Separation of Alpha-Stable Random Vectors[☆]

Mathieu Fontaine

Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

Roland Badeau

LTCI, Télécom Paris, Institut Polytechnique de Paris, Paris, France

Antoine Liutkus

Inria, LIRMM, University of Montpellier, France

Abstract

Source separation aims at decomposing a vector into additive components. This is often done by first estimating source parameters before feeding them into a *filtering method*, often based on ratios of covariances. The whole pipeline is traditionally rooted in some probabilistic framework providing both the likelihood for parameter estimation and the separation method. While Gaussians are ubiquitous for this purpose, many studies showed the benefit of heavy-tailed models for estimation. However, there is no counterpart filtering method to date exploiting such formalism, so that related studies revert to covariance-based filtering after estimation is finished.

Here, we introduce a new multivariate separation technique, that fully exploits the flexibility of α-stable heavy-tailed distributions. We show how a *spatial* representation can be exploited, which decomposes the observation as an infinite sum of contributions originating from all directions. Two methods for separation are derived. The first one is non-linear and similar to a beamforming technique, while the second one is linear, but minimizes a covariation criterion, which is the counterpart of the covariance for α-stable vectors. We evaluate the proposed techniques in a large number of challenging and adverse situations on synthetic experiments, demonstrating their performance for the extraction of signals from strong interferences.

Keywords: alpha-stable distribution, separation theory, additive models, measure theory, optimization

[☆]This work was partly supported by the research programmes KAMoulox (ANR-15-CE38-0003-01) and EDiSon3D (ANR-13-CORD-0008-01) funded by ANR, the French State agency for research.

Email addresses: fontaine.mathieu2@gmail.com (Mathieu Fontaine), roland.badeau@telecom-paris.fr (Roland Badeau), antoine.liutkus@inria.fr (Antoine Liutkus)

1. Introduction

Source separation is the task that consists in decomposing a signal into additive components. It is a very active research area in signal processing, notably because of its numerous applications. In audio for instance, it is the natural paradigm for denoising (Godsill and Rayner [17], Godsill et al. [16], Fontaine et al. [14]) or for the demixing of music and speech recordings (Rafii et al. [33], Vincent et al. [39]). It also finds applications in image processing and biological signal processing (Damon et al. [10], Cavalcant et al. [8]), to name just a few (Comon and Jutten [9]).

Although there was some successful recent research on end-to-end methods that directly produce source estimates when fed with a mixture (Wang et al. [42], Venkataramani et al. [38]), the most common separation processing pipeline consists in two steps done sequentially. First, the mixture is fed into some *estimation* system. This results in a set of parameters that are used in a second step to design a source-specific *filter,* which is applied to the mixture to produce estimates. Formally, this filtering operation boils down to computations performed on many mixture vectors in an independent manner. For instance, the observed (multivariate) mixture samples may directly be assumed independent as routinely done in biological signal processing. In other cases like audio where temporal dependencies cannot be neglected, it is common to apply some Time-Frequency (TF) analysis first and then to assume independence in this transformed domain (Benaroya et al. [4], Duong et al. [11], Liutkus et al. [22]).

Separating an observed vector into additive components requires further assumptions that are usually encoded into a probabilistic model permitting inference. More precisely, the required feature of such a model is to allow derivation of the posterior distribution of one source given the observation of the mixture and the knowledge of the model parameters. An ubiquitous example of such a model is the Gaussian case, for which each source is described through a covariance matrix, and separation is easily performed in an analytical form. It turns out that such a model subsumes the popular, yet degenerate case where each source lies in a linear subspace, corresponding to its *direction of arrival* (Duong et al. [11]). In any case, this whole linear estimation theory enjoys a rich history whose roots can be traced back to the work of N. Wiener in the 1940s (Wiener [43]). From a broader perspective, we see that the core challenge faced by most source separation methods is strongly related to additive probabilistic models (Duvenaud et al. [12], Febrero-Bande and González-Manteiga [13], Wood et al. [44], Marra and Wood [25]). The particular twist in this respect is that the models chosen for separation should provide a way to recover the additive sources.

Although covariance modeling for source separation has enjoyed a strong popularity due to the simplicity and effectiveness of the separation procedure, experience shows that it also suffers from some weaknesses. First and foremost,

Gaussian processes realizations may not explore more than a few standard deviations, which means that Bayesian inference in these models is intrinsically very sensitive to initialization, since the probability mass is almost everywhere negligible. A common workaround is to further constrain covariance models through shallow (Ozerov et al. [31]) or deep (Nugraha et al. [30]) parametric constraints, but another complementary route is to simply opt for heavy-tail models, for which much more robust inference is possible. For instance, multivariate Laplace filters (Wang et al. [41]) were successfully used for robust detection and result from Bayesian inference in a state-space model where some variables are Laplace distributed. In the same vein, a Student's t filter (Roth et al. [34]) was also proposed for a tracking scenario and exploits the heavy-tailed nature of the Student's t distribution. Likewise, this distribution was also already considered for robust estimation of source separation parameters (Kitamura et al. [20], Yoshii et al. [45]).

Although the Laplace and Student's t distributions mentioned above are characterized as featuring heavy tails and are thus suitable for robust estimation of sources parameters, their density is not stable under convolution, which means that the distribution of sums of such random variables does not belong to the same family. As a consequence, they do not straightforwardly lead to convenient filters that may be used for the separation stage. In this context, a natural solution is to take the target signal as deterministic, and only pick a heavy-tailed distribution for the noise term. However, such an approach breaks down when uncertainty is to be considered for the target, that becomes stochastic, or when more than two components are mixed, which is for instance common in source separation. Consequently, the strategy employed, e.g. in (Kitamura et al. [20], Yoshii et al. [45]) is to use robust models for estimation only, and then revert to a covariance separation approach. The only principled separation method we are aware of, that is based on heavy tailed modeling, is the α-Wiener filter presented in (Liutkus and Badeau [21]) and further developed in (Fontaine et al. [14]). It is based on α-stable distributions but is however restricted to the scalar case.

In this paper, we build on the scalar α-Wiener filter (Liutkus and Badeau [21]) to extend it to the multivariate case and hence propose for the first time a filter based on multivariate α-stable distributions. Doing so, we enable the use of such heavy-tail models not only for parameter estimation, but also for separation. The α-stable distributions and processes (Samoradnitsky and Taqqu [35]) are defined as the only class of distributions that are stable under convolution, and thus under addition of independent realizations. They hence naturally appear in the generalized version of the central limit theorem, which is applicable even when the random variables under consideration do not have finite moments. This is useful for modeling very volatile noise or signals, or to allow iterative parameter estimation strategies even with bad initialization. They were first put forward by Mandelbrot to model financial time series (Mandelbrot [24]) and have found widespread applications ever since: impulse noise modeling in landline connections (Stuck and Kleiner [36]), modeling of background speckle patterns in SAR images (Achim et al. [1]), and audio noise modeling (Bassiou

et al. [2]), to name a few.

The dependencies between the entries of α-stable random vectors are not encoded in covariance matrices as in the Gaussian case. Instead, they are described through a unique measure defined on the hypersphere (Hardin J. [18]), which can be understood as providing the strength of each *direction of arrival*. This makes the expressive power of the model much larger than in the Gaussian case, that is limited to ellipsoidal profiles. Throughout this paper, we call this object the *spatial density*[1]. That object regularly attracts some attention. It has for instance been considered for independent component analysis (Kidmose [19], Wang et al. [40]), for exchange rates estimation in financial data (Nolan et al. [29]), and more recently for audio source localization (Fontaine et al. [15]). As a mathematical object defined on the hypersphere, some authors also proposed alternative equivalent representations, notably through spherical harmonics (Pivato and Seco [32]).

In this paper, we show how to design digital filters specifically for α-stable random vectors. For this purpose, we go further than both (Kidmose [19]), that focused on linear determined mixtures, and (Liutkus and Badeau [21]), which is limited to the scalar case. First, we present some theoretical results in Section 2, that develop a *spatial spectrum* representation for α-stable random vectors. Based on this representation, we propose two different filtering methods. In Section 3, multivariate observations are decomposed into their spatial spectrum, whose components are then filtered individually for reconstruction. An alternative approach is presented in Section 4, where the sources are estimated directly as a combination of linear filters, but with a design involving the spatial density. Both methods are evaluated in Section 5 and compared to their Gaussian counterpart.

Notation

Throughout the paper, scalars are denoted with a normal, light font, e.g. $\alpha \in (0, 2)$ or $K \in \mathbb{N}$. Vectors are indicated with bold lowercase letters, e.g. $\boldsymbol{x} \in \mathbb{C}^K$ and matrices with bold uppercase letters, e.g. $\boldsymbol{P} \in \mathbb{C}^{K \times K}$. We furthermore use the following notation:

- $\mathbb{K}$: denotes either $\mathbb{C}$ or $\mathbb{R}$.
- $\mathbb{K}^K$: set of $\mathbb{K}$-valued vectors of dimension K.
- $\mathcal{S}^K = \left\{(\theta_1 \dots \theta_K) \in \mathbb{K}^K, \sum_{k=1}^K |\theta_k|^2 = 1\right\}$: $K-1$ dimensional hypersphere.
- $\mathcal{B}\left(\mathcal{S}^K\right)$: set of Borelian sets on $\mathcal{S}^K$.

[1] In the literature, the spatial density is rather called *spectral measure* (Samoradnitsky and Taqqu [35]). We have deliberately chosen to avoid the term "spectral" here for two reasons. First, we believe that it may bring some confusion for a signal processing audience. Second, we think that calling it spatial better highlights the fact that it encodes dependencies between covariates.

- $\boldsymbol{\theta}$: vector from the hypersphere $\boldsymbol{\theta} \in \mathcal{S}^K$, also called a *direction,* with entries θ_k.
- Θ: A partition $\{\Theta_1, \ldots, \Theta_P\}$ of $\mathcal{S}^K$, for which:
  - All *cells* Θ_p have the same area Δ_Θ.
  - $\boldsymbol{\theta}_p \in \Theta_p$ denotes an element of cell Θ_p.
- $\Re$: real part of a complex number.
- $.^\star$: Hermitian transposition (resp. conjugation) of a complex vector (resp. complex number).
- $\langle ., . \rangle$: inner product on $\mathbb{K}^K$: $\langle \boldsymbol{w}, \boldsymbol{x} \rangle = \boldsymbol{w}^\star \boldsymbol{x}$.
- y_{jk}: k^{th} component of a vector $\boldsymbol{y}_j$.
- $\boldsymbol{e}_k$: k^{th} vector from the canonical basis.
- $.^{\langle . \rangle}$: signed power function (Samoradnitsky and Taqqu [35], Nikias and Shao [28]): $\forall z \in \mathbb{K}$, $z^{\langle \alpha \rangle} = z\,|z|^{\alpha-1}$
- $\triangleq$: "equal by definition to".
- $\leftarrow$: assignment of a value to a variable in an algorithm.

2. The multivariate α-stable probabilistic model

2.1. Isotropic symmetric α-stable random variables

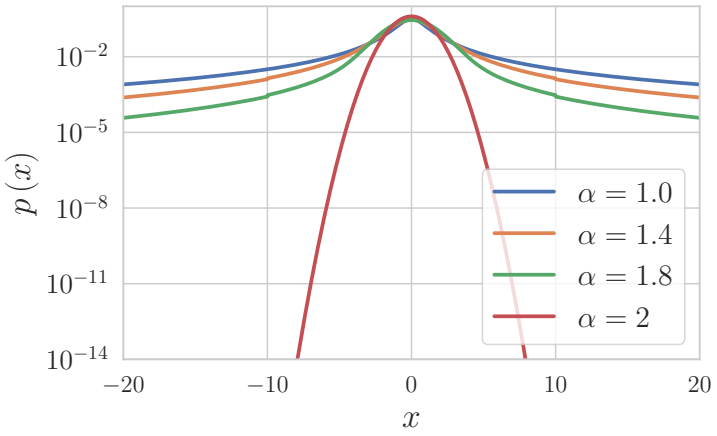


Figure 1: $S\alpha S_c^1$ density probability functions in the real case with $\sigma = 1$

An isotropic (circular) symmetric α-stable scalar random variable $x \sim S\alpha S_c^1(\sigma_x)$ can be defined by its characteristic function (chf.) $\varphi_x : u \in \mathbb{K} \mapsto \mathbb{E}(\exp(i\Re(u^\star x)))$ as follows:

$$\forall u \in \mathbb{K},\ \varphi_x(u) = \exp(-|u|^\alpha \sigma_x^\alpha)$$

where σ_x is called a *scale factor*. The real number $\alpha \in (0, 2]$ is called the *characteristic exponent* and defines the heaviness of the tails in a stable distribution: the smaller α, the heavier the tails. Examples of probability density functions (pdf.) for such scalar random variables are represented in the real case in Fig. 1. It is remarkable that such a pdf. is not available in closed-form in general, but only in some particular cases, like the Cauchy ($\alpha = 1$) and Gaussian ($\alpha = 2$) distributions.

