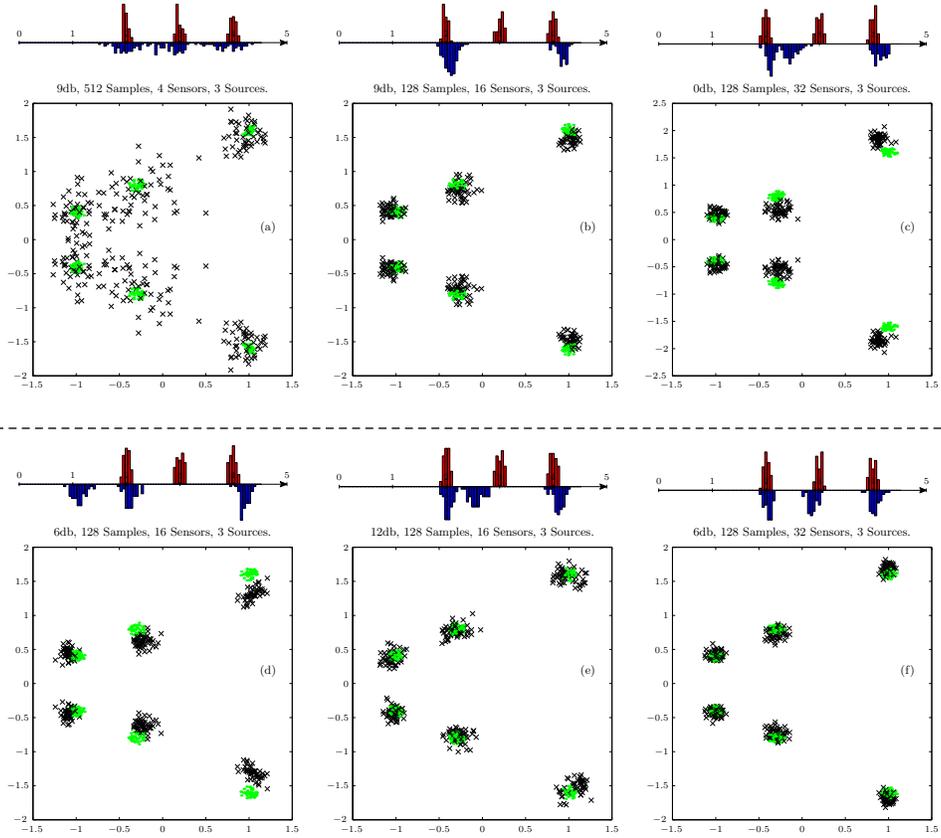


Fig. 14: Simulations show the effectiveness of the separable ML estimation of statistics for a 3-source problem under various noise conditions and array arrangements.

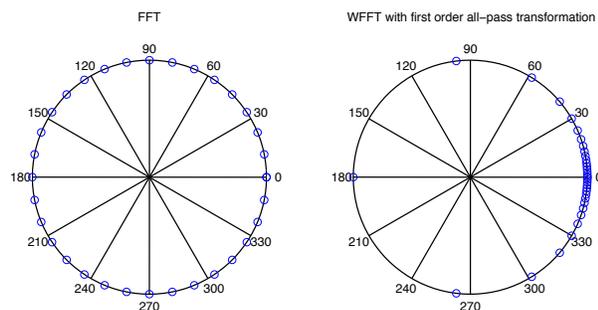


Fig. 15: Sample frequency transformation of a 1st order all-pass filter

5 SUB-GAUSSIAN AND GAUSSIAN ML LOCALIZATION COMPARISONS ON REAL DATA

In order to test the localization algorithm with some real data, we constructed two synthetic microphone arrays: using the 10.2 channel system and ProTools we played back several (dry) signals (Trumpet, Cello, a female voice in English, and a female voice in Danish). These audio channels were played together in various combinations through the loudspeakers at 48kHz, and 2 microphones were shifted forming a linear array. The synchronized playback-recording feature of ProTools, confirmed by the addition of chirp synchronization signals at the start of the recording, ensured that the array was accurately created.

