α-MNMF with a Spatial Measure Representation

October 5, 2020

1 The model

Let assume N, F, T, M be the number of sources, frequency bins, time frame and microphones respectively. The sources are admiting the following probablistic model:

$$\boldsymbol{x}_{nft} \sim \mathcal{S}\alpha S_{\mathbb{C}}\left(\mathfrak{C}_{nft}\right)$$

with (the next result is true because of the Radon-Nikodym-Lebesgue theorem applied on a positive measure):

$$\mathfrak{C}_{nft}\left(d\boldsymbol{\theta}\right) \triangleq \lambda_{nft}\Gamma_{nf}\left(d\boldsymbol{\theta}\right) \triangleq \sum_{k} w_{nfk}h_{nkt}\Gamma_{nf}\left(d\boldsymbol{\theta}\right)$$

and where $d\boldsymbol{\theta}$ is a small portion of the hypersphere $\mathcal{S}_{\mathbb{C}}^{M}$. Then we have

$$\boldsymbol{x}_{nft} \sim \mathcal{S}\alpha S_{\mathbb{C}}\left(\lambda_{nft}\boldsymbol{\Gamma}_{nf}\right)$$

and:

$$\boldsymbol{x}_{ft} \triangleq \sum_{n=1}^{N} \boldsymbol{x}_{nft} \sim \mathcal{S}\alpha S_{\mathbb{C}}\left(\sum_{n=1}^{N} \lambda_{nft}\boldsymbol{\Gamma}_{nf}\right)$$

In order to estimate λ_{nft} and Γ_{nf}, we use an EM approach.

2 EM algorithm

Let's introduce some notations:

- P: number of sphere sampling
- $\boldsymbol{\theta}_{f1}, \ldots, \boldsymbol{\theta}_{fP}$: sphere sampling
- $\tilde{\mathbf{x}}_{ft} \in \mathbb{R}_{+}^{P}$: The Levy exponent where $\tilde{\mathbf{x}}_{ft} = \left[\tilde{x}_{ft1}, \ldots, \tilde{x}_{ftP}\right]^{\top}$

Theoretically, we have the following equality:

$$\tilde{\mathbf{x}}_{ft}\left(\boldsymbol{\theta}\right)=\sum_{n}\int_{\boldsymbol{\theta'}_{f}}\left|\langle\boldsymbol{\theta}_{f},\boldsymbol{\theta}'_{f}\rangle\right|^{\alpha}\lambda_{nft}\boldsymbol{\Gamma}_{nf}\left(d\boldsymbol{\theta}'_{f}\right)$$

An estimation of $\tilde{\mathbf{x}}_{f[t-\Delta t,t+\Delta t]}$ (**Levy exponent estimator** along the interval $[t-\Delta t,t+\Delta t]$) is given as follow:

$$\forall\boldsymbol{\theta}_{p}\in\mathbb{C}^{K},\tilde{\boldsymbol{x}}_{f[t-\Delta t,t+\Delta t]}\left(\boldsymbol{\theta}_{p}\right)\triangleq\tilde{x}_{ftp}\simeq-2\log\left|\frac{1}{2\Delta_{t}}\sum_{t\in[t'-\Delta t,t'+\Delta t]}\exp\left(i\frac{\Re\left[\boldsymbol{\theta}_{p}^{\star}\boldsymbol{x}_{ft'}\right]}{2^{1/\alpha}}\right)\right|$$

The estimator will be considered by doing a moving average along the time axis.

2.1 M-Step

It can be shown that it exists the following relation:

$$\tilde{\boldsymbol{x}}_{ft}\simeq\boldsymbol{\Psi}_{f}\sum_{n}\lambda_{nft}\hat{\boldsymbol{\Gamma}}_{nf}$$

where for the entry p,p' of the matrix $\boldsymbol{\Psi}$ we set

$$\left[\boldsymbol{\Psi}_{f}\right]_{p,p'}\triangleq\left|\langle\boldsymbol{\theta}_{fp},\boldsymbol{\theta}_{fp'}\rangle\right|^{\alpha}$$