2.2. α-stable random vectors

In this paper, we limit our attention to symmetric $\alpha-$stable random vectors, i.e. α-stable random vectors $\boldsymbol{x}$ such that $\boldsymbol{x}$ and $-\boldsymbol{x}$ have the same distribution, defined as in (Samoradnitsky and Taqqu [35]):

Definition 1. *Let $\boldsymbol{x}$ be a random vector in $\mathbb{K}^K$ associated to its characteristic function $\varphi_{\boldsymbol{x}} : \boldsymbol{u} \in \mathbb{K}^K \mapsto \mathbb{E}(\exp(i\Re\langle \boldsymbol{u}, \boldsymbol{x}\rangle))$. A symmetric isotropic α-stable distribution with $\alpha \in (0, 2)$ is fully described by the unique representation:*

$$\forall \boldsymbol{u} \in \mathbb{K}^K,\ \varphi_{\boldsymbol{x}}(\boldsymbol{u}) = \exp\left(-\int_{\boldsymbol{\theta}\in\mathcal{S}^K} |\langle \boldsymbol{u}, \boldsymbol{\theta}\rangle|^\alpha \Gamma_{\boldsymbol{x}}(d\boldsymbol{\theta})\right), \tag{1}$$

where $\Gamma_{\boldsymbol{x}}$ is a symmetric measure on the sphere $\mathcal{S}^K$ called the spatial density[2]*. Henceforth, we note $\boldsymbol{x} \sim S\alpha S_c^K(\Gamma_{\boldsymbol{x}})$ whenever $\boldsymbol{x}$ follows a symmetric α-stable distribution with* spatial density $\Gamma_{\boldsymbol{x}}$.[3]

Remark 2. A symmetric $\alpha-$stable vector $\boldsymbol{x} \sim S\alpha S_c^K(\Gamma_{\boldsymbol{x}})$ belongs to a larger class: the *elliptically multivariate contoured* (EMC) distribution (Cambanis et al. [6]). In many cases, the sum of EMC vectors $\boldsymbol{x}, \boldsymbol{y}$, respectively associated to the so-called scatter matrices $\boldsymbol{R}_{\boldsymbol{x}}$ and $\boldsymbol{R}_{\boldsymbol{y}}$, is still an EMC vector. However, the parameter $\boldsymbol{R}_{\boldsymbol{x}+\boldsymbol{y}}$ called *scatter matrix* of the mix is usually not equal to $\boldsymbol{R}_{\boldsymbol{x}} + \boldsymbol{R}_{\boldsymbol{y}}$. In the $\alpha-$stable case, Definition 1 ensures that the *spatial density* of the mix $\Gamma_{\boldsymbol{x}+\boldsymbol{y}}$ is equal to $\Gamma_{\boldsymbol{x}} + \Gamma_{\boldsymbol{y}}$.

We highlight the fact that the Gaussian ($\alpha = 2$) case is omitted in Definition 1, because the representation (1) is not unique in that case.

---

[2] See Footnote 1 on page 4.

[3] $\Gamma_{\boldsymbol{x}}$ is a symmetric measure on $\mathcal{S}^K$ in the sense that for any continuous function f defined on $\mathcal{S}^K$ and for any $z \in \mathbb{K}$ such that $|z| = 1$, we have $\int_{\boldsymbol{\theta}\in\mathcal{S}^K} f(\boldsymbol{\theta})\Gamma_{\boldsymbol{x}}(z d\boldsymbol{\theta}) = \int_{\boldsymbol{\theta}\in\mathcal{S}^K} f(\boldsymbol{\theta})\Gamma_{\boldsymbol{x}}(d\boldsymbol{\theta})$.

2.3. Spatial spectrum and spatial representation

In this section, we show how the spatial density $\Gamma_{\boldsymbol{x}}$, featured in Definition 1, of a symmetric α -stable distribution can be understood through a so-called *spatial representation*, which is central to the filtering methods we propose later in Sections 3 and 4.

Let $\mathcal{X}$ be an independently scattered α-stable random measure on $\mathcal{S}^K$, with control measure $\Gamma_{\boldsymbol{x}}$ (Samoradnitsky and Taqqu [35]). This means that:

1. for any Borelian set $A \subset \mathcal{B}\left(\mathcal{S}^K\right)$, the scalar random variable $\mathcal{X}(A) \in \mathbb{K}$ is distributed as $\mathcal{X}(A) \sim S\alpha S_c\left(\Gamma_{\boldsymbol{x}}(A)\right)$,
2. $\mathcal{X}(A)$ is independent from $\mathcal{X}(B)$ whenever $A \cap B = \emptyset$ for any two Borelian subsets A and $B \subset \mathcal{B}\left(\mathcal{S}^K\right)$.

We call $\mathcal{X}$ the *spatial spectrum* of the distribution $S\alpha S_c^K\left(\Gamma_{\boldsymbol{x}}\right)$.

Remark 3. In (Ma and Nikias [23]), the $\alpha-$*spectrum* terminology is used in a different way as the *spatial spectrum.* It describes the so-called covariation (see Section 4.1 for further details) between the input single-channel signal and the output single-channel signal.We have the following first result:

Theorem 4. *Let* $\boldsymbol{x} \sim S\alpha S_c^K\left(\Gamma_{\boldsymbol{x}}\right)$*, with spatial density* $\Gamma_{\boldsymbol{x}}$*. Then* $\boldsymbol{x}$ *admits the following* spatial representation*:*

$$\boldsymbol{x} \overset{d}{=} \int_{\boldsymbol{\theta} \in \mathcal{S}^K} \boldsymbol{\theta} \mathcal{X}(d\boldsymbol{\theta}), \tag{2}$$

where $\overset{d}{=}$ *means "equal in distribution" and* $\mathcal{X}$ *is the* spatial spectrum *with control measure* $\Gamma_{\boldsymbol{x}}$*.*

The proof of this result is given in Appendix 1. The representation theorem means that an $S\alpha S_c^K$ random vector is distributed as the sum of infinitely many contributions, coming from all directions $\boldsymbol{\theta} \in \mathcal{S}^K$ on the sphere. $\Gamma_{\boldsymbol{x}}(d\boldsymbol{\theta})$ may thus be interpreted as the *scale factor* of the contributions pointing in direction $\boldsymbol{\theta}$. Very interestingly, the spatial representation in Theorem 4 provides a straightforward way to generate samples from $S\alpha S_c^K\left(\Gamma_{\boldsymbol{x}}\right)$ random vectors: first, generate the spatial spectrum $\mathcal{X}$, and then use (2) to construct the $S\alpha S_c^K$ random vector $\boldsymbol{x}$. The method is summarized in Algorithm 1.

The integration over the real or complex sphere, appearing in (2), is replaced by finite sums. This is done by simply constructing a regular partition Θ of the sphere $\mathcal{S}^K$, and substituting the integration by the corresponding sum carried over the Θ_p's. Let f be a function defined on the hypersphere and M be a measure on the sphere. For P large enough, the approximation goes as:

$$\int_{\mathcal{S}^K} \boldsymbol{\theta} f(\boldsymbol{\theta}) M(d\boldsymbol{\theta}) \approx \sum_p \boldsymbol{\theta}_p f(\boldsymbol{\theta}_p) M(\Theta_p), \tag{3}$$

where $M(\Theta_p)$ may further be approximated as $m(\boldsymbol{\theta}_p)\Delta_\Theta$ (where Δ_Θ is the Lebesgue measure of Θ_p), whenever M is dominated by the Lebesgue measure and thus equal to $M(d\boldsymbol{\theta}) = m(\boldsymbol{\theta})\, d\boldsymbol{\theta}$ for some measurable function m.

Algorithm 1 Sampling of $S\alpha S_c^K(\Gamma_{\boldsymbol{x}})$ random vectors through their spatial representation.

1. **Input**
   - Number N of desired realizations
   - Partition $\Theta = \{\Theta_1, \ldots, \Theta_P\}$ of $\mathcal{S}^K$ as described in Section 1
   - Characteristic exponent $\alpha \in (0, 2)$
   - Spatial density $\Gamma_{\boldsymbol{x}}$
2. **Spatial spectrum generation**
   $\forall n, p,\ \mathcal{X}_{np} \sim S\alpha S_c(\Gamma_{\boldsymbol{x}}(\Theta_p))$ where $\Gamma_{\boldsymbol{x}}(\Theta_p)$ is the restriction of $\Gamma_{\boldsymbol{x}}$ to the set Θ_p up to renormalization.
3. **Synthesis** $\forall n, x_n \leftarrow \sum_p \boldsymbol{\theta}_p \mathcal{X}_{np}$

Given our model for α-stable vectors, we now discuss the distribution of mixtures of such random vectors.

2.4. Mixtures of α-stable random vectors

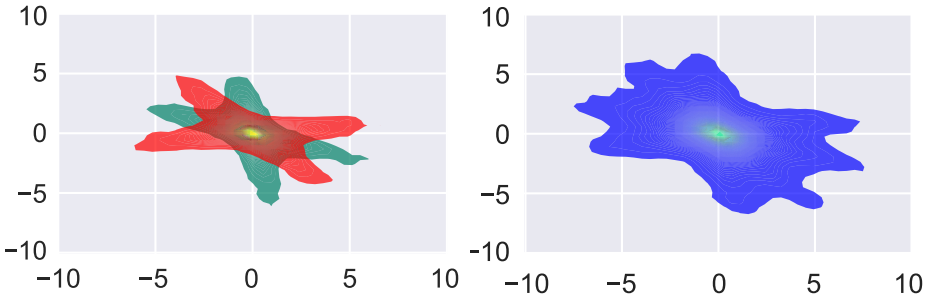


Figure 2: Spatial densities in the real case with $K = 2$. **On the left**, density plots of $\boldsymbol{y}_1 \sim S\alpha S_c^K(\Gamma_1)$ (in green) and $\boldsymbol{y}_2 \sim S\alpha S_c^K(\Gamma_2)$ (in red), where the maxima of Γ_1 and Γ_2 are reached for $\left\{\frac{5\pi}{8}, \frac{5\pi}{8} + \frac{\pi}{3}\right\}$ and $\left\{\frac{5\pi}{8} + \frac{\pi}{6}, \frac{5\pi}{8} + \frac{\pi}{2}\right\}$. These plots show the influence of the spatial density on the dependence patterns between covariates. **On the right**, a density plot for the mixture $\boldsymbol{x} = \boldsymbol{y}_1 + \boldsymbol{y}_2$ shows the additive property of spatial densities.

In the filtering and signal processing literature, it is common to assume that the observed vector $\boldsymbol{x} \in \mathbb{K}^K$ is the sum of J components $\boldsymbol{y}_j \in \mathbb{K}^K$ that we want to recover. Here, we take each component $\boldsymbol{y}_j \sim S\alpha S_c^K(\Gamma_j)$ as described above, with its own spatial density Γ_j:

$$\begin{cases} \boldsymbol{x} = \sum_{j=1}^J \boldsymbol{y}_j \\ \forall j,\ \boldsymbol{y}_j \sim S\alpha S_c^K(\Gamma_j)\,. \end{cases} \tag{4}$$

Because $\boldsymbol{x}$ is the sum of α-stable vectors, then $\boldsymbol{x}$ itself follows an α-stable distribution, with the following spatial density:

$$\begin{cases} \boldsymbol{x} \sim S\alpha S_c^K\left(\Gamma_{\boldsymbol{x}}\right) \\ \Gamma_{\boldsymbol{x}} = \sum_j \Gamma_j. \end{cases}$$

By invoking the spatial representation Theorem 4 on each latent vector $\boldsymbol{y}_j$, we get: $\forall j,\ \boldsymbol{y}_j \stackrel{d}{=} \int_{\boldsymbol{\theta}\in\mathcal{S}^K} \boldsymbol{\theta}\mathcal{Y}_j\left(d\boldsymbol{\theta}\right)$, where $\mathcal{Y}_j$ denotes the spatial spectrum of $\boldsymbol{y}_j$. Moreover, $\mathcal{X} \triangleq \sum_j \mathcal{Y}_j$ also defines a spatial spectrum associated to $\boldsymbol{x}$. Informally, it simply means that $\boldsymbol{x}$ is also the sum of infinitely many contributions $\mathcal{X}\left(d\boldsymbol{\theta}\right)$, coming from all directions $\boldsymbol{\theta} \in \mathcal{S}^K$ on the sphere, each one of them being in turn the sum of the contributions $\mathcal{Y}_j\left(d\boldsymbol{\theta}\right)$ for all components that come from this particular direction. An illustration of the relationships between the spatial densities $\Gamma_{\boldsymbol{x}}, \Gamma_j$ is given in Fig. 2.