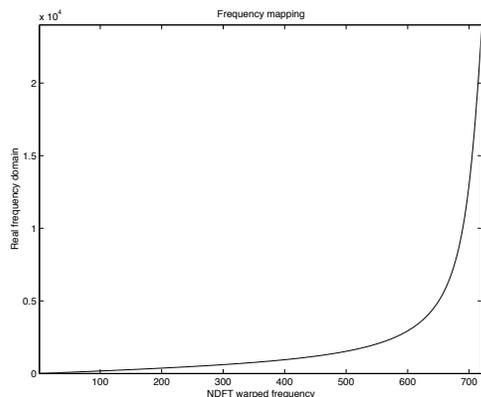


Fig. 16: Actual frequency mapping used in the real signal ML localization experiments

The ML function for the following cases was evaluated over all frequencies, by re-calculating the transformation matrix $\underline{\underline{A}}$ for all possible $(\underline{\theta}, f)$ combinations, which is a computationally expensive process. For the localization part, a *Non-linear FFT* (NFFT) [2, 19, 22] was used

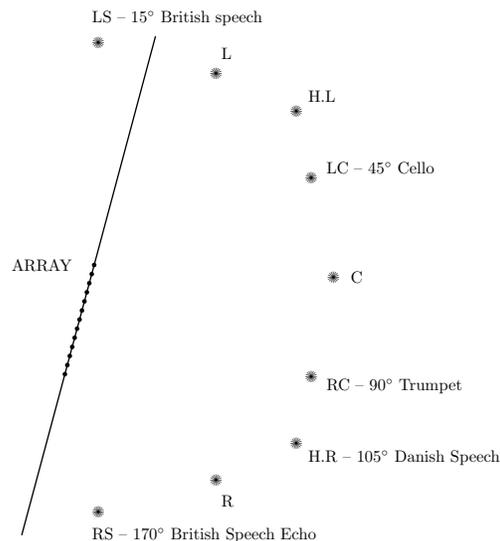


Fig. 17: Experiment setup for the 20-microphone array. Angles shown are relative to the center of the array arrangement. Only the Cello and Trumpet sources could be reliably localized due to the inaccuracies in sensor placement with the 20-microphone array.

in order to keep the resulting frequency domain signals. Specifically, we employed the method described by Mitra *et al* in [19], with a first order all-pass filter and a 30ms window (1440 samples). The resulting frequency mapping is shown on Fig. 16, while a more visual representation of the first-order mapping is shown on Fig. 15 with fewer taps.

5.1 20-Microphone Array

In the 20-microphone array case, the aperture was 38cm and the intersensor spacing was 2cm, while 4 (originally dry) signals (Trumpet, Cello, a female voice in English, and a female voice in Danish) and an artificial echo of the cello were used. These 5 channels were played together in various combinations, although the results shown here are based on localization of the sources when two signals were active (the Cello and Trumpet at 45° and 90° respectively). This array was not very accurately spaced and the error rate from the part where all 5 channels were active was very large. The array setup is shown on Fig. 17.

Results of localization demonstrate that the sub-Gaussian based ML method performs significantly better than its Gaussian counterpart. Fig. 18 shows 7s of the signal where only the cello and trumpet are being played. Each frame of the segment corresponds to a sliding win-

	Gaussian	Sub-Gaussian
45° angle RMS error	16	5
90° angle RMS error	22.5	10
Overall RMS error	19.5	8

Table 2: Errors for the Gaussian based ML method are more than double those of the sub-Gaussian based ML.

dow of 30ms, and the sources were placed at 45° and at 90°. As can be observed, the sub-Gaussian ML method works significantly better. Table 2 shows the RMS error for this localization experiment, and reveals that the performance of the Gaussian based ML is significantly worse than that of the sub-Gaussian based ML.