We consider the β-divergence as a cost function (for $\beta\leq1$):

$$\begin{aligned}
&d_{\beta}\left(\tilde{\boldsymbol{x}}_{ft}\mid\sum_{n=1}^{N}\lambda_{nft}\boldsymbol{\Psi}_{f}\hat{\boldsymbol{\Gamma}}_{nf}\right)\\
&=\sum_{f,t,p}\frac{1}{\beta\left(\beta-1\right)}\left(\left\{\tilde{\boldsymbol{x}}_{ft}\right\}_{p}^{\beta}+\left(\beta-1\right)\left[\sum_{n,k=1}^{N,K}w_{nfk}h_{nkt}\left\{\boldsymbol{\Psi}_{f}\hat{\boldsymbol{\Gamma}}_{nf}\right\}_{p}\right]^{\beta}-\beta\left\{\tilde{\boldsymbol{x}}_{ft}\right\}_{p}\left[\sum_{n,k=1}^{N,K}w_{nfk}h_{nkt}\left\{\boldsymbol{\Psi}_{f}\hat{\boldsymbol{\Gamma}}_{nf}\right\}_{p}\right]^{\beta-1}\right)\\
&\leq\sum_{f,t,p}\left(\beta\left(\beta-1\right)\tilde{x}_{ftp}^{\beta}+\pi_{ftp}^{\beta-1}\left(\sum_{n,k=1}^{N,K}w_{nfk}h_{nkt}\tilde{g}_{nfp}-\pi_{ftp}\right)+\frac{\pi_{ftp}^{\beta}}{\beta}-\frac{1}{\beta-1}\tilde{x}_{ftp}\sum_{n,k=1}^{N,K}\omega_{ftnkp}\left(\frac{w_{nfk}^{\beta-1}h_{nkt}^{\beta-1}\tilde{g}_{nfp}^{\beta-1}}{\omega_{ftnkp}^{\beta-1}}\right)\right)\\
&\leq\sum_{f,t,p}\left(\beta\left(\beta-1\right)\tilde{x}_{ftp}^{\beta}+\pi_{ftp}^{\beta-1}\left(\sum_{n,k,p'=1}^{N,K,P}w_{nfk}h_{nkt}\psi_{fpp'}\gamma_{nfp'}-\pi_{ftp}\right)+\frac{\pi_{ftp}^{\beta}}{\beta}-\frac{1}{\beta-1}\tilde{x}_{ftp}\sum_{n,k,p'=1}^{N,K,P}\omega_{ftnkp}^{2-\beta}\rho_{nfpp'}^{2-\beta}w_{nfk}^{\beta-1}h_{nkt}^{\beta-1}\psi_{fpp'}^{\beta-1}\gamma_{nfp}^{\beta-1}\right.\\
&\triangleq\mathcal{L}_{+}\left(\tilde{\boldsymbol{x}}_{f}\mid\sum_{n=1}^{N}\lambda_{nft}\boldsymbol{\Psi}_{f}\hat{\boldsymbol{\Gamma}}_{nf},\boldsymbol{\Pi},\boldsymbol{\Omega},\rho\right)
\end{aligned}$$

with $y_{ftp}=\sum_{n,k}w_{nkf}h_{nkt}\left\{\boldsymbol{\Psi}_{f}\hat{\boldsymbol{\Gamma}}_{nf}\right\}_{p}$ and $\tilde{g}_{nfp}=\left\{\boldsymbol{\Psi}_{f}\hat{\boldsymbol{\Gamma}}_{nf}\right\}_{p}=\sum_{p'}\psi_{fpp'}\gamma_{nfp'}$.
The equalities hold when

$$\omega_{ftnkp}=w_{nkf}h_{nkt}\left\{\boldsymbol{\Psi}_{f}\hat{\boldsymbol{\Gamma}}_{nf}\right\}_{p}\left[\sum_{n,k}w_{nkf}h_{nkt}\left\{\boldsymbol{\Psi}_{f}\hat{\boldsymbol{\Gamma}}_{nf}\right\}_{p}\right]^{-1}\triangleq w_{nkf}h_{nkt}\tilde{g}_{nfp}y_{ftp}^{-1}$$

$$\pi_{ftp}=y_{ftp},\rho_{nfpp'}=\psi_{fpp'}\gamma_{nfp'}\tilde{g}_{nfp}^{-1}$$

2.1.1 Estimation of Γ_{nf}

We assume that λ_{nft} is known. We can then derive $\mathcal{L}_+\left(\tilde{\boldsymbol{x}}_{ft} \mid \sum_{n=1}^{N}\lambda_{nft}\boldsymbol{\Psi}_f\hat{\boldsymbol{\Gamma}}_{nf},\boldsymbol{\Pi}\right)$ along $\hat{\boldsymbol{\Gamma}}_{nf}$ to get:

$$\begin{aligned}
&\frac{\partial\mathcal{L}_+\left(\tilde{\boldsymbol{x}}_{ft} \mid \sum_{n=1}^{N}\lambda_{nft}\boldsymbol{\Psi}_f\hat{\boldsymbol{\Gamma}}_{nf},\boldsymbol{\Pi},\boldsymbol{\Omega}\right)}{\partial\gamma_{nfp''}}\\
&=\sum_{t,p,k}\left(\pi_{ftp}^{\beta-1}w_{nfk}h_{nkt}\psi_{fpp''}-\tilde{x}_{ftp}\omega_{ftnkp}^{2-\beta}\rho_{nfpp''}^{2-\beta}w_{nfk}^{\beta-1}h_{nkt}^{\beta-1}\psi_{fpp''}^{\beta-1}\gamma_{nfp''}^{\beta-2}\right)
\end{aligned}$$