The next two sections make different uses of this spatial representation to devise *filters*, which aim at estimating the latent vectors $\boldsymbol{y}_j$ given $\boldsymbol{x}$, provided the spatial densities Γ_j, which are parameters, are *known*. This allows us to dissociate the actual filtering problem from the question of estimating signal parameters. This strategy is for instance classical in the literature focusing on Wiener-based filter design (Wiener [43], Duong et al. [11]).

3. Spatial spectrum filter (SSF)

Our objective is to estimate each latent vector $\boldsymbol{y}_j$, such that $\sum \boldsymbol{y}_j = \boldsymbol{x}$. The strategy we discuss here proceeds in two steps. Firstly, we estimate the spatial spectrum $\mathcal{X}$ of the mixture, as described in Section 3.1. Secondly, this estimate is used to reconstruct the desired latent vectors $\boldsymbol{y}_j$, as detailed in Section 3.2.

3.1. Spatial spectrum estimation

We assume that the observation $\boldsymbol{x} \sim S\alpha S_c^K\left(\Gamma_{\boldsymbol{x}}\right)$ and its spatial density Γ_x are known. The first step in the filter we discuss here is to estimate the spatial spectrum $\mathcal{X}$. For this purpose, we choose the *a posteriori* expectation $\widehat{\mathcal{X}}\left(d\boldsymbol{\theta}\right)$ of $\mathcal{X}\left(d\boldsymbol{\theta}\right)$ given $\boldsymbol{x}$, defined in the following sense: for any continuous function ψ on $\mathcal{S}^K$, satisfying $\int_{\boldsymbol{\theta}\in\mathcal{S}^K} \left|\psi\left(\boldsymbol{\theta}\right)\right|^{\alpha} \Gamma_{\boldsymbol{x}}\left(d\boldsymbol{\theta}\right) < +\infty$, we have:

$$\mathbb{E}\left[\int_{\boldsymbol{\theta}\in\mathcal{S}^K} \psi\left(\boldsymbol{\theta}\right)\mathcal{X}\left(d\boldsymbol{\theta}\right) \mid \boldsymbol{x}\right] = \int_{\boldsymbol{\theta}\in\mathcal{S}^K} \psi\left(\boldsymbol{\theta}\right)\widehat{\mathcal{X}}\left(d\boldsymbol{\theta}\right). \tag{5}$$

Note that it is such that $\int_{\boldsymbol{\theta}\in\mathcal{S}^K} \boldsymbol{\theta}\widehat{\mathcal{X}}\left(d\boldsymbol{\theta}\right) = \boldsymbol{x}$. It turns out that in the particular case of an $S\alpha S_c^K\left(\Gamma_{\boldsymbol{x}}\right)$ observation $\boldsymbol{x}$, $\widehat{\mathcal{X}}\left(d\boldsymbol{\theta}\right)$ has the following form:

Proposition 5. *Under the previous assumptions, $\widehat{\mathcal{X}}\left(d\boldsymbol{\theta}\right)$ can be rewritten as:*

$$\widehat{\mathcal{X}}\left(d\boldsymbol{\theta}\right) = g_{_{\mathcal{X}}}\left(\boldsymbol{x},\boldsymbol{\theta}\right)\Gamma_{\boldsymbol{x}}\left(d\boldsymbol{\theta}\right) \quad a.s. \tag{6}$$

with:

$$\forall \boldsymbol{\theta} \in \mathcal{S}^K,\ g_{_{\mathcal{X}}}\left(\boldsymbol{x},\boldsymbol{\theta}\right) = \frac{N\left(\boldsymbol{x},\boldsymbol{\theta}\right)}{D\left(\boldsymbol{x}\right)}, \tag{7}$$

where

$$N(\boldsymbol{x},\boldsymbol{\theta}) = \alpha i \int_{\boldsymbol{u}\in\mathbb{K}^K} \langle \boldsymbol{\theta},\boldsymbol{u}\rangle^{\langle\alpha-1\rangle} \varphi_{\boldsymbol{x}}(\boldsymbol{u})\; e^{-i\Re(\langle \boldsymbol{u},\boldsymbol{x}\rangle)} d\boldsymbol{u} \tag{8}$$

and

$$D(\boldsymbol{x}) = \int_{\boldsymbol{u}\in\mathbb{K}^K} \varphi_{\boldsymbol{x}}(\boldsymbol{u})\; e^{-i\Re(\langle \boldsymbol{u},\boldsymbol{x}\rangle)} d\boldsymbol{u}. \tag{9}$$

If $\alpha > 1$, the density $g_{\mathcal{X}}(\boldsymbol{x},\boldsymbol{\theta})$ is a continuous map on $\mathcal{S}^K$, and hence the measure $\widehat{\mathcal{X}}(d\boldsymbol{\theta})$ is dominated by $\Gamma_{\boldsymbol{x}}(d\boldsymbol{\theta})$.

Proposition 5 is proved in Appendix 2.

Although $g_{\mathcal{X}}(\boldsymbol{x},\boldsymbol{\theta})$ has the closed-form expression (7), its computation is not straightforward because it requires two integrations over $\mathbb{K}^K$. Let $I_{\boldsymbol{x}}(\boldsymbol{u}) \triangleq -\ln(\varphi_{\boldsymbol{x}}(\boldsymbol{u}))$ be the Levy exponent of $\boldsymbol{x}$ (Unser and Tafti [37]). From (1), it is given by: $I_{\boldsymbol{x}}(\boldsymbol{u}) = \int_{\boldsymbol{\theta}\in\mathcal{S}^K} |\langle \boldsymbol{u},\boldsymbol{\theta}\rangle|^{\alpha} \Gamma_{\boldsymbol{x}}(d\boldsymbol{\theta})$. We have the following result, with proofs given in Appendix 2:

Proposition 6. *Let $\beta = 1$ if $\mathbb{K} = \mathbb{R}$, or $\beta = 2$ if $\mathbb{K} = \mathbb{C}$. Then (8) is equivalent to*

$$N(\boldsymbol{x},\boldsymbol{\theta}) = \int_{\boldsymbol{\theta}'\in\mathcal{S}^K} \frac{\langle \boldsymbol{\theta},\boldsymbol{\theta}'\rangle^{\langle\alpha-1\rangle} \langle \boldsymbol{\theta}',\boldsymbol{x}\rangle}{I_{\boldsymbol{x}}(\boldsymbol{\theta}')^{\frac{\beta K+\alpha}{\alpha}}} \eta\left(\frac{|\langle \boldsymbol{\theta}',\boldsymbol{x}\rangle|^2}{I_{\boldsymbol{x}}(\boldsymbol{\theta}')^{\frac{2}{\alpha}}}\right) d\boldsymbol{\theta}', \tag{10}$$

where:

$$\eta(\rho) = \sum_{n=0}^{+\infty} \frac{(-1)^n f_\Gamma\left(\frac{2n+\beta K+\alpha}{\alpha}\right)}{2^{2n+1} n!\,(n+1)!} \rho^n. \tag{11}$$

with f_Γ the Gamma function. In the same way, (9) is equivalent to

$$D(\boldsymbol{x}) = \frac{\left\| \int_{\mathcal{S}^K} \boldsymbol{\theta} N(\boldsymbol{x},\boldsymbol{\theta})\, \Gamma_x(d\boldsymbol{\theta}) \right\|_1}{\|\boldsymbol{x}\|_1}. \tag{12}$$

Proposition 6 is proved in Appendix 2.

Note that in (12), any norm could be picked for the computation and yield the same result. The ℓ_1-norm $\|.\|_1$ turns out to be a good compromise between computational cost and numerical stability.

Equations (10) and (12) in Proposition 6 provide an estimator of $g_{\mathcal{X}}$, which is computationally tractable via integrations on the compact set $\mathcal{S}^K$, as opposed to (7) that requires integration over $\mathbb{K}^K$.

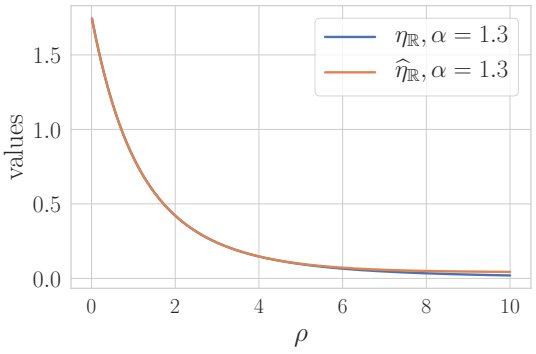
The quantity $\eta(\rho)$ is a power series with an infinite radius of convergence when $\alpha > 1$. It is a smooth function of ρ and independent of the data and the model parameters. It can hence be computed beforehand.

The main computational burden for this method lies in the computation of the alternating power series η in (11). It converges slowly and goes through extreme values, requiring a fairly large numerical precision in practice. However, there are some cases where a closed-form expression is available, for instance

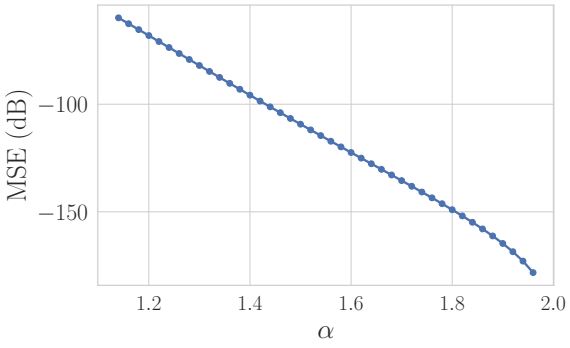
when $\alpha = 2$, $K = 2$, $\beta = 2$, where $\eta(\rho) = \frac{1}{16}\left(\rho^2 - 20\rho + 64\right) e^{-\rho/4}$. Inspired by this result, we decided to approximate η in all cases as:

$$\eta(\rho) \approx \widehat{\eta}(\rho) = \left(a\rho^2 + b\rho + c\right) e^{-d\rho} \tag{13}$$

where $a, b, c, d \in \mathbb{R}$ are model parameters. Using such a parameterized version allows us to avoid the on-demand time-consuming evaluations of η. We estimated the parameters that minimize the mean-square-error (MSE) $\|\eta - \hat{\eta}\|_2^2$. We highlight that both the choice of this parametric model (13) and the choice of the MSE criterion are driven by ad-hoc considerations, indeed, there is no theoretical result we are aware of that would justify the convergence of the power series to an exponential function. However, the plots for η and $\hat{\eta}$ for $\beta = 1$ are displayed in Fig. 3(a), and the error of fit for η for $\beta = 2$ as a function of α is displayed in Fig. 3(b) and quality is very good. Results were similar in the real and complex cases. η was computed for 50 regularly spaced values $\alpha \in (1, 2)$, and for $\rho \in (0, 10)$, because ρ only has positive values in (10). For getting a suitable convergence, η was calculated up to order 10^5.



(a) η and $\hat{\eta}$ for $\alpha = 1.3$ and $\beta = 1$.



(b) MSE as a function of α.

Figure 3: Performance of a parametric fit $\hat{\eta}$ for η in (11) .

3.2. Signal reconstruction

Once the spatial spectrum $\mathcal{X}(d\boldsymbol{\theta})$ of the observation is estimated, or equivalently $g_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{\theta})$ is computed through (7), (10) and (12), our next step is to construct an estimate for the components $\boldsymbol{y}_j$ of interest. We pick the *a posteriori* expectation: $\widehat{\boldsymbol{y}}_j \triangleq \mathbb{E}\left[\boldsymbol{y}_j \mid \boldsymbol{x}\right]$, which is given as follows:

Theorem 7. *Let $\boldsymbol{x}$ be the sum of $S\alpha S_c^K$ random vectors $\boldsymbol{y}_j \sim S\alpha S_c^K(\Gamma_j)$, each with known spatial density Γ_j. Then the* a posteriori *expectation of each component $\boldsymbol{y}_j$ given $\boldsymbol{x}$ is:*

$$\widehat{\boldsymbol{y}}_j \triangleq \mathbb{E}\left[\boldsymbol{y}_j \mid \boldsymbol{x}\right] = \int_{\boldsymbol{\theta} \in \mathcal{S}^K} \boldsymbol{\theta} g_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{\theta}) \Gamma_j(d\boldsymbol{\theta}), \tag{14}$$

where $g_{\mathcal{X}}$ was defined in (7).

The proof of this theorem is also in Appendix 2. A summary of this filtering technique is given in Algorithm 2:

Algorithm 2 $\alpha-$**SSF**: multivariate α-stable filtering through a spatial spectrum estimation.