5.2 41-Microphone Array

In the 41-microphone array the recording conditions are similar to the previous case. However, the inter-sensor spacing is 1cm, the array is much more accurately spaced than the previous one, and the sources are the two speech signals used in the previous section placed at 48° and 110°. In addition, the arrangement is such that a strong echo is created at 90°. Fig. 19 shows the positions of the sources, the array, and the flat screen,² which as we expect causes a strong sound reflection.

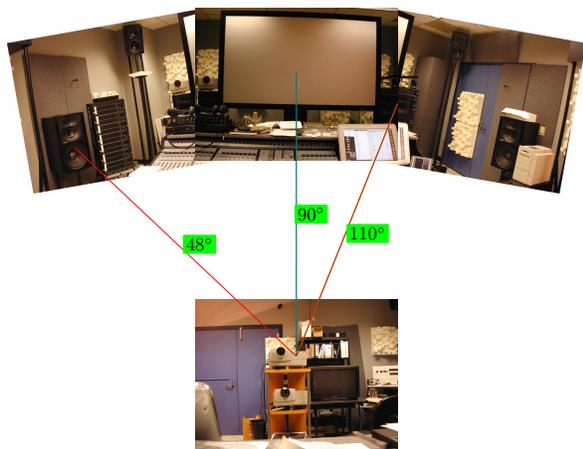


Fig. 19: Arrangement of 41-microphone array.

We note that, in addition to a superior performance of the sub-Gaussian based ML, the errors of the sub-Gaussian tend to be more reasonable. In other words, the sub-

²The screen is made from a synthetic material that is highly reflective.

	Gaussian	Sub-Gaussian
48° angle RMS error	11.1	9.3
90° angle RMS error	13.1	6.9
110° angle RMS error	17.7	6.6
Overall RMS error	24.6	13.3

Table 3: Errors for the Gaussian based ML method are much higher than those of the sub-Gaussian based ML, but compare better under these conditions of the larger array than in the case of the 20-microphone array.

Gaussian algorithm incorrectly localizes sources mostly in the range 50°-90°, which we believe corresponds to the reflections off the console, while the Gaussian based ML is severely influenced by the noise impulsiveness and locates sources more indiscriminately. Nevertheless, the performance difference decreases as the array size grows, a similar conclusion with the performance difference gap narrowing at increasing SNR's in the simulations. The RMS error of localization for the two methods is shown on Table 3.

6 SIGNAL RECOVERY

The accurate extraction of sound from specific locations in the room is a possible application of a large array. We attempt to reconstruct the original English speech signal from the 41-microphone array with the simplest of methods to demonstrate the concept.

An overview of the signal reconstruction process is given on Fig. 21. First, the time aligned signal blocks are transformed in the frequency domain. The transformation into the frequency domain is the same as the one for the ML localization, and as such, it will add no additional computational expense. Vectors equal to the array size are constructed from each frequency and filtered through the frequency dependent steering vectors \underline{w} . The resulting collection of $\underline{w}(f)^T \underline{S}(f)$ coefficients forms the frequency domain of the required signal. However, as inverse FFT is not sufficient to recover the sound since it causes clicks in the signal, a phase and amplitude corrected sinusoid is created instead from each frequency coefficient. The resulting sinusoid is windowed to smooth the transition and minimize the clicks in the final signal. Finally, the collection of overlapping sinusoids over all frequencies are added to produce the sound signal estimate. For improved signal quality, interpolation in the frequency domain (both in amplitude and phase) can be used before the creation of the sinusoids to produce a better signal recovery.