we zeroing and get:

$$\sum_{t,p,k}\pi_{ftp}^{\beta-1}w_{nfk}h_{nkt}\psi_{fpp''}=\sum_{t,p,k}\tilde{x}_{ftp}\omega_{ftnkp}^{2-\beta}\rho_{nfpp''}^{2-\beta}w_{nfk}^{\beta-1}h_{nkt}^{\beta-1}\psi_{fpp''}^{\beta-1}\gamma_{nfp''}^{\beta-2}$$

i.e:

$$\gamma_{nfp''}\leftarrow\left(\frac{\sum_{t,p}\tilde{x}_{ftp}\omega_{ftnkp}^{2-\beta}\rho_{nfpp''}^{2-\beta}w_{nfk}^{\beta-1}h_{nkt}^{\beta-1}\psi_{fpp''}^{\beta-1}}{\sum_{t,p,k}\pi_{ftp}^{\beta-1}w_{nfk}h_{nkt}\psi_{fpp''}}\right)^{e(\beta)}$$

with substition of the auxiliary variables, we get:

$$\gamma_{nfp''}\leftarrow\gamma_{nfp''}.\left(\frac{\sum_{t,p}\tilde{x}_{ftp}\lambda_{nft}\tilde{g}_{nfp}^{2-\beta}y_{ftp}^{\beta-2}\psi_{pp''}^{2-\beta}\tilde{g}_{nfp}^{\beta-2}\psi_{fpp''}^{\beta-1}}{\sum_{t,p,k}y_{ftp}^{\beta-1}w_{nfk}h_{nkt}\psi_{fpp''}}\right)^{e(\beta)}$$

i.e.:

$$\gamma_{nfp''}\leftarrow\gamma_{nfp''}.\left(\frac{\sum_{t,p}\tilde{x}_{ftp}\lambda_{nft}y_{ftp}^{\beta-2}\psi_{fpp''}}{\sum_{t,p}y_{ftp}^{\beta-1}\lambda_{nft}\psi_{fpp''}}\right)^{e(\beta)}$$

$$\hat{\boldsymbol{\Gamma}}_{nf}\leftarrow\hat{\boldsymbol{\Gamma}}_{nf}.\left(\frac{\sum_t\lambda_{nft}\boldsymbol{\Psi}_f^{\top}\left(\boldsymbol{y}_{ft}^{\odot\beta-2}\odot\tilde{\boldsymbol{x}}_{ft}\right)}{\sum_t\lambda_{nft}\boldsymbol{\Psi}_f^{\top}\boldsymbol{y}_{ft}^{\odot\beta-1}}\right)^{\odot e(\beta)} \tag{1}$$

where

$$e(\beta)=\begin{cases}\frac{1}{2-\beta} & \text{if } \beta<1\\ 1 & \text{if } 1\le\beta\le 2\end{cases}$$

and

$$\begin{aligned}\boldsymbol{y}_{ft}&\triangleq\left[y_{ft1},\ldots,y_{ftP}\right]^{\top}\\&=\sum_n\lambda_{nft}\boldsymbol{\Psi}\hat{\boldsymbol{\Gamma}}_{nf}\end{aligned}$$

.

2.1.2 Estimation of λ_{nft}

We assume that the spatial measures Γ_{nf} are known. We get:

$$\frac{\partial \mathcal{L}_+ \left(\tilde{\boldsymbol{x}}_{ft} \mid \sum_{n=1}^{N} \lambda_{nft} \boldsymbol{\Psi} \hat{\boldsymbol{\Gamma}}_{nf}, \boldsymbol{\Pi}, \boldsymbol{\Omega} \right)}{\partial w_{nfk}} = \sum_{t,p} \left(\pi_{ftp}^{\beta-1} h_{nkt} \tilde{g}_{nfp} - \sum_{p'} \tilde{x}_{ftp} \omega_{ftnkp}^{2-\beta} \rho_{nfpp'}^{2-\beta} w_{nfk}^{\beta-2} h_{nkt}^{\beta-1} \psi_{fpp'}^{\beta-1} \gamma_{nfp'}^{\beta-1} \right)$$

i.e we have:

$$\sum_{t,p=1}^{T,P} \pi_{ftp}^{\beta-1} h_{nkt} \tilde{g}_{nfp} = w_{nfk}^{\beta-2} \sum_{t,p,p'=1}^{T,P,P} \tilde{x}_{fp} \omega_{ftnkp}^{2-\beta} \rho_{nfpp'}^{2-\beta} h_{nkt}^{\beta-1} \psi_{fpp'}^{\beta-1} \gamma_{nfp'}^{\beta-1}$$