1. **Input**
   - Observation $\boldsymbol{x}$ of size K
   - Regular partition $\Theta = \{\Theta_1, \ldots, \Theta_P\}$ of $\mathcal{S}^K$
   - Characteristic exponent $\alpha \in (1, 2)$
   - spatial densities Γ_j
2. **Spatial spectrum estimation**
   - Using $\hat{\eta}$ in (13), compute $\forall p$, $N(\boldsymbol{x}, \boldsymbol{\theta}_p)$ in (10)
   - Compute $D(\boldsymbol{x})$ in (12)
   - Compute $g_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{\theta}_p) = \frac{N(\boldsymbol{x}, \boldsymbol{\theta}_p)}{D(\boldsymbol{x})}$
3. **Reconstruction:** $\widehat{\boldsymbol{y}}_j = \sum_p \boldsymbol{\theta}_p g_{\mathcal{X}}(x, \boldsymbol{\theta}_p) \Gamma_j(\Theta_p)$

4. Covariation-minimizing filter (CMF)

Despite a relatively simple expression of $g_{\mathcal{X}}$, the estimation technique in Section 3 is computationally demanding, due to several numerical integrations. In this section, the spatial representation in Theorem 4 will be exploited differently, leading to a faster filtering method.

4.1. Covariation between stable variables

Many signal processing studies exploit second-order statistics for the design of digital filters (Cardoso [7], Duong et al. [11], Moussaoui et al. [27]). This convenient strategy finds a straightforward probabilistic interpretation through

Gaussian processes ($\alpha = 2$), that are characterized by their covariance functions. However, this strategy breaks down for α-stable processes with $\alpha < 2$, because the moments of order $p \geq \alpha$ are infinite. For this reason, the *covariation* was introduced (Samoradnitsky and Taqqu [35], Nikias and Shao [28, p. 87]) as a substitute of the covariance, with many similar properties.

Definition 8. The covariation between two random variables (x_1, x_2), jointly distributed as: $(x_1, x_2) \triangleq \boldsymbol{x} \sim S\alpha S_c^2(\Gamma_{\boldsymbol{x}})$ for $\alpha > 1$, is defined as:

$$[x_1, x_2]_\alpha \triangleq \int_{\boldsymbol{z}=(z_1,z_2)\in\mathcal{S}^2} z_1^* z_2^{\langle\alpha-1\rangle} \Gamma_{\boldsymbol{x}}(d\boldsymbol{z}).$$

Moreover, the *covariation norm* (Samoradnitsky and Taqqu [35, p 95]) of $u \sim S\alpha S_c^1$ is:

$$\|u\|_\alpha = ([u,u]_\alpha)^{1/\alpha}.$$

Remark 9. The covariation is always anti-linear in its left argument. It is also linear in the right argument if and only if all terms of the linear combination on the right-hand side are mutually independent: if (x, x_1, x_2) are jointly $S\alpha S_c^3$ with x_1 and x_2 independent, then $[x, x_1 + x_2]_\alpha = [x, x_1]_\alpha + [x, x_2]_\alpha$. In addition, if x_1 and x_2 are independent then $[x_1, x_2]_\alpha = 0$, and if $x \sim S\alpha S_c^1(\sigma_x)$, then $\|x\|_\alpha = \sigma_x$.

4.2. Covariation minimization filtering (CMF) technique

Our objective in this section is to build a filter to extract the component $\boldsymbol{y}_j$ from the mixture $\boldsymbol{x}$. Motivated by (Masry [26]), we seek filtering vectors $\boldsymbol{w}_{jk} \in \mathbb{K}^K$ such that $\widehat{y}_{jk} = \langle \boldsymbol{w}_{jk}, \boldsymbol{x}\rangle$ minimizes the covariation norm $\|y_{jk} - \widehat{y}_{jk}\|_\alpha^\alpha$. Additionally, we enforce $\sum_j \boldsymbol{w}_{jk} = \boldsymbol{e}_k$ to guarantee perfect reconstruction of the mixture: $\boldsymbol{x} = \sum_j \widehat{\boldsymbol{y}}_j$. For each k, this results in the following optimization problem with linear equality constraints:

$$\begin{array}{ll} \text{minimize} & \sum_j \|y_{jk} - \langle \boldsymbol{w}_{jk}, \boldsymbol{x}\rangle\|_\alpha^\alpha \text{ w.r.t. } \boldsymbol{w}_{jk} \\ \text{subject to} & \sum_j \boldsymbol{w}_{jk} = \boldsymbol{e}_k. \end{array} \tag{15}$$

The constraints and covariation norm have convenient properties. Firstly, the constraints are linear. Secondly, the criterion is a differentiable function whose derivative is continuous, and it is convex. Thirdly, the covariation norm is coercive. By invoking the Karush, Khun and Tucker theorem (Boyd and Vandenberghe [5]), this optimization problem thus has a unique solution.

We apply the spatial representation in Theorem 4 to get $\boldsymbol{x} \overset{d}{=} \int \boldsymbol{\theta}\mathcal{X}(d\boldsymbol{\theta})$ and $\boldsymbol{y}_j \overset{d}{=} \int \boldsymbol{\theta}\mathcal{Y}_j(d\boldsymbol{\theta})$, with integrations done over S^K. The Lagrangian of the

problem (15) is

$$\mathcal{L}\left(\{\boldsymbol{w}_{jk}\}_j, \boldsymbol{\lambda}_k\right) = \sum_j \|y_{jk} - \langle \boldsymbol{w}_{jk}, \boldsymbol{x}\rangle\|_\alpha^\alpha + \alpha\Re\left(\boldsymbol{\lambda}_k^*\left(\sum_j \boldsymbol{w}_{jk} - \boldsymbol{e}_k\right)\right) \tag{16}$$

where the factor α in the second line is introduced to simplify the following calculations. Now, thanks to properties of the covariation given in Remark 9, the development of $\|y_{jk} - \widehat{y}_{jk}\|_\alpha^\alpha$ for all j, k yields:

$$\begin{aligned}
&\|y_{jk} - \widehat{y}_{jk}\|_\alpha^\alpha \\
&= \left\| y_{jk} - \sum_{j'} \langle \boldsymbol{w}_{jk}, \boldsymbol{y}_{j'}\rangle \right\|_\alpha^\alpha \\
&= \left\| \int \left(\theta_k - \langle \boldsymbol{w}_{jk}, \boldsymbol{\theta}\rangle\right) \mathcal{Y}_j\left(d\boldsymbol{\theta}\right) \right\|_\alpha^\alpha + \sum_{j'\neq j} \left\| \int \langle \boldsymbol{w}_{jk}, \boldsymbol{\theta}\rangle \, \mathcal{Y}_{j'}\left(d\boldsymbol{\theta}\right) \right\|_\alpha^\alpha \\
&= \int |\theta_k - \langle \boldsymbol{w}_{jk}, \boldsymbol{\theta}\rangle|^\alpha \, \Gamma_j\left(d\boldsymbol{\theta}\right) + \sum_{j'\neq j} \int |\langle \boldsymbol{w}_{jk}, \boldsymbol{\theta}\rangle|^\alpha \, \Gamma_{j'}\left(d\boldsymbol{\theta}\right) \\
&= \int \left(|\theta_k - \langle \boldsymbol{w}_{jk}, \boldsymbol{\theta}\rangle|^\alpha - |\langle \boldsymbol{w}_{jk}, \boldsymbol{\theta}\rangle|^\alpha\right) \Gamma_j\left(d\boldsymbol{\theta}\right) + \int |\langle \boldsymbol{w}_{jk}, \boldsymbol{\theta}\rangle|^\alpha \, \Gamma_x(d\boldsymbol{\theta}).
\end{aligned} \tag{17}$$

By substituting (17) in (16), and by zeroing the gradient of $\mathcal{L}$ w.r.t. $\boldsymbol{w}_{jk}$, we get for all j:

$$\begin{aligned}
\boldsymbol{\lambda}_k = &\int \boldsymbol{\theta}\left(\left(\theta_k^* - \langle \boldsymbol{\theta}, \boldsymbol{w}_{jk}\rangle\right)^{\langle\alpha-1\rangle} + \langle \boldsymbol{\theta}, \boldsymbol{w}_{jk}\rangle^{\langle\alpha-1\rangle}\right) \Gamma_j\left(d\boldsymbol{\theta}\right) \\
&- \int \boldsymbol{\theta}\, \langle \boldsymbol{\theta}, \boldsymbol{w}_{jk}\rangle^{\langle\alpha-1\rangle} \Gamma_x\left(d\boldsymbol{\theta}\right).
\end{aligned} \tag{18}$$

By noting that $z^{\langle\alpha-1\rangle} = \frac{z}{|z|^{2-\alpha}}$, (18) can be written as:

$$\boldsymbol{\lambda}_k = -\boldsymbol{P}_{jk}\boldsymbol{w}_{jk} + \boldsymbol{r}_{jk}, \tag{19}$$

where the $K \times K$ matrix $\boldsymbol{R}_{jk}$ and the vector $\boldsymbol{r}_{jk} \in \mathbb{K}^K$ are defined as:

$$\begin{aligned}
\boldsymbol{P}_{jk} = &\int \left(\frac{\boldsymbol{\theta}\boldsymbol{\theta}^\star}{|\theta_k - \langle \boldsymbol{w}_{jk}, \boldsymbol{\theta}\rangle|^{2-\alpha}} - \frac{\boldsymbol{\theta}\boldsymbol{\theta}^\star}{|\langle \boldsymbol{w}_{jk}, \boldsymbol{\theta}\rangle|^{2-\alpha}}\right) \Gamma_j\left(d\boldsymbol{\theta}\right) \\
&+ \int \frac{\boldsymbol{\theta}\boldsymbol{\theta}^\star}{|\langle \boldsymbol{w}_{jk}, \boldsymbol{\theta}\rangle|^{2-\alpha}} \Gamma_x\left(d\boldsymbol{\theta}\right).
\end{aligned} \tag{20}$$

$$\boldsymbol{r}_{jk} = \int \frac{\boldsymbol{\theta}\theta_k^\star}{|\theta_k - \langle \boldsymbol{w}_{jk}, \boldsymbol{\theta}\rangle|^{2-\alpha}} \Gamma_j\left(d\boldsymbol{\theta}\right). \tag{21}$$

Therefore $\boldsymbol{w}_{jk}$ is a fixed point of the following equation:

$$\boldsymbol{w}_{jk} = \boldsymbol{P}_{jk}^{-1}\left(\boldsymbol{r}_{jk} - \boldsymbol{\lambda}_k\right). \tag{22}$$

A sum over j in (22) leads to:

$$\boldsymbol{\lambda}_k = \left(\sum_j \boldsymbol{P}_{jk}^{-1} \right)^{-1} \left(\left(\sum_j \boldsymbol{P}_{jk}^{-1} \boldsymbol{r}_{jk} \right) - \boldsymbol{e}_k \right). \tag{23}$$

Putting together the above results, we propose to design the filters $\boldsymbol{w}_{jk}$ based on a fixed-point approach where $\boldsymbol{P}_{jk}$ in (20), $\boldsymbol{w}_{jk}$ in (22) and $\boldsymbol{\lambda}_k$ in (23) are updated in turn. This is summarized in Algorithm 3. The integrals are replaced by a discrete sum as in (3).