Much research has been performed in estimating a signal

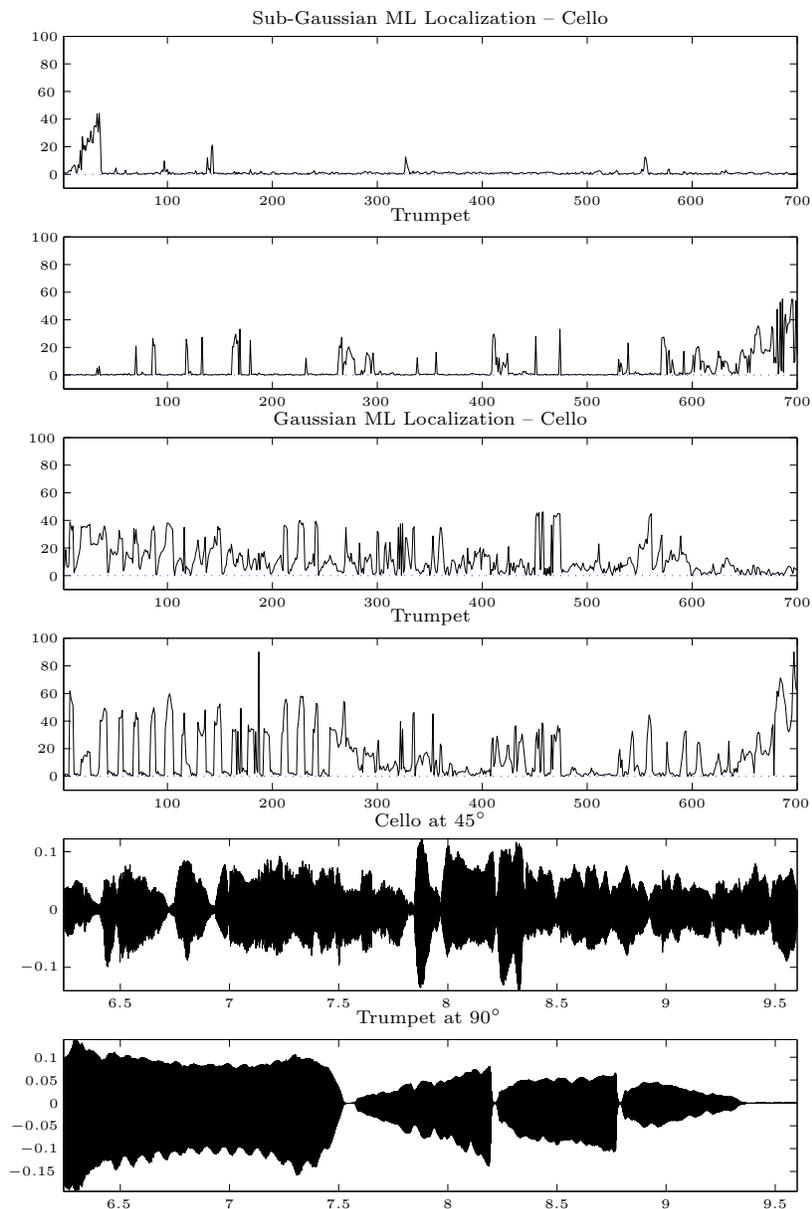


Fig. 18: The audio array ML DOA comparison for the localization of two sources at 45° and at 90° using the sub-Gaussian and Gaussian ML based methods. The Gaussian based ML appears to suffer significantly from reverberation effects. The two original sound signals are plotted at the bottom two graphs, and we can see the correlation of the error rising when the amplitude of the trumpet dies off at the end. Note that, although not visible on the graph, there is actually a low amplitude signal until the very end of the recording.

