SUPPORTING DOCUMENT FOR ALPHA-STABLE MULTICHANNEL AUDIO SOURCE SEPARATION

Simon Leglaive[1]*, Umut Şimşekli*[1]*, Antoine Liutkus*[2]*, Roland Badeau*[1]*, Gaël Richard*[1]

1: LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France
2: Inria, Speech Processing Team, Villers-lès-Nancy, France

This document provides additional materials for the reference paper [1], in which we propose a probabilistic model for multichannel audio source separation based on a class of heavy-tailed distributions. In this approach the observed mixtures and the latent sources are jointly modeled by using the complex multivariate elliptically contoured stable distribution. The inference procedure for estimating the source signals relies on a Monte Carlo Expectation Maximization (MCEM) algorithm. The aim of the present document is to give further details on the derivation of this algorithm.

1. REMINDER OF THE MODEL

In this section we briefly describe the model introduced in [1]. We work in the Short-Time Fourier Transform (STFT) domain. For all $(f, n) \in \{0, ..., F-1\} \times \{0, ..., N-1\}$, we consider J audio sources signals $\mathbf{s}_{fn} \in \mathbb{C}^J$ mixed together to form I mixtures $\mathbf{x}_{fn} \in \mathbb{C}^I$. The observations and the sources are jointly modeled as following an elliptically contoured stable distribution [2], denoted by $\mathcal{E}\alpha\mathcal{S}_c$. Using the fact that $\mathcal{E}\alpha\mathcal{S}_c$ distributions are conditionally Gaussian if we introduce an *impulse variable* $\phi_{fn} \in \mathbb{R}_+$ [3] we have:

$$\begin{aligned} \phi_{fn} &\sim \mathcal{P}\frac{\alpha}{2}\mathcal{S}\left(2\Big(\cos\frac{\pi\alpha}{4}\Big)^{2/\alpha}\right), \\ \begin{pmatrix}\mathbf{x}_{fn}\\ \mathbf{s}_{fn}\end{pmatrix} |\phi_{fn} &\sim \mathcal{N}_c(0, \phi_{fn}\boldsymbol{\Sigma}_{fn}), \end{aligned} \tag{1}$$

where $\mathcal{P}\frac{\alpha}{2}\mathcal{S}$ denotes the positive α-stable distribution (see [1] for more details), with $\alpha \in (0, 2)$, and $\mathcal{N}_c$ denotes the multivariate complex isotropic Gaussian distribution. $\boldsymbol{\Sigma}_{fn} \in \mathbb{C}^{(I+J)\times(I+J)}$ is a positive definite matrix called the *shape matrix* and has the following structure:

$$\boldsymbol{\Sigma}_{fn} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{x},fn} & \mathbf{A}_f\boldsymbol{\Sigma}_{\mathbf{s},fn} \\ \boldsymbol{\Sigma}_{\mathbf{s},fn}\mathbf{A}_f^\star & \boldsymbol{\Sigma}_{\mathbf{s},fn} \end{pmatrix}, \tag{2}$$

where

$$\boldsymbol{\Sigma}_{\mathbf{x},fn} = \mathbf{A}_f\boldsymbol{\Sigma}_{\mathbf{s},fn}\mathbf{A}_f^\star + \boldsymbol{\Sigma}_{\mathbf{b},f}. \tag{3}$$

and $\cdot^\star$ denotes Hermitian conjugation. $\mathbf{A}_f = [a_{ij,f}]_{ij} \in \mathbb{C}^{I\times J}$ is called the mixing matrix, $\boldsymbol{\Sigma}_{\mathbf{b},f} = \sigma_{b,f}^2\mathbf{I}_I$ with $\sigma_{b,f}^2 > 0$, and $\mathbf{I}_I$ is the identity matrix of size $I \times I$. The source covariance matrices $\boldsymbol{\Sigma}_{\mathbf{s},fn}$ are further parametrized by using a Non-negative Matrix Factorization (NMF) model:

$$\boldsymbol{\Sigma}_{\mathbf{s},fn} = \mathrm{diag}([v_{j,fn}]_j), \text{ with } v_{j,fn} = [\mathbf{W}_j\mathbf{H}_j]_{fn}, \tag{4}$$

where $\mathbf{W}_j \in \mathbb{R}_+^{F\times K_j}$ and $\mathbf{H}_j \in \mathbb{R}_+^{K_j\times N}$ are called the dictionary and activation matrices of source j, respectively.