Algorithm 3 $\alpha-$**CMF**: multivariate α-stable filtering through covariation minimization

---

1. **Input**
   - Observation $\boldsymbol{x}$ of size K
   - Regular partition $\Theta = \{\Theta_1, \ldots, \Theta_P\}$ of $\mathcal{S}^K$
   - Characteristic exponent $\alpha \in (1, 2)$
   - Spatial densities Γ_j
2. **Initialization**
   - $\forall j, k,\ \boldsymbol{w}_{jk} = \frac{1}{J} \boldsymbol{e}_k$
   - $\forall j, k,$ the entries of $\boldsymbol{P}_{j,k}$ are independently drawn from the standard normal distribution
   - $\forall k, \boldsymbol{\lambda}_k = 0$
3. **Updates of $\boldsymbol{P}_{jk}$, $\boldsymbol{w}_{j,k}$ and $\boldsymbol{\lambda}_k$**
   - $\forall j, k,$ update $\boldsymbol{P}_{jk}$ as in (20)
   - $\forall k,\ \boldsymbol{\lambda}_k \leftarrow \boldsymbol{\lambda}_k + \left(\sum_j \boldsymbol{P}_{jk}^{-1} \right)^{-1} \left(\left(\sum_j \boldsymbol{w}_{jk} \right) - \boldsymbol{e}_k \right)$
   - $\forall j, k,$ update $\boldsymbol{w}_{jk}$ as in (22) and (21)
4. **Reconstruction:** $\forall j, k,\ \widehat{y}_{jk} = \langle \boldsymbol{w}_{jk}, \boldsymbol{x} \rangle$

4.3. The Gaussian case ($\alpha = 2$)

Although the derivations done above focus on the case $\alpha \in (1, 2)$, it is interesting to note the limiting behaviour of the proposed filtering method when $\alpha \to 2$. We get:

$$\forall j, k,\ \boldsymbol{P}_{jk} = \boldsymbol{P} = \int \boldsymbol{\theta} \boldsymbol{\theta}^\star \Gamma_{\boldsymbol{x}} \left(d\boldsymbol{\theta} \right), \tag{24}$$

which is the covariance matrix of the mixture $\boldsymbol{x}$. Indeed, exploiting the chf. representation in (1), we have:

$$\begin{aligned}\forall \boldsymbol{u} \in \mathbb{K}^K,\ \varphi_{\boldsymbol{x}}(\boldsymbol{u}) &= \exp\left(-\int |\langle \boldsymbol{u}, \boldsymbol{\theta}\rangle|^2 \, \Gamma_{\boldsymbol{x}}(d\boldsymbol{\theta})\right), \\ &= \exp\left(-\boldsymbol{u}^\star \boldsymbol{P} \boldsymbol{u}\right),\end{aligned}$$

which is the chf. of a Gaussian vector of covariance matrix $\boldsymbol{P}$. It is straightforward to show that $\boldsymbol{P} = \sum \boldsymbol{P}_j$, where $\boldsymbol{P}_j$ is the covariance matrix of the j^{th} component, given as in (24) but by integrating against Γ_j:

$$\boldsymbol{P}_j = \int \boldsymbol{\theta}\boldsymbol{\theta}^\star \Gamma_j(d\boldsymbol{\theta}). \tag{25}$$

Then, $\boldsymbol{r}_{jk}$ in (21) becomes the k^{th} column of $\boldsymbol{P}_j$. Consequently, when $\alpha \to 2$, the estimates $\widehat{\boldsymbol{y}}_j$ for the components become:

$$\widehat{\boldsymbol{y}}_j = \boldsymbol{P}_j \left(\sum_{j'} \boldsymbol{P}_{j'}\right)^{-1} \boldsymbol{x}, \tag{26}$$

which is exactly the classical *multichannel Wiener filter* (MWF), but with parameters computed by exploiting the spatial densities Γ_j. This is the linear filtering method which minimizes the MSE between the sources and their estimates when second-order moments are available. As expected, the CMF technique presented in Section 4.2 is a generalization of the MWF to α-stable distributions.

Finally, we will propose a last estimator, which is another generalization of the MWF to α-stable distributions. Because second-order moments are not defined for $\alpha < 2$, matrices $\boldsymbol{P}_j$ cannot be defined through $\boldsymbol{P}_j = \mathbb{E}\left[\boldsymbol{y}_j \boldsymbol{y}_j^\star\right]$ as in the case $\alpha = 2$. Nevertheless, it is remarkable that the expression (25) remains computable whatever $\alpha \in (1, 2)$, making it an interesting method to estimate the parameters of what becomes an *ad-hoc* filter (26), which we call MWF through an abuse of notation, since it is only equivalent to MWF when $\alpha = 2$.

5. Evaluation

In this section, we assess the performance of the three filtering methods presented in Sections 3, 4.2 and 4.3. In this regard, we stress that this paper only addresses the design of filters associated to α-stable processes with *known* parameters Γ_j. Hence, the estimation of those parameters is kept out of the scope of the present study. The reader can however find spatial measure estimation techniques in (Nolan et al. [29]) and (Pivato and Seco [32]). The rationale for disentangling filtering and parameter estimation is to provide a grounded basis for the last filtering step of whole processing pipelines involving α-stable processes, as is routinely done for Gaussian processes with the MWF (Duong et al. [11], Liutkus et al. [22]).

Consequently, this evaluation focuses on the performance of the proposed filters on synthetic data only. The procedure is always the same: firstly, we generate the realizations $\boldsymbol{y}_j$ according to multivariate symmetric α-stable distributions with known spatial densities Γ_j as in Algorithm 1, then, these realizations are summed to produce the observations $\boldsymbol{x}$ to be filtered, and the performance scores are computed by comparing the true $\boldsymbol{y}_j$ with their estimates. That said, the set of parameters considered spans a wide range of configurations of various difficulties, allowing us to assess the strengths and weaknesses of the proposed methods.

5.1. Experimental setup

Evaluated methods. We investigate the performance of $\alpha-$SSF, as presented in Section 3, and of $\alpha-$CMF with 50 iterations, as presented in Section 4.2. We compare them with the MWF method described in Section 4.3, i.e. with parameters $\boldsymbol{P}_j$ computed as in (25) using the true Γ_j.

Metrics. Because $\alpha > 1$, the relative root mean-square error does not exist (except for $\alpha = 2$). Thus, we only consider the relative mean-absolute error (MAE) defined as:

$$\mathrm{MAE}\left(\boldsymbol{y}, \widehat{\boldsymbol{y}}\right) = \frac{\sum_j \mathbb{E}\left(\left|\boldsymbol{y}_j - \widehat{\boldsymbol{y}}_j\right|\right)}{\sum_j \mathbb{E}\left(\left|\boldsymbol{y}_j\right|\right)}, \tag{27}$$

where expectations are approximated by the empirical mean.

The Von-Mises Fisher distribution. As mentioned above, we evaluate all methods in the case of known spatial densities Γ_j. As a running example, we will take the Γ_j as mixtures of Von-Mises Fisher (VMF) distributions, written $\mathcal{V}_{\boldsymbol{\mu},\kappa}$, which are defined as:

$$\mathcal{V}_{\boldsymbol{\mu},\kappa}\left(d\boldsymbol{\theta}\right) \propto \exp\left(\kappa \boldsymbol{\mu}^\top \boldsymbol{\theta}\right) d\boldsymbol{\theta}, \tag{28}$$

where $\boldsymbol{\mu} \in \mathcal{S}^K$ is the *mean direction* and $\kappa > 0$ is a *concentration parameter*, which is higher when the mass concentrates close to $\boldsymbol{\mu}$. Since the spatial densities are symmetric, we sample them on the hyper-hemisphere only. Although any other choice of a distribution over $\mathcal{S}^K$ could be made, we picked the VMF because it is the maximum entropy distribution of a random variable on the sphere with known location and spread parameters.

5.2. Performance versus the spatial distance of components

In this first round of experiments, we focus on the spatial resolution of each algorithm, because filtering out components that are spatially close has many applications, e.g. in audio processing. For this purpose, we study the filtering performance of mixtures of $J = 2$ real-valued sources of dimension $K = 2$, so that a direction $\boldsymbol{\theta} \in \mathcal{S}^K$ can be understood as a point on the unit circle.

We take each component $\boldsymbol{y}_j$ as originating mostly from one direction $\boldsymbol{\mu}_j$, with both components sharing the same concentration parameter $\kappa = 15$, so

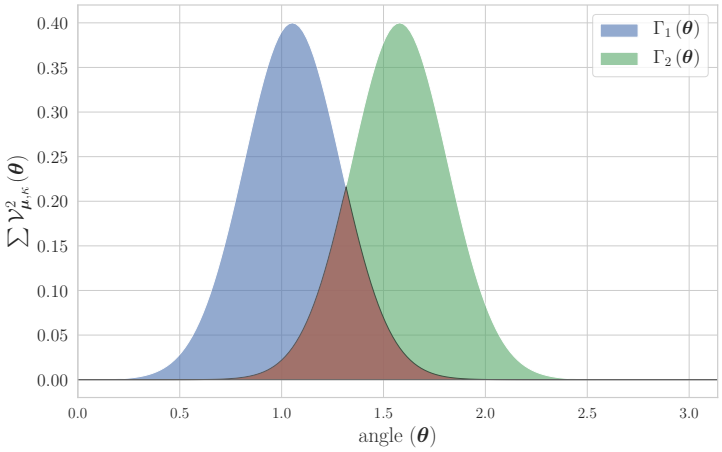


Figure 4: Two Spatial Von-Mises densities Γ_1, Γ_2 on the semicircle with respective mean directions $\mu_1 = \frac{\pi}{3}, \mu_2 = \frac{\pi}{2}$ and concentration $\kappa = 15$. The red area indicates the overlap between the spatial densities.

that $\Gamma_j = \mathcal{V}_{\boldsymbol{\mu}_j,\kappa}$. Depending on the choice of the directions $\boldsymbol{\mu}_j$, this allows us to create some overlap between the Γ_j's. We set a 0.2 step-size for $\alpha \in [1.2, 2]$, and the mean directions $\boldsymbol{\mu}_j$ for the sources are separated by $\{5, 15, \ldots, 85\}$ degrees, with $\boldsymbol{\mu}_1$ randomly positioned on the semi-circle. An example is given in Fig. 4.

Algorithms $\alpha-$SSF and $\alpha-$CMF were run with partitions Θ composed of $P = 180$ regions. For each configuration of the $\boldsymbol{\mu}_j$, the performance of all methods was evaluated on the filtering of $N = 2000$ independent realizations. 100 different such experiments were conducted to report the scores.

The corresponding MAE (27) values for $\alpha = 1.6$ are displayed in Fig. 5. As can be seen on this figure, $\alpha-$SSF globally outperforms the other methods for all the angular distances between the sources, followed by $\alpha-$CMF. While these two methods behave similarly, they both outperform MWF.

In Fig. 6, we show the evolution of these scores as a function of α, for a fixed deviation of 25 degrees between the sources. As can be seen, smaller values for α lead to a degradation of the performance, due to the extremely heavy tails of the distributions. However, we notice that all proposed methods remain quite robust. Hence, even MWF, the proposed approach that exploits the spatial densities Γ_j to build the MWF filters as discussed in Section 4.3, is remarkably effective. In practice, decreasing P often causes instability for MWF, while increasing P does not significantly change scores.

5.3. Performance versus the number of components

In this second set of experiments, we evaluate our proposed filtering methods in the complex case with $K = 2$. Our objective here is to assess the performance with a varying number J of components to separate. Hence, for $J \in \{2, \ldots, 8\}$, $\alpha \in [1.2, 2]$, and $N = 2000$ independent realizations, we run 100 independents

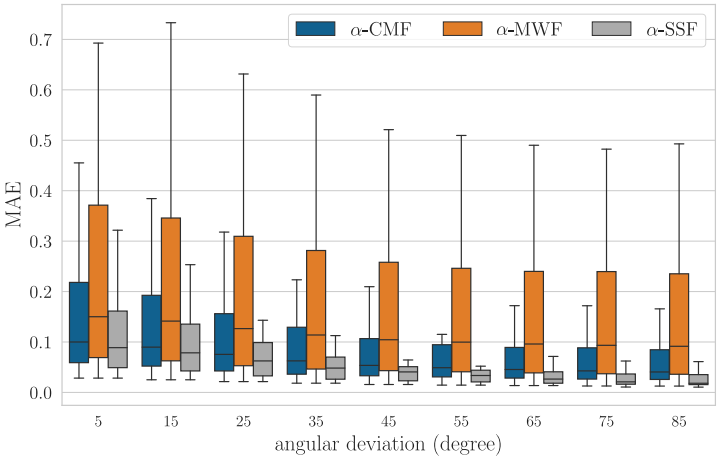


Figure 5: MAE box plot (lower is better) of MWF, $\alpha-$CMF and $\alpha-$SSF methods for several angular deviations between spatial densities and $\alpha = 1.6$. The box plots shows the minimum and maximum for whiskers, and 75th/25th percentiles for boxes.

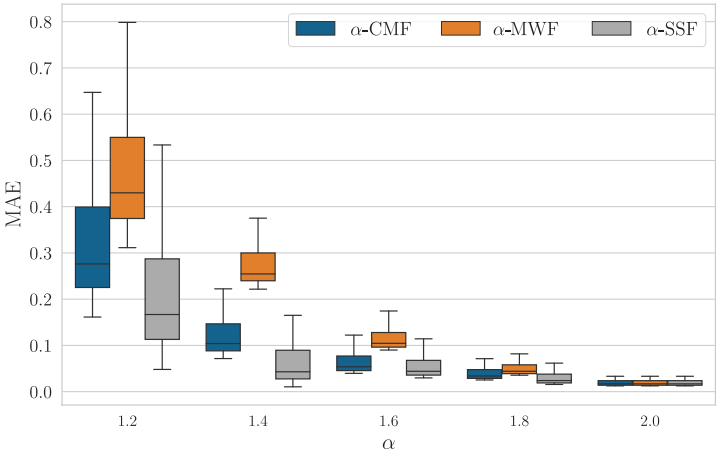


Figure 6: MAE performance (lower is better) as a function of α, for a distance of 25 degrees between the sources.