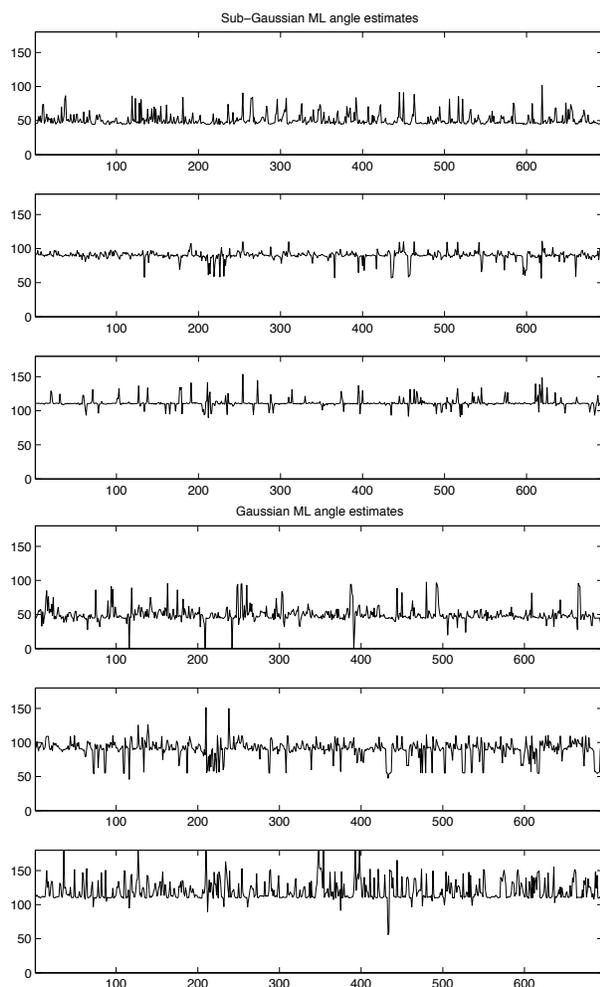


Fig. 20: Angle estimates for Gaussian and sub-Gaussian based ML methods for the 41-microphone setup.

from its short-term Fourier transform, mostly due to the need for good compression algorithms such as MPEG. Since compression is not an issue in our case, we used the simple sinusoidal addition with a reconstruction window identical to the window used in the time domain (Hanning window). A similar algorithm is described by Griffin and Lim in [13]. Further analysis on the subject is given in [24, 26], however this material is beyond the scope of this paper.

The signal recovery process was successful to a certain extent, although the array size was restrictive. Traditional design of the focusing vector w is based on the concept that a signal can be classified as source, interference and noise. However, in an environment where reverbera-

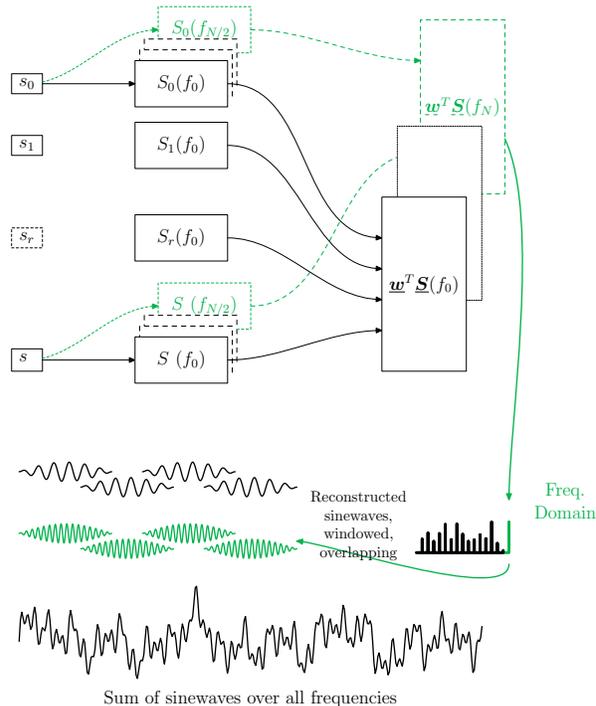


Fig. 21: Signal reconstruction process.

tion is both the noise and interference, the “noise” level is extremely high. Nevertheless, we have used a constrained algorithm that minimizes the overall response at high frequencies, and places additional constraints at lower frequencies (at 90° and 110°). These correspond to the *Minimum Variance Distortionless Response* (MVDR) algorithm and the *Linearly Constrained Minimum Variance* (LCMV) [5, 6, 9, 10], respectively. Again, these methods are based on second order measures, and better algorithms could be used to allow for impulsive signals.