2. DERIVATION DETAILS FOR THE MONTE CARLO EXPECTATION-MAXIMIZATION ALGORITHM

In this section we give further details on the derivation of Monte Carlo Expectation-Maximization (MCEM) algorithm presented in [1]. Let $\mathbf{X} = \{\mathbf{x}_{fn}\}_{f,n}$ be the set of observed data while $\mathbf{S} = \{\mathbf{s}_{fn}\}_{f,n}$ and $\boldsymbol{\Phi} = \{\phi_{fn}\}_{f,n}$ denote the set of hidden variables. $\boldsymbol{\Theta} = \{\{\mathbf{W}_j, \mathbf{H}_j\}_j, \{\boldsymbol{\Sigma}_{\mathbf{b},f}, \mathbf{A}_f\}_f\}$ denotes the set of parameters to be estimated.

2.1. E-step

At the E-step of the MCEM algorithm we have to compute the conditional expectation of the complete data log-likelihood:

$$\begin{aligned} Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}') &= \mathbb{E}_{\mathbf{S},\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}\Big[\ln p(\mathbf{X},\mathbf{S},\boldsymbol{\Phi}|\boldsymbol{\Theta})\Big] \\ &= \mathbb{E}_{\mathbf{S},\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}\Big[\ln p(\mathbf{X}|\mathbf{S},\boldsymbol{\Phi};\boldsymbol{\Theta}) + \ln p(\mathbf{S}|\boldsymbol{\Phi};\boldsymbol{\Theta}) + \ln p(\boldsymbol{\Phi}|\boldsymbol{\Theta})\Big] \\ &\stackrel{c}{=} -\sum_{f=0}^{F-1}\sum_{n=0}^{N-1}\mathbb{E}_{\mathbf{S},\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}\Big[\ln\det(\pi\phi_{fn}\boldsymbol{\Sigma}_{\mathbf{b},f}) + (\mathbf{x}_{fn}-\mathbf{A}_f\mathbf{s}_{fn})^\star(\phi_{fn}\boldsymbol{\Sigma}_{\mathbf{b},f})^{-1}(\mathbf{x}_{fn}-\mathbf{A}_f\mathbf{s}_{fn})\Big] \\ &\quad -\sum_{j=1}^{J}\sum_{f=0}^{F-1}\sum_{n=0}^{N-1}\mathbb{E}_{\mathbf{S},\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}\left[\ln(\pi\phi_{fn}v_{j,fn}) + \frac{|s_{j,fn}|^2}{\phi_{fn}v_{j,fn}}\right] \\ &\stackrel{c}{=} -N\sum_{f=0}^{F-1}\Big[\ln\det(\boldsymbol{\Sigma}_{\mathbf{b},f}) + \mathrm{tr}\Big(\boldsymbol{\Sigma}_{\mathbf{b},f}^{-1}\hat{\mathbf{R}}_{\phi\mathbf{x}\mathbf{x},f} - \boldsymbol{\Sigma}_{\mathbf{b},f}^{-1}\mathbf{A}_f\hat{\mathbf{R}}_{\phi\mathbf{x}\mathbf{s},f}^\star - \boldsymbol{\Sigma}_{\mathbf{b},f}^{-1}\hat{\mathbf{R}}_{\phi\mathbf{x}\mathbf{s},f}\mathbf{A}_f^\star + \boldsymbol{\Sigma}_{\mathbf{b},f}^{-1}\mathbf{A}_f\hat{\mathbf{R}}_{\phi\mathbf{s}\mathbf{s},f}\mathbf{A}_f^\star\Big)\Big] \\ &\quad -\sum_{j=1}^{J}\sum_{f=0}^{F-1}\sum_{n=0}^{N-1}\left[\ln(v_{j,fn}) + \frac{\hat{p}_{j,fn}}{v_{j,fn}}\right], \end{aligned} \tag{5}$$

where $\stackrel{c}{=}$ denotes equality up to an additive constant and

- ▷ $\hat{\mathbf{R}}_{\phi\mathbf{x}\mathbf{x},fn} = \mathbb{E}_{\mathbf{S},\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}[\phi_{fn}^{-1}\mathbf{x}_{fn}\mathbf{x}_{fn}^\star]$;
- ▷ $\hat{\mathbf{R}}_{\phi\mathbf{x}\mathbf{s},fn} = \mathbb{E}_{\mathbf{S},\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}[\phi_{fn}^{-1}\mathbf{x}_{fn}\mathbf{s}_{fn}^\star]$;
- ▷ $\hat{\mathbf{R}}_{\phi\mathbf{s}\mathbf{s},fn} = \mathbb{E}_{\mathbf{S},\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}[\phi_{fn}^{-1}\mathbf{s}_{fn}\mathbf{s}_{fn}^\star]$;
- ▷ $\hat{p}_{j,fn} = \mathbb{E}_{\mathbf{S},\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}[\phi_{fn}^{-1}|s_{j,fn}|^2]$;
- ▷ $\hat{\mathbf{R}}_{\phi\cdot\cdot,f} = \dfrac{1}{N}\sum_{n=0}^{N-1}\hat{\mathbf{R}}_{\phi\cdot\cdot,fn}$.