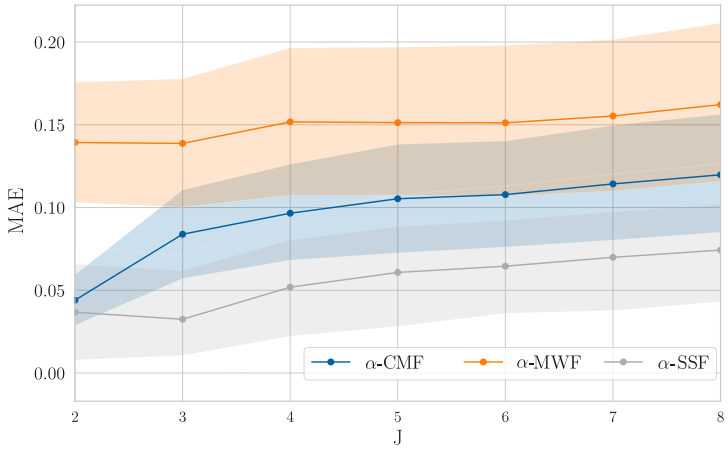


Figure 7: MAE averaged over all sources for $\alpha \in [1.2, 2]$. The solid lines display the median, and the light areas the standard deviation of each method.

experiments, where the spatial densities are VMF distributions with random parameters κ_j and $\boldsymbol{\mu}_j$.

Regarding the computation of our integrals as in (3) for this complex case, they were performed with $P = 400$ values for the partition Θ.

The results shown in Fig. 7 are in line with those for $J = 2$ reported above, and suggest that the $\alpha-$SSF method globally outperforms the three other methods for all configurations.

	MWF	$\alpha-$CMF	$\alpha-$SSF
$J = 2$	**0.02**	0.18	1.02
$J = 3$	**0.02**	0.20	1.11
$J = 5$	**0.02**	0.45	1.12
$J = 8$	**0.02**	0.65	1.16

Table 1: Elapsed time (in sec., lower is better) for each filtering method.

Now, we show that this increase of performance comes at the price of an increased computational cost. Indeed, the computational complexity of the $\alpha-$SSF method is $\mathcal{O}\left(P^4\left(P'\right)^4 K^5 N^3 + JP^3K^2N\right)$, where P' is the number of samples for the discretization of integrals in (10) wrt. the Lebesgue measure, while for $\alpha-$CMF it is $\mathcal{O}\left(IJK^4P^4N\right)$, where I denotes the number of iterations performed in Algorithm 3. The number P and P' of cells used to sample $\mathcal{S}^K$ inherently depends on K, and increase exponentially according to the curse of dimensionality (see Bellman [3]). This explains why evaluations are only performed when $K = 2$, and suggests an important research direction to scale

the proposed method to higher spatial dimensions. In Table 1, we report the average computing time for the different methods to process $N = 2000$ samples, as observed with our Python implementation running on a regular small laptop computer with an i7-4810MQ CPU and 32 GB of RAM. We observe that the MWF method is the fastest, followed by α-CMF and by α-SSF. Analyzing the methods further, we observe that a large part of the computing time for $\alpha-$SSF is spent on computing $g_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{\theta})$, so that separating additional components J doesn't yield a significant increase in computational load, as for MWF.

5.4. Filtering spatially scattered sources

In the preceding sections, we estimated components whose spatial density Γ_j was a simple VMF distribution, hence corresponding to only one direction of arrival μ_j, with some variations brought in by the concentration parameter κ_j. We now investigate more diverse spatial densities, where each (real) source is characterized by several directions of arrival. This is done by taking each Γ_j as a mixture of C VMF distributions:

$$\forall j,\ \Gamma_j(d\boldsymbol{\theta}) = \sum_c w_{jc} \mathcal{V}_{\boldsymbol{\mu}_{jc}, \kappa_{jc}}(d\boldsymbol{\theta}),$$

where $\mathcal{V}_{\boldsymbol{\mu}_{jc}, \kappa_{jc}}$ is defined in (28) and $w_{jc} \in [0, 1]$ are weight parameters, such that $\forall j,\ \sum_c w_{jc} = 1$. Examples of realizations for such models are depicted in Fig. 2 for $K = 2$.

We considered 5 regularly spaced $\alpha \in [1.2, 2]$ and the separation of $N = 2000$ i.i.d. samples from $J = 4$ components, whose spatial densities Γ_j are mixtures of $C = 2, 3, 4$ VMF distributions, with parameters w_{jc}, $\boldsymbol{\mu}_{jc}$ and κ_{jc} drawn randomly anew for each of the 100 experiments. The circle is uniformly divided into $P = 360$ arcs.

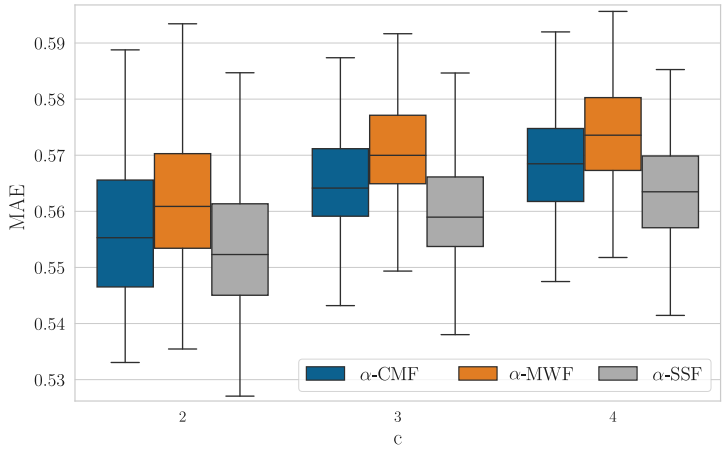


Figure 8: MAE boxplot for $\alpha = 1.4$.

Results for this experiment are depicted in Fig. 8, giving the MAE as a function of the number C of directions of arrivals for each component. We see that $\alpha-$SSF slightly outperforms the other proposed methods and that the performance is overall not so sensitive to the number of components. We believe that the slight gain brought in by α-SSF can be explained by the fact that it is a non-linear filter and may hence better handle more sophisticated spatial models.

6. Conclusion & Future works

In this paper, we showed how the multivariate symmetric α-stable ($S\alpha S$) distribution can be used in filtering applications.

An $S\alpha S$ distribution features remarkably *heavy tails,* that permit the modeling of signals with very large dynamics. As we showed, an $S\alpha S$ vector is characterized by a *spatial density*, that indicates the amount of energy originating from every direction in space. It hence naturally relaxes the common assumption of deterministic directions of arrival made for multivariate observations. One key asset of this model is then to straightforwardly extend to mixtures of such vectors, owing to the stability property of their distributions.

Equipped with such a powerful multivariate probabilistic model, we proposed several filters able to recover $S\alpha S$ vectors from the observation of their sum. The first one relies on a *spatial spectrum* decomposition and may be understood as the combination of a nonlinear beamformer followed by a scalar filter. The second one enforces *linear* filtering and minimizes the *covariation* of the difference between estimates and target. As we show, these filters generalize the classical multivariate Wiener filter to heavy-tailed signals.

Throughout the paper, we considered the separation of real and complex $S\alpha S$ vectors. A very straightforward application of these developments is the filtering of multivariate time series, via their short-time Fourier transforms. Indeed, this would simply mean generalizing the recently proposed α-harmonizable processes (Liutkus and Badeau [21]) to the multivariate case.

A natural route for future work is the combination of such filters with effective parameter estimation techniques similar to those presented in (Fontaine et al. [15]). All together, they would form a principled processing pipeline for heavy-tailed and impulsive multivariate signals. Furthermore, finding a way to proceed to the required integrations over the hypersphere without a brute-force partitioning as done here would allow the use of the proposed method in high-dimensional settings.

AppendixA. Proofs

1− Proof of spatial representation theorem

If $\mathbb{K} = \mathbb{C}$, by substituting in the theorem 6.3.4. in (Samoradnitsky and Taqqu [35, p. 284]), the terms:

$$E = \mathcal{S}^K,\ \boldsymbol{x} = \boldsymbol{\theta},\ M\left(d\boldsymbol{x}\right) = \mathcal{X}\left(d\boldsymbol{\theta}\right),\ m\left(dx\right) = \Gamma_x\left(d\boldsymbol{\theta}\right),\ d = K,$$
$$j = k, z_j = u_k, T = \left[1 \dots K\right],\ t_j = k,\ f_{t_j}\left(x\right) = x_{t_j} = \theta_k$$

we get that the chf. of $\int_{\boldsymbol{\theta}\in\mathcal{S}^K} \boldsymbol{\theta}\mathcal{X}\left(d\boldsymbol{\theta}\right)$ is:

$$\forall \boldsymbol{u}\in\mathbb{C}^K,\ \mathbb{E}\left[\exp\left(i\Re\left\langle \boldsymbol{u}, \int_{\mathcal{S}^K}\boldsymbol{\theta}\,\mathcal{X}\left(d\boldsymbol{\theta}\right)\right\rangle\right)\right] = \exp\left(-\int_{\mathcal{S}^K}\left|\langle \boldsymbol{u},\boldsymbol{\theta}\rangle\right|^{\alpha}\Gamma_x\left(d\boldsymbol{\theta}\right)\right). \tag{A.1}$$

The right side of (A.1) is exactly the chf. of $\boldsymbol{x}$. This identification achieves the proof of Theorem 4.

If $\mathbb{K}=\mathbb{R}$, (2) can be demonstrated by applying the theorem 3.5.6. in (Samoradnitsky and Taqqu [35, p 131]).

2– Proof of spatial spectrum estimation

Proof of Proposition 5

Let $\forall v\in\mathbb{K}$, $c\left(v\right)\triangleq\varphi_{\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\psi(\boldsymbol{\theta})\mathcal{X}(d\boldsymbol{\theta})|\boldsymbol{x}}\left(v\right)$. Note that if the first derivative exists (in the sense of the Wirtinger derivatives), we have:

$$\mathbb{E}\left[\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\psi\left(\boldsymbol{\theta}\right)\mathcal{X}\left(d\boldsymbol{\theta}\right)\mid\boldsymbol{x}\right] = -i\frac{dc}{dv^{\star}}\left(v=0\right). \tag{A.2}$$

In order to find a suitable form of c, we start by calculating the joint chf. of $\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\psi\left(\boldsymbol{\theta}\right)\mathcal{X}\left(d\boldsymbol{\theta}\right)$ and $\boldsymbol{x}$, $\forall v\in\mathbb{K}$, $\forall\boldsymbol{u}\in\mathbb{K}^K$:

$$\begin{aligned}\varphi_{\left(\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\psi(\boldsymbol{\theta})\mathcal{X}(d\boldsymbol{\theta}),\boldsymbol{x}\right)}\left(v,\boldsymbol{u}\right) &\triangleq \mathbb{E}\left[e^{i\Re\left(v^*\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\psi(\boldsymbol{\theta})\mathcal{X}(d\boldsymbol{\theta})+\langle\boldsymbol{u},\boldsymbol{x}\rangle\right)}\right]\\ &= \mathbb{E}\left[e^{i\Re\left(\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\left(v^*\psi(\boldsymbol{\theta})+\langle\boldsymbol{u},\boldsymbol{\theta}\rangle\right)\mathcal{X}(d\boldsymbol{\theta})\right)}\right]\\ &= \varphi_{\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\left(v^*\psi(\boldsymbol{\theta})+\langle\boldsymbol{u},\boldsymbol{\theta}\rangle\right)\mathcal{X}(d\boldsymbol{\theta})}\left(1\right)\\ &= e^{-\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\left|v^*\psi(\boldsymbol{\theta})+\langle\boldsymbol{u},\boldsymbol{\theta}\rangle\right|^{\alpha}\Gamma_{\boldsymbol{x}}(d\boldsymbol{\theta})},\end{aligned}$$

where the last equality holds because

$$\begin{aligned}&\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\left(v^*\psi\left(\boldsymbol{\theta}\right)+\langle\boldsymbol{u},\boldsymbol{\theta}\rangle\right)\,\mathcal{X}\left(d\boldsymbol{\theta}\right)\\ &\sim S\alpha S_c^K\left(\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\left|v^*\psi(\boldsymbol{\theta})+\langle\boldsymbol{u},\boldsymbol{\theta}\rangle\right|^{\alpha}\Gamma_{\boldsymbol{x}}\left(d\boldsymbol{\theta}\right)\right).\end{aligned}$$