However, automated design fails for both methods as they are designed for placing specific restrictions at specific locations. In addition, the constraints often lead to weight vectors of extreme amplitude range, which cause amplification of microphone placement errors. Work in literature such as [14, 15, 20], attempts to tackle these issues.

The transformation at the frequency domain was attempted with both the linear and non-linear FFT’s. In the case that we use the NFFT method, the resulting signal’s frequency content must be weighted with the in-

verse density of the frequency mapping, *i.e.*,

$$\text{Emphasis} = \frac{\delta f_{out}}{\max(\delta f_{out})} \quad (68)$$

Fig. 22 and Fig. 23 show the obtained beampattern using the MVDR method and a non-linear frequency transformation. Although the constraints are sufficient for a large part of the frequency range, they can be too restrictive at certain frequencies, or not restrictive enough at other frequencies. As an example, the constrain of $C = 1$ at 48° will cause a very narrow beampattern at high frequencies, while a plateau constrain in a range $48^\circ \pm 5^\circ$ will cause extreme ripples; nevertheless, the same problem does not appear at low frequencies.

In the design of the coefficients of Fig. 22 and Fig. 23, we did not place any constraints on the amplitude of the coefficients; however, by using only two constraints (zero at 110° and unity at 48° , and vice versa), the resulting set of vectors w was well behaved.

The smaller amplitude variation on the weights clearly produces better quality sounds at the cost of poor source separation. In fact, attempts to use significantly constrained beampatterns gave a high pitch noise that made the resulting signal incomprehensible, while the ones of Fig. 22 and Fig. 23 gave a better quality signal but with far poorer separation. The common assumptions in the literature that the interference source is white noise fail to allow for these kinds of problems, which we encounter with real signals. Clearly, we need to further explore methods of automatic creation of the weight space for more constraints.

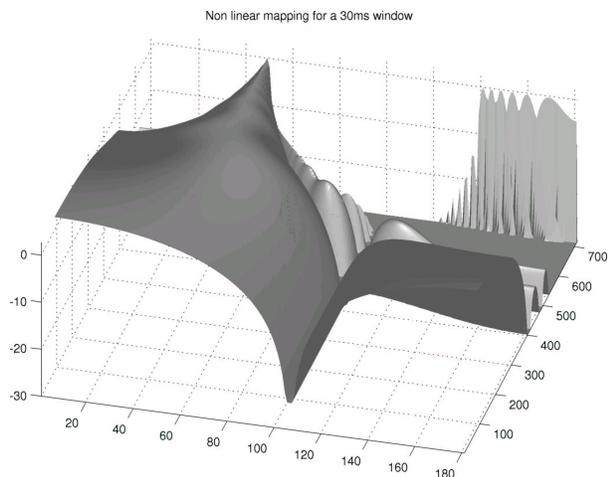


Fig. 22: Beampattern for recovery of the 48° source (English speech) using the NFFT mapping.

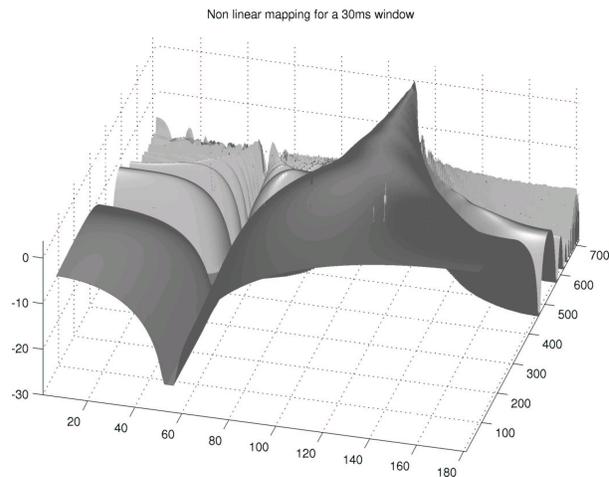


Fig. 23: Beampattern for recovery of the 110° source (Danish speech) using the NFFT mapping.