We can further write the previous conditional expectations as:

$$\mathbb{E}_{\mathbf{S},\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}[\cdot] = \mathbb{E}_{\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}[\mathbb{E}_{\mathbf{S}|\boldsymbol{\Phi},\mathbf{X},\boldsymbol{\Theta}'}[\cdot]]. \tag{6}$$

Moreover, from (1) to (3), the conditional distribution according to which the inner expectation in (6) is taken is given as follows:

$$\mathbf{s}_{fn}|\mathbf{x}_{fn},\phi_{fn};\boldsymbol{\Theta} \sim \mathcal{N}_c\Big(\hat{\mathbf{s}}_{fn}, \phi_{fn}\boldsymbol{\Sigma}_{\mathbf{s},fn}^{cond}\Big), \tag{7}$$

with $\hat{\mathbf{s}}_{fn} = \boldsymbol{\Sigma}_{\mathbf{s},fn}\mathbf{A}_f^\star\boldsymbol{\Sigma}_{\mathbf{x},fn}^{-1}\mathbf{x}_{fn}$ and $\boldsymbol{\Sigma}_{\mathbf{s},fn}^{cond} = (\mathbf{I}_J - \boldsymbol{\Sigma}_{\mathbf{s},fn}\mathbf{A}_f^\star\boldsymbol{\Sigma}_{\mathbf{x},fn}^{-1}\mathbf{A}_f)\boldsymbol{\Sigma}_{\mathbf{s},fn}$. Interestingly, $\hat{\mathbf{s}}_{fn}$ does not depend on ϕ_{fn}. Finally, from (6) and (7) the statistics in (5) can be further written as:

- ▷ $\hat{\mathbf{R}}_{\phi\mathbf{x}\mathbf{x},fn} = \mathbb{E}_{\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}[\phi_{fn}^{-1}]\mathbf{x}_{fn}\mathbf{x}_{fn}^\star$;
- ▷ $\hat{\mathbf{R}}_{\phi\mathbf{x}\mathbf{s},fn} = \mathbf{x}_{fn}\mathbb{E}_{\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}\Big[\phi_{fn}^{-1}\mathbb{E}_{\mathbf{S}|\boldsymbol{\Phi},\mathbf{X},\boldsymbol{\Theta}'}\big[\mathbf{s}_{fn}^\star\big]\Big] = \mathbf{x}_{fn}\hat{\mathbf{s}}_{fn}^\star\mathbb{E}_{\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}[\phi_{fn}^{-1}]$;
- ▷ $\hat{\mathbf{R}}_{\phi\mathbf{s}\mathbf{s},fn} = \mathbb{E}_{\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}\Big[\phi_{fn}^{-1}\mathbb{E}_{\mathbf{S}|\boldsymbol{\Phi},\mathbf{X},\boldsymbol{\Theta}'}\big[\mathbf{s}_{fn}\mathbf{s}_{fn}^\star\big]\Big] = \boldsymbol{\Sigma}_{\mathbf{s},fn}^{cond} + \hat{\mathbf{s}}_{fn}\hat{\mathbf{s}}_{fn}^\star\mathbb{E}_{\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}[\phi_{fn}^{-1}]$;
- ▷ $\hat{p}_{j,fn} = \mathbb{E}_{\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}\Big[\phi_{fn}^{-1}\mathbb{E}_{\mathbf{S}|\boldsymbol{\Phi},\mathbf{X},\boldsymbol{\Theta}'}\big[|s_{j,fn}|^2\big]\Big] = \Big[\hat{\mathbf{R}}_{\phi\mathbf{s}\mathbf{s},fn}\Big]_{j,j}$.

2.2. Markov Chain Monte Carlo algorithm

The previous statistics computed at the E-step involve the computation of $\mathbb{E}_{\boldsymbol{\Phi}|\mathbf{X},\boldsymbol{\Theta}'}[\phi_{fn}^{-1}]$. This is achieved in [1] using a Markov Chain Monte Carlo (MCMC) algorithm. More precisely, we develop a Metropolis-Hastings (MH) algorithm that generates samples from $\pi(\phi_{fn}) = p(\phi_{fn}|\mathbf{X},\boldsymbol{\Theta})$. As explained in [1], this MH algorithm relies on the computation of the following acceptance probability:

$$\mathrm{acc}(\phi_{fn} \to \phi'_{fn}) = \min\left\{1, \frac{q(\phi_{fn})\pi(\phi'_{fn})}{q(\phi'_{fn})\pi(\phi_{fn})}\right\}, \tag{8}$$