Thus, c has the following form: $\forall v\in\mathbb{K}$,

$$\begin{aligned}c(v) &= \frac{\int_{\boldsymbol{u}\in\mathbb{K}^K}\varphi_{\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\psi(\boldsymbol{\theta})\mathcal{X}(d\boldsymbol{\theta}),\boldsymbol{x}}\left(v,\boldsymbol{u}\right)e^{-i\Re(\langle\boldsymbol{u},\boldsymbol{x}\rangle)}d\boldsymbol{u}}{\int_{\boldsymbol{u}\in\mathbb{K}^K}\varphi_{\boldsymbol{x}}\left(\boldsymbol{u}\right)e^{-i\Re(\langle\boldsymbol{u},\boldsymbol{x}\rangle)}d\boldsymbol{u}}\\ &= \frac{\int_{\boldsymbol{u}\in\mathbb{K}^K}e^{-\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\left|v^*\psi(\boldsymbol{\theta})+\langle\boldsymbol{u},\boldsymbol{\theta}\rangle\right|^{\alpha}\Gamma_{\boldsymbol{x}}(d\boldsymbol{\theta})}e^{-i\Re(\langle\boldsymbol{u},\boldsymbol{x}\rangle)}d\boldsymbol{u}}{\int_{\boldsymbol{u}\in\mathbb{K}^K}\varphi_{\boldsymbol{x}}\left(\boldsymbol{u}\right)e^{-i\Re(\langle\boldsymbol{u},\boldsymbol{x}\rangle)}d\boldsymbol{u}}.\end{aligned} \tag{A.3}$$

The existence of $\frac{dc}{dv^{\star}}\left(v=0\right)$ is because

$$\nu\mapsto\int_{\boldsymbol{u}\in\mathbb{K}^K}\varphi_{\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\psi(\boldsymbol{\theta})\mathcal{X}(d\boldsymbol{\theta}),\boldsymbol{x}}\left(v,\boldsymbol{u}\right)e^{-i\Re(\langle\boldsymbol{u},\boldsymbol{x}\rangle)}d\boldsymbol{u}$$

is the Fourier transform of a characteristic function whose first derivative exists because $\alpha>1$ (if a random vector admits a 1^{st} order moment, then its characteristic function is continuously differentiable at zero). Consequently, by using

the differentiation under the integral sign theorem (the domination is induced by the fact that a chf. is bounded by 1) and by combining (A.2) and (A.3) we obtain:

$$\begin{aligned} & \mathbb{E}\left[\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\psi\left(\boldsymbol{\theta}\right)\mathcal{X}\left(d\boldsymbol{\theta}\right)\mid\boldsymbol{x}\right] \\ = & \frac{\alpha i\int_{\boldsymbol{u}\in\mathbb{K}^K}\left(\int_{\boldsymbol{\theta}\in\mathcal{S}^K}\psi(\boldsymbol{\theta})\langle\boldsymbol{\theta},\boldsymbol{u}\rangle^{\langle\alpha-1\rangle}\Gamma_{\boldsymbol{x}}(d\boldsymbol{\theta})\right)\varphi_{\boldsymbol{x}}(\boldsymbol{u})e^{-i\Re(\langle\boldsymbol{u},\boldsymbol{x}\rangle)}du}{\int_{\boldsymbol{u}\in\mathbb{K}^K}\varphi_{\boldsymbol{x}}(\boldsymbol{u})e^{-i\Re(\langle\boldsymbol{u},\boldsymbol{x}\rangle)}d\boldsymbol{u}} \\ = & \int_{\boldsymbol{\theta}\in\mathcal{S}^K}\psi\left(\boldsymbol{\theta}\right)\left[\frac{\alpha i\int_{\boldsymbol{u}\in\mathbb{K}^K}\langle\boldsymbol{\theta},\boldsymbol{u}\rangle^{\langle\alpha-1\rangle}\varphi_{\boldsymbol{x}}(\boldsymbol{u})\,e^{-i\Re(\langle\boldsymbol{u},\boldsymbol{x}\rangle)}d\boldsymbol{u}}{\int_{\boldsymbol{u}\in\mathbb{K}^K}\varphi_{\boldsymbol{x}}(\boldsymbol{u})\,e^{-i\Re(\langle\boldsymbol{u},\boldsymbol{x}\rangle)}d\boldsymbol{u}}\right]\Gamma_{\boldsymbol{x}}\left(d\boldsymbol{\theta}\right) \\ = & \int_{\boldsymbol{\theta}\in\mathcal{S}^K}\psi\left(\boldsymbol{\theta}\right)g_{_{\mathcal{X}}}\left(\boldsymbol{x},\boldsymbol{\theta}\right)\Gamma_{\boldsymbol{x}}\left(d\boldsymbol{\theta}\right). \end{aligned} \tag{A.4}$$

where we have used (7). Equation (6) is obtained by identifying (A.4) with (5) for any continuous function $\psi\left(\boldsymbol{\theta}\right)$.

Proof of Proposition 5 about continuity

$\forall\boldsymbol{u}\in\mathbb{K}^K$, $\boldsymbol{\theta}\mapsto h\left(\boldsymbol{\theta},\boldsymbol{u}\right)=\langle\boldsymbol{\theta},\boldsymbol{u}\rangle^{\langle\alpha-1\rangle}\varphi_{\boldsymbol{x}}\left(\boldsymbol{u}\right)e^{-i\Re(\langle\boldsymbol{u},\boldsymbol{x}\rangle)}$ is a continuous function. Moreover, the probability density function (pdf.) of an $S\alpha S_c$ (non-degenerated) distribution is infinitely continuous. As a result, the characteristic function $\varphi_{\boldsymbol{x}}\left(\boldsymbol{u}\right)$ for all $\boldsymbol{u}\in\mathbb{K}^K$(which is the Fourier transform of a pdf.) decreases faster than any power of $\|u\|$ (where $\|.\|$ is any norm on $\mathbb{K}^K$). In particular, $\boldsymbol{u}\mapsto f\left(\boldsymbol{u}\right)=\|\boldsymbol{u}\|^{\alpha-1}\left|\varphi_x\left(\boldsymbol{u}\right)\right|$ is integrable. Besides, $\forall\boldsymbol{\theta}\in\mathcal{S}^K$, $\left|\langle\boldsymbol{\theta},\boldsymbol{u}\rangle^{\langle\alpha-1\rangle}\right|\leq\|\boldsymbol{u}\|^{\alpha-1}$ and $\forall\boldsymbol{u}\in\mathbb{K}^K$, $\left|h\left(\boldsymbol{\theta},\boldsymbol{u}\right)\right|\leq f\left(\boldsymbol{u}\right)$. By applying the theorem of continuity under the integral sign, we conclude that $g_{_{\mathcal{X}}}\left(\boldsymbol{x},\boldsymbol{\theta}\right)$ is a continuous function of $\boldsymbol{\theta}$.

Proof of Proposition 6

We consider the numerator $N\left(\boldsymbol{x},\boldsymbol{\theta}\right)$ of $g_{_{\mathcal{X}}}\left(\boldsymbol{x},\boldsymbol{\theta}\right)$ in (8) and apply the change of variables $\boldsymbol{u}\triangleq r\boldsymbol{\theta'}\in\mathbb{K}^K$ where $r\in\mathbb{R}^+$and $\boldsymbol{\theta'}\in\mathcal{S}^K$, which is such that $d\boldsymbol{u}=r^{\beta K-1}drd\boldsymbol{\theta'}$. Then we get:

$$N\left(\boldsymbol{x},\boldsymbol{\theta}\right)=i\int_{\boldsymbol{\theta'}\in\mathcal{S}^K}\left\langle\boldsymbol{\theta},\boldsymbol{\theta'}\right\rangle^{\langle\alpha-1\rangle}\left(\int_{r\in\mathbb{R}_+}\left(\alpha r^{\alpha-1}e^{-r^\alpha I_{\boldsymbol{x}}(\boldsymbol{\theta'})}\right)\left(r^{\beta K-1}e^{-ir\Re(\langle\boldsymbol{\theta'},\boldsymbol{x}\rangle)}\right)dr\right)d\boldsymbol{\theta'}.$$

Applying an integration by parts to $\int_{r\in\mathbb{R}_+}\left(\alpha r^{\alpha-1}e^{-r^\alpha I_{\boldsymbol{x}}(\boldsymbol{\theta'})}\right)\left(r^{\beta K-1}e^{-ir\Re(\langle\boldsymbol{\theta'},\boldsymbol{x}\rangle)}\right)dr$ yields:

$$N\left(\boldsymbol{x},\boldsymbol{\theta}\right)=A\left(\boldsymbol{\theta},\boldsymbol{x}\right)+\left(\beta K-1\right)iB\left(\boldsymbol{\theta},\boldsymbol{x}\right)$$

where

$$A(\boldsymbol{\theta},\boldsymbol{x}) = \int_{\boldsymbol{\theta}'\in\mathcal{S}^K} \frac{\langle\boldsymbol{\theta},\boldsymbol{\theta}'\rangle^{\langle\alpha-1\rangle}\Re\left(\langle\boldsymbol{\theta}',\boldsymbol{x}\rangle\right)}{I_{\boldsymbol{x}}(\boldsymbol{\theta}')} \left(\int_{r\in\mathbb{R}_+} r^{\beta K-1}e^{-r^\alpha I_{\boldsymbol{x}}(\boldsymbol{\theta}')}e^{-ir\Re(\langle\boldsymbol{\theta}',\boldsymbol{x}\rangle)}dr\right)d\boldsymbol{\theta}'$$

$$B(\boldsymbol{\theta},\boldsymbol{x}) = \int_{\boldsymbol{\theta}'\in\mathcal{S}^K} \frac{\langle\boldsymbol{\theta},\boldsymbol{\theta}'\rangle^{\langle\alpha-1\rangle}}{I_{\boldsymbol{x}}(\boldsymbol{\theta}')} \left(\int_{r\in\mathbb{R}_+} r^{\beta K-2}e^{-r^\alpha I_{\boldsymbol{x}}(\boldsymbol{\theta}')}e^{-ir\Re(\langle\boldsymbol{\theta}',\boldsymbol{x}\rangle)}dr\right)d\boldsymbol{\theta}'.$$

By developing the complex exponential as a power series and by applying the identity $\int_0^{+\infty} x^n e^{-ax^b}dx = \frac{1}{b}a^{-\frac{n+1}{b}}f_\Gamma\left(\frac{n+1}{b}\right)$ to both integrals $A(\boldsymbol{\theta},\boldsymbol{x})$ and $B(\boldsymbol{\theta},\boldsymbol{x})$, we get:

$$A(\boldsymbol{\theta},\boldsymbol{x}) = \frac{1}{\alpha}\int_{\boldsymbol{\theta}'\in\mathcal{S}^K}\sum_{n=0}^{+\infty}\frac{(-i)^n f_\Gamma\left(\frac{n+\beta K}{\alpha}\right)}{n!}\frac{\langle\boldsymbol{\theta},\boldsymbol{\theta}'\rangle^{<\alpha-1>}\frac{1}{2^{n+1}}\mathcal{B}_{n+1}\left(\langle\boldsymbol{\theta}',\boldsymbol{x}\rangle,\langle\boldsymbol{x},\boldsymbol{\theta}'\rangle\right)}{I_{\boldsymbol{x}}(\boldsymbol{\theta}')^{\frac{n+\beta K+\alpha}{\alpha}}}d\boldsymbol{\theta}' \tag{A.5}$$

$$B(\boldsymbol{\theta},\boldsymbol{x}) = \frac{1}{\alpha}\int_{\boldsymbol{\theta}'\in\mathcal{S}^K}\sum_{n=0}^{+\infty}\frac{(-i)^n f_\Gamma\left(\frac{n+\beta K-1}{\alpha}\right)}{n!}\frac{\langle\boldsymbol{\theta},\boldsymbol{\theta}'\rangle^{\langle\alpha-1\rangle}\frac{1}{2^n}\mathcal{B}_n\left(\langle\boldsymbol{\theta}',\boldsymbol{x}\rangle,\langle\boldsymbol{x},\boldsymbol{\theta}'\rangle\right)}{I_{\boldsymbol{x}}(\boldsymbol{\theta}')^{\frac{n+\beta K-1+\alpha}{\alpha}}}d\boldsymbol{\theta}'. \tag{A.6}$$

where $\mathcal{B}_n(u,v) = (u+v)^n = \sum_{k=0}^n \binom{n}{k}u^n v^{n-k}$. Finally, we remark from (8) that $N(z\boldsymbol{x},\boldsymbol{\theta}) = zN(\boldsymbol{x},\boldsymbol{\theta}), \forall z\in\mathcal{S}^1_{\mathbb{K}}$, which shows that in (A.5), all odd values of n, and all values of k different from $\frac{n}{2}+1$, can be discarded (the corresponding terms vanish when integrated). In the same way, in (A.6), all even values of n, and all values of k different from $\frac{n+1}{2}$, can be discarded, which finally leads to (10) and (11). Equation (11) defines a power series with an infinite radius of convergence when $\alpha > 1$, which is smooth and independent the mixing model.