Acoustically, the best separation occurred with the NFFT approach using emphasis as in eq. (68), and with a block size of 30ms. However, with the Danish speech source being much louder originally and much more reverberant,³ the result in recovering only the English speech is poor. Nevertheless, we were able to achieve a pretty good separation in the case of removing the English speech from the Danish one. Rough estimates from periods of silence in one of the two signals suggest a minimum of 10dB attenuation of the English speech. Much better separation can be achieved with the use of non-linearly spaced arrays and 2-D arrangements such as in a cross pattern.

7 CONCLUSIONS

We have presented in this work a model designed to account for signals that are dependent and impulsive in nature. Such signals are often encountered in many disciplines including audio. Our present research was motivated by existing work demonstrating the impulsiveness of sound and by the observation that reverberation is highly dependent on the original source.

The ML solution of this model was given under a sensor array scenario, and its separable solution was derived. The separable solution assumes known statistics to localize the directions-of-arrival and known directions-of-arrival to find the statistics of the underlying processes.

³The console is located between the 110° loudspeaker and the array, and the Danish speech is still much louder than the English one at the receiving sensors. This suggests that the Danish reverberant component is comparable in intensity to the direct component of the English speech.

Although the statistics estimator could not be derived as a closed form expression, the resulting form allows for a fast iterative solution. The directions-of-arrival estimator still requires a search, but of a much smaller space.

Simulations have demonstrated the robustness of the sub-Gaussian based ML, and encourage us to further develop methods employing the sub-Gaussian, rather than the Gaussian, model. Additionally, the performance loss of the sub-Gaussian based ML in the case that signals are Gaussian is insignificant, which further enforces our robustness claim.

Real world measurements were conducted with two large arrays (20 and 41 microphones) in our audio lab, a room with the acoustics of a typical living room. These experiments have also supported the advantages of the new model. The sub-Gaussian based ML exhibits an improvement in localization up to a factor of 3 in the RMS error versus the Gaussian ML.

ACKNOWLEDGEMENTS

This research was funded in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement No. EEC-9529152. The project was also sponsored in part by the U.S. Army. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