where $q(\phi_{fn})$ is called the proposal distribution. In practice we choose the prior distribution of ϕ_{fn} as the proposal distribution: $q(\phi_{fn}) = p(\phi_{fn})$. The acceptance probability in (8) can thus be computed according to:

$$\begin{aligned}
\ln \frac{q(\phi_{fn})\pi(\phi'_{fn})}{q(\phi'_{fn})\pi(\phi_{fn})} &= \ln \frac{p(\phi_{fn})p(\phi'_{fn}|\mathbf{x}_{fn};\boldsymbol{\Theta})}{p(\phi'_{fn})p(\phi_{fn}|\mathbf{x}_{fn};\boldsymbol{\Theta})} \\
&= \ln \frac{p(\mathbf{x}_{fn}|\phi'_{fn};\boldsymbol{\Theta})p(\phi'_{fn})p(\phi_{fn})}{p(\mathbf{x}_{fn}|\phi_{fn};\boldsymbol{\Theta})p(\phi_{fn})p(\phi'_{fn})} \\
&= \ln \frac{N_c(\mathbf{x}_{fn}; 0, \phi'_{fn}\boldsymbol{\Sigma}_{\mathbf{x},fn})}{N_c(\mathbf{x}_{fn}; 0, \phi_{fn}\boldsymbol{\Sigma}_{\mathbf{x},fn})} \\
&= -\ln\det(\pi\phi'_{fn}\boldsymbol{\Sigma}_{\mathbf{x},fn}) - \mathbf{x}^\star_{fn}(\phi'_{fn}\boldsymbol{\Sigma}_{\mathbf{x},fn})^{-1}\mathbf{x}_{fn} + \ln\det(\pi\phi_{fn}\boldsymbol{\Sigma}_{\mathbf{x},fn}) + \mathbf{x}^\star_{fn}(\phi_{fn}\boldsymbol{\Sigma}_{\mathbf{x},fn})^{-1}\mathbf{x}_{fn} \\
&= -I\ln\phi'_{fn} - \frac{1}{\phi'_{fn}}\mathbf{x}^\star_{fn}\boldsymbol{\Sigma}^{-1}_{\mathbf{x},fn}\mathbf{x}_{fn} + I\ln\phi_{fn} + \frac{1}{\phi_{fn}}\mathbf{x}^\star_{fn}\boldsymbol{\Sigma}^{-1}_{\mathbf{x},fn}\mathbf{x}_{fn} \\
&= I\ln\frac{\phi_{fn}}{\phi'_{fn}} + \left(\frac{1}{\phi_{fn}} - \frac{1}{\phi'_{fn}}\right)\mathbf{x}^\star_{fn}\boldsymbol{\Sigma}^{-1}_{\mathbf{x},fn}\mathbf{x}_{fn}. \qquad (9)
\end{aligned}$$

2.3. M-step

The M-step aims to maximize the function $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}')$ given in (5), in order to obtain a new estimate of the parameters. Zeroing the gradient of $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}')$ with respect to the mixing matrix $\mathbf{A}_f$ and the noise shape matrix $\boldsymbol{\Sigma}_{\mathbf{b},f}$ leads to the updates given in [1]. The expression of $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}')$ in (5) involves the Itakura-Saito (IS) divergence [4] between $\hat{p}_{j,fn}$ and $v_{j,fn} = [\mathbf{W}_j\mathbf{H}_j]_{fn}$. Therefore, the matrices $\mathbf{W}_j$ and $\mathbf{H}_j$ can be updated by solving the following optimization problem:

$$\mathbf{W}_j, \mathbf{H}_j = \underset{\mathbf{W}_j,\mathbf{H}_j \geq 0}{\arg\min} \sum_{f=0}^{F-1}\sum_{n=0}^{N-1} d_{IS}\big(\hat{p}_{j,fn}, [\mathbf{W}_j\mathbf{H}_j]_{fn}\big), \qquad (10)$$

where the IS divergence is given by:

$$d_{IS}(x,y) = \frac{x}{y} - \ln\frac{x}{y} - 1. \qquad (11)$$

The minimization problem (10) can be solved efficiently by using the standard multiplicative update rules given in [4].

3. REFERENCES

[1] S. Leglaive, U. Şimşekli, A. Liutkus, R. Badeau, and G. Richard, "Alpha-stable multichannel audio source separation," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, 2017.

[2] J. P. Nolan, "Multivariate elliptically contoured stable distributions: theory and estimation," *Computational Statistics*, vol. 28, no. 5, pp. 2067–2089, 2013.

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

[4] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.