The estimate of $\boldsymbol{x}\to D(\boldsymbol{x})$ is obtained by noting that:

$$\boldsymbol{x} = \int_{\boldsymbol{\theta}\in\mathcal{S}^K}\boldsymbol{\theta}g_{\mathcal{X}}(\boldsymbol{x},\boldsymbol{\theta})\Gamma_x(d\boldsymbol{\theta}) = \int_{\boldsymbol{\theta}\in\mathcal{S}^K}\boldsymbol{\theta}\frac{N(\boldsymbol{x},\boldsymbol{\theta})}{D(\boldsymbol{x})}\Gamma_{\boldsymbol{x}}(d\boldsymbol{\theta}).$$

Proof of Theorem 7

We first determine the joint chf. of $\boldsymbol{y}_j$ and $\boldsymbol{x}$: $\forall\boldsymbol{v},\boldsymbol{u}\in\mathbb{K}^K$,

$$\begin{aligned}\varphi_{(\boldsymbol{y}_j,\boldsymbol{x})}(\boldsymbol{v},\boldsymbol{u}) &\triangleq \mathbb{E}\left[e^{i\Re\left(\langle\boldsymbol{v},\boldsymbol{y}_j\rangle+\langle\boldsymbol{u},\boldsymbol{x}\rangle\right)}\right]\\ &= \mathbb{E}\left[e^{i\Re\left(\langle\boldsymbol{v}+\boldsymbol{u},\boldsymbol{y}_j\rangle+\sum_{j'\neq j}\langle\boldsymbol{u},\boldsymbol{y}_{j'}\rangle\right)}\right]\\ &= \varphi_{\boldsymbol{y}_j}(\boldsymbol{v}+\boldsymbol{u})\prod_{j'\neq j}\varphi_{\boldsymbol{y}_{j'}}(\boldsymbol{u})\\ &= e^{-\int_{\boldsymbol{\theta}\in\mathcal{S}^K}|\langle\boldsymbol{v}+\boldsymbol{u},\boldsymbol{\theta}\rangle|^\alpha\Gamma_j(d\boldsymbol{\theta})-\sum_{j'\neq j}\int_{\boldsymbol{\theta}\in\mathcal{S}^K}|\langle\boldsymbol{u},\boldsymbol{\theta}\rangle|^\alpha\Gamma_{j'}(d\boldsymbol{\theta})},\end{aligned}$$

which permits us to deduce the chf. of $\boldsymbol{y}_j$ given $\boldsymbol{x}$, $\forall \boldsymbol{v} \in \mathbb{K}^K$:

$$\varphi_{\boldsymbol{y}_j|\boldsymbol{x}}\left(\boldsymbol{v}\right) = \frac{\int_{\boldsymbol{u}\in\mathbb{K}^K} e^{-\int_{\boldsymbol{\theta}\in\mathcal{S}^K} |\langle \boldsymbol{v}+\boldsymbol{u},\boldsymbol{\theta}\rangle|^{\alpha}\Gamma_j(d\boldsymbol{\theta}) - \sum_{j'\neq j}\int_{\boldsymbol{\theta}\in\mathcal{S}^K}|\langle \boldsymbol{u},\boldsymbol{\theta}\rangle|^{\alpha}\Gamma_{j'}(d\boldsymbol{\theta})} e^{-i\Re(\langle \boldsymbol{u},\boldsymbol{x}\rangle)} d\boldsymbol{u}}{\int_{\boldsymbol{u}\in\mathbb{K}^K}\varphi_{\boldsymbol{x}}(\boldsymbol{u}) e^{-i\Re(\langle \boldsymbol{u},\boldsymbol{x}\rangle)} d\boldsymbol{u}}.$$

The proof of the existence of $\nabla_{\boldsymbol{v}^\star}\varphi_{\boldsymbol{y}_j|\boldsymbol{x}}$ is exactly the same as in Proposition 5. Thus, we get:

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{y}_j \mid \boldsymbol{x}\right] &= -i\nabla_{\boldsymbol{v}^\star}\varphi_{\boldsymbol{y}_j|\boldsymbol{x}}\left(\boldsymbol{v}=\boldsymbol{0}\right) \\
&= \int_{\boldsymbol{\theta}\in\mathcal{S}^K} \boldsymbol{\theta}\left[\frac{\alpha i \int_{\boldsymbol{u}\in\mathbb{K}^K}\langle \boldsymbol{u},\boldsymbol{\theta}\rangle^{\langle\alpha-1\rangle}\varphi_{\boldsymbol{x}}\left(\boldsymbol{u}\right) e^{-i\Re(\langle \boldsymbol{u},\boldsymbol{x}\rangle)} d\boldsymbol{u}}{\int_{\boldsymbol{u}\in\mathbb{K}^K}\varphi_{\boldsymbol{x}}\left(\boldsymbol{u}\right) e^{-i\Re(\langle \boldsymbol{u},\boldsymbol{x}\rangle)} d\boldsymbol{u}}\right]\Gamma_j\left(d\boldsymbol{\theta}\right) \\
&= \int_{\boldsymbol{\theta}\in\mathcal{S}^K} \boldsymbol{\theta}\, g_{\mathcal{X}}\left(\boldsymbol{x},\boldsymbol{\theta}\right)\Gamma_j\left(d\boldsymbol{\theta}\right),
\end{aligned}$$

which completes the proof.

---

[1] Achim, A., Kuruoglu, E. E., Zerubia, J., 2006. SAR image filtering based on the heavy-tailed Rayleigh model 15 (9), 2686–2693.

[2] Bassiou, N., Kotropoulos, C., Pitas, I., 2014. Greek folk music denoising under a symmetric α-stable noise assumption. In: 10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine). IEEE, pp. 18–23.

[3] Bellman, R., 2013. Dynamic programming. Courier Corporation.

[4] Benaroya, L., Bimbot, F., Gribonval, R., 2006. Audio source separation with a single sensor. IEEE Transactions on Audio, Speech, and Language Processing 14 (1), 191–199.

[5] Boyd, S., Vandenberghe, L., 2004. Convex optimization. Cambridge university press.

[6] Cambanis, S., Huang, S., Simons, G., 1981. On the theory of elliptically contoured distributions. Journal of Multivariate Analysis 11 (3), 368–385.

[7] Cardoso, J.-F., 1998. Blind signal separation: statistical principles 86 (10), 2009–2025.

[8] Cavalcant, Y., Oberlin, T., Dobigeon, N., Févotte, C., Stute, S., Taube, C., 2019. Unmixing dynamic pet images: combining spatial heterogeneity and non-gaussian noise. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 1373–1377.

[9] Comon, P., Jutten, C., 2010. Handbook of Blind Source Separation: Independent component analysis and applications. Academic press.

[10] Damon, C., Liutkus, A., Gramfort, A., Essid, S., 2013. Non-negative matrix factorization for single-channel eeg artifact rejection. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 1177–1181.

[11] Duong, N. Q. K., Vincent, E., Gribonval, R., Sept. 2010. Under-determined reverberant audio source separation using a full-rank spatial covariance model 18 (7), 1830 –1840.

[12] Duvenaud, D., Nickisch, H., Rasmussen, C., 2011. Additive gaussian processes. In: Advances in neural information processing systems. pp. 226–234.

[13] Febrero-Bande, M., González-Manteiga, W., 2013. Generalized additive models for functional data. Test 22 (2), 278–292.

[14] Fontaine, M., Liutkus, A., Girin, L., Badeau, R., 2017. Explaining the parameterized Wiener filter with alpha-stable processes. In: Proc. of Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). IEEE.

[15] Fontaine, M., Vanwynsberghe, C., Liutkus, A., Badeau, R., Aug. 2017. Scalable source localization with multichannel alpha-stable distributions. In: Proc. of 25th European Signal Processing Conference (EUSIPCO). pp. 11–15.

[16] Godsill, S., Rayner, P., Cappé, O., 2002. Digital audio restoration. Applications of digital signal processing to audio and acoustics, 133–194.

[17] Godsill, S. J., Rayner, P. J. W., 1995. A Bayesian approach to the restoration of degraded audio signals 3 (4), 267–278.

[18] Hardin J., C. D., 1982. On the spectral representation of symmetric stable processes. Journal of Multivariate Analysis 12 (3), 385–401.

[19] Kidmose, P., 2001. Blind separation of heavy tail signals. Ph.D. thesis, Institute of Mathematical Modelling, Technical University of Denmark, Kongens Lyngby, Denmark.

[20] Kitamura, K., Bando, Y., Itoyama, K., Yoshii, K., 2016. Student's t multichannel nonnegative matrix factorization for blind source separation. In: 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC). IEEE, pp. 1–5.

[21] Liutkus, A., Badeau, R., 2015. Generalized Wiener filtering with fractional power spectrograms. In: Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 266–270.

[22] Liutkus, A., Badeau, R., Richard, G., 2011. Gaussian processes for underdetermined source separation 59 (7), 3155–3167.

[23] Ma, X., Nikias, C., 1995. Parameter estimation and blind channel identification in impulsive signal environments. IEEE Transactions on signal processing 43 (12), 2884–2897.

[24] Mandelbrot, B., May 1960. The Pareto-Lévy law and the distribution of income. International Economic Review 1 (2), 79–106.

[25] Marra, G., Wood, S., 2011. Practical variable selection for generalized additive models. Computational Statistics & Data Analysis 55 (7), 2372–2387.

[26] Masry, E., 2000. Alpha-stable signals and adaptive filtering 48 (11), 3011–3016.

[27] Moussaoui, S., Hauksdottir, H., Schmidt, F., Jutten, C., Chanussot, J., Brie, D., Douté, S., Benediktsson, J. A., 2008. On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation. Neurocomputing 71 (10-12), 2194–2208.

[28] Nikias, C. L., Shao, M., 1995. Signal processing with alpha-stable distributions and applications. Adaptive and learning systems for signal processing, communications, and control. Wiley.

[29] Nolan, J. P., Panorska, A. K., McCulloch, J. H., 2001. Estimation of stable spectral measures. Mathematical and Computer Modelling 34 (9), 1113–1122.

[30] Nugraha, A., Liutkus, A., Vincent, E., 2018. Deep neural network based multichannel audio source separation. In: Audio Source Separation. Springer, pp. 157–185.

[31] Ozerov, A., Févotte, C., Vincent, E., 2018. An introduction to multichannel nmf for audio source separation. In: Audio Source Separation. Springer, pp. 73–94.

[32] Pivato, M., Seco, L., 2003. Estimating the spectral measure of a multivariate stable distribution via spherical harmonic analysis. Journal of Multivariate Analysis 87 (2), 219–240.

[33] Rafii, Z., Liutkus, A., Stoter, F.-R., Mimilakis, S., FitzGerald, D., Pardo, B., 2018. An overview of lead and accompaniment separation in music. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP) 26 (8), 1307–1335.

[34] Roth, M., Özkan, E., Gustafsson, F., 2013. A Student's t filter for heavy tailed process and measurement noise. In: Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 5770–5774.

[35] Samoradnitsky, G., Taqqu, M., 1994. Stable non-Gaussian random processes: stochastic models with infinite variance. Vol. 1. CRC Press.

[36] Stuck, B. W., Kleiner, B., 1974. A statistical analysis of telephone noise. Bell Labs Technical Journal 53 (7), 1263–1320.

[37] Unser, M., Tafti, P., 2014. An introduction to sparse stochastic processes. Cambridge University Press.

[38] Venkataramani, S., Casebeer, J., Smaragdis, P., 2018. End-to-end source separation with adaptive front-ends. In: 2018 52nd Asilomar Conference on Signals, Systems, and Computers. IEEE, pp. 684–688.

[39] Vincent, E., Virtanen, T., Gannot, S., 2018. Audio source separation and speech enhancement. John Wiley & Sons.

[40] Wang, B., Kuruoglu, E., Zhang, J., 2009. Ica by maximizing non-stability. In: International Conference on Independent Component Analysis and Signal Separation. Springer, pp. 179–186.

[41] Wang, D., Zhang, C., Zhao, X., 2008. Multivariate Laplace filter: a heavy-tailed model for target tracking. In: Proc. of 19th International Conference on Pattern Recognition (ICPR). IEEE, pp. 1–4.

[42] Wang, Z.-Q., Roux, J. L., Wang, D., Hershey, J., 2018. End-to-end speech separation with unfolded iterative phase reconstruction. arXiv preprint arXiv:1804.10204.

[43] Wiener, N., 1949. Extrapolation, interpolation, and smoothing of stationary time series. Vol. 7. MIT press Cambridge, MA.

[44] Wood, S., Goude, Y., Shaw, S., 2015. Generalized additive models for large data sets. Journal of the Royal Statistical Society: Series C (Applied Statistics) 64 (1), 139–155.

[45] Yoshii, K., Itoyama, K., Goto, M., 2016. Student's t nonnegative matrix factorization and positive semidefinite tensor factorization for single-channel audio source separation. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 51–55.