8 REFERENCES

- [1] R. Adler, R. E. Feldman, and M. S. Taqqu, editors. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhäuser Editions, 1998.
- [2] S. Bagchi and S. K. Mitra. *Nonuniform Discrete Fourier Transform and its Signal Processing Applications*. Kluwer, Norwell, MA, 1999.
- [3] O. Besson, F. Vincent, P. Stoica, and A. B. Gershman. Approximate maximum likelihood estimators for array processing in multiplicative noise environments. *IEEE Trans. Signal Processing*, 48(9):2506–2518, 2000.
- [4] 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.*, volume 4, pages 2819–22, 1986.
- [5] K. M. Buckley and L. J. Griffiths. An adaptive generalized sidelobe canceller with derivative constraints. *IEEE Trans. Antennas Prop.*, 34:311–319, Mar. 1986.
- [6] K. M. Buckley and L. J. Griffiths. Linearly-constrained beamforming: A relationship between phase center location and beampattern. In *Conference Record of The Nineteenth Asilomar Conference on Signals, Systems and Computers*, pages 234–238, Pacific Grove, CA, 1986.
- [7] S. Cambanis and G. Miller. Linear problems in p^{th} order and stable processes. *SIAM Journal on Applied Mathematics*, 41(1):43–69, August 1981.
- [8] Stamatis Cambanis, Gennady Samorodnitsky, and Murad S. Taqqu, editors. *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.
- [9] M. H. Er and A. Cantoni. A new approach to the design of broad-band element space antenna array processors. *IEEE J. Oceanic Eng.*, 10:231–240, July 1985.
- [10] O. L. Frost III. An algorithm for linearly constrained adaptive antenna array processing. In *Proc. IEEE*, volume 60, pages 926–935, Aug. 1972.
- [11] P. G. Georgiou, P. Tsakalides, and C. Kyriakakis. Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise. *IEEE Transactions on Multimedia*, 1(3):291–301, September 1999.
- [12] A. B. Gershman, C. F. Mecklenbrauker, and J. F. Bohme. Matrix fitting approach to direction of arrival estimation with imperfect spatial coherence of wavefronts. *IEEE Trans. Signal Processing*, 45(7):1894–1899, 1997.
- [13] D. W. Griffin and J. S. Lim. Signal estimation from modified short-time fourier transform. *IEEE Trans. Acoust., Speech, and Signal Process.*, 32(2):236–43, Apr. 1984.
- [14] O. Hoshuyama and A. Sugiyama. An adaptive microphone array with good sound quality using auxiliary fixed beamformers and its DSP implementation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing.*, volume 2, pages 949–952, 1999.
- [15] O. Hoshuyama, A. Sugiyama, and A. Hirano. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive. *IEEE Trans. Signal Processing*, 47(10):2677–2684, 1999.
- [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.*, volume 5, pages 2893–2896, 1988.
- [17] Don H. Johnson and Dan E. Dudgeon. *Array signal processing: concepts and techniques*. Signal Processing Series. Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [18] P. Kidmose. Alpha-stable distributions in signal processing of audio signals. In *41st Conference on Simulation and Modelling, SIMS2000*, pages 87–94. Scandinavian Simulation Society, 2000.
- [19] A. Makur and S. K. Mitra. Warped discrete-fourier transform: theory and applications. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 48(9):1086–1093, Sep. 2001.
- [20] W. H. Neo and B. Farhang-Boroujeny. Robust microphone arrays using subband adaptive filters. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing.*, volume 6, pages 3721–3724, 2001.
- [21] C. L. Nikias and M. Shao. *Signal Processing with Alpha-Stable Distributions and Applications*. John Wiley and Sons, New York, 1995.
- [22] A. Oppenheim and D. Johnson. Computation of spectra with unequal resolution using the fast fourier transform. In *Proc. IEEE*, volume 59, pages 299–301, 1971.
- [23] Athanasios Papoulis. *Probability, random variables, and stochastic processes*. McGraw-Hill, New York, 3rd edition, 1991.
- [24] R. J. McAulay T. F. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans. Acoust., Speech, and Signal Process.*, 34(4):744–54, Aug. 1986.
- [25] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York - London, 1994.
- [26] Xavier Serra and Julius Smith III. Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4), 1990.
- [27] M. Shao and C. L. Nikias. Signal processing with fractional lower order moments: Stable processes and their applications. *Proc. of the IEEE*, 81(7):986–1010, July 1993.
- [28] D. Starer and A. Nehorai. Newton algorithms for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise. *IEEE Trans. Signal Processing*, 40(6):1528–1534, June 1992.
- [29] H. Stark and J. W. Woods. *Probability, Random Processes and Estimation Theory for Engineers*. Prentice Hall, Englewood Cliffs, New Jersey, 2nd edition, 1994.
- [30] P. Stoica, O. Besson, and A. B. Gershman. Direction-of-arrival estimation of an amplitude-distorted wavefront. *IEEE Trans. Signal Processing*, 49(2):269–276, 2001.
- [31] P. Tsakalides, R. Raspanti, and C. L. Nikias. Joint target angle and doppler estimation in stable impulsive interference. *IEEE Trans. Aerosp. Electron. Syst.*, 32, 1999.
- [32] V. M. Zolotarev. *One-dimensional stable distributions*, volume 65 of *Translations of mathematical monographs*. American Mathematical Society, Providence, R.I., 1986.