$$w_{nfk} \leftarrow \left(\frac{\sum_{t,p,p'} \omega_{ftnkp}^{2-\beta} \rho_{nfpp'}^{2-\beta} h_{nkt}^{\beta-1} \psi_{fpp'}^{\beta-1} \gamma_{nfp'}^{\beta-1} \tilde{x}_{fp}}{\sum_{t,p} h_{nkt} \pi_{ftp}^{\beta-1} \tilde{g}_{nfp}} \right)^{e(\beta)} \tag{2}$$

ie we get:

$$w_{nfk} \leftarrow w_{nfk} \left(\frac{\sum_{t,p} h_{nkt} y_{ftp}^{\beta-2} \tilde{g}_{nfp} \tilde{x}_{fp}}{\sum_{t,p} h_{nkt} \pi_{ftp}^{\beta-1} \tilde{g}_{nfp}} \right)^{e(\beta)}$$

$$h_{nkt} \leftarrow h_{nkt} \left(\frac{\sum_{f,p} w_{nfk} y_{ftp}^{\beta-2} \tilde{g}_{nfp} \tilde{x}_{fp}}{\sum_{f,p} w_{nfk} y_{ftp}^{\beta-1} \tilde{g}_{nfp}} \right)^{e(\beta)} \tag{3}$$

2.2 E Step

A simple and less time consuming filtering than using covariation distance, directly deriving from a linear form of the posterior $\mathbb{E}\left[\boldsymbol{y}_{nft} \mid \boldsymbol{x}_{ft}\right]$, is:

$$\hat{y}_{nft} = \boldsymbol{W}_{nft} \boldsymbol{x}_{ft}$$

where

$$\boldsymbol{W}_{nft} = \lambda_{nft} M \int_{\boldsymbol{\theta}} \Xi_{ft}(\boldsymbol{\theta}) \Gamma_{nf}(d\boldsymbol{\theta}) \tag{4}$$

with:

$$\Xi_{ft}(\boldsymbol{\theta}) = \boldsymbol{\theta} \left(\frac{\int_{\boldsymbol{\theta}'} \boldsymbol{\theta}' \frac{\langle \boldsymbol{\theta}', \boldsymbol{\theta} \rangle^{\langle \alpha-1 \rangle}}{\tilde{\boldsymbol{x}}_{ft}(\boldsymbol{\theta}')^{\frac{2M+\alpha}{\alpha}}} d\boldsymbol{\theta}'}{\int_{\boldsymbol{\theta}'} \tilde{\boldsymbol{x}}_{ft}(\boldsymbol{\theta}')^{-\frac{2M}{\alpha}} d\boldsymbol{\theta}'} \right)^{\mathrm{H}} \tag{5}$$

we can rewrite $\Xi_{ft}(\boldsymbol{\theta})$ as:

$$\Xi_{ft}(\boldsymbol{\theta}) \simeq \boldsymbol{\theta} \left(\frac{\int_{\boldsymbol{\theta}'} \boldsymbol{\theta}' \frac{\langle \boldsymbol{\theta}', \boldsymbol{\theta} \rangle^{\langle \alpha-1 \rangle}}{\left(\sum_n \lambda_{nft} \boldsymbol{\Psi}_f \hat{\boldsymbol{\Gamma}}_{nf} \right)^{\frac{2M+\alpha}{\alpha}}} d\boldsymbol{\theta}'}{\int_{\boldsymbol{\theta}'} \left(\sum_n \lambda_{nft} \boldsymbol{\Psi}_f \hat{\boldsymbol{\Gamma}}_{nf} \right)^{-\frac{2M}{\alpha}} d\boldsymbol{\theta}'} \right)^{\mathrm{H}}$$

3 Acoustic Model

The main problem for the proposed method is about $\boldsymbol{\theta}_{fp}$. We have to sample the hypersphere of dimension $2M$... We can instead consider the $\boldsymbol{\theta}_{fp}$ as a steering vector. If a farfield region assumption is for instance assumed, we put:

$$[\boldsymbol{\theta}_{fp}]_m = \frac{1}{r_{mp}} \exp\left(-\frac{i\omega_f r_{mp}}{c_0}\right)$$

where r_{mp} is the (euclidean) distance between the microphone m and a point r, ω_f the angular frequency of f and c_0 the speed of sound in the air. In this case, we sample $\mathbb{R}^3$ and not S^{2M}.

Other extensions: A farfield region assumption where we sample S^2 (azimuth and elevation $(\theta, \varphi) \in [0, 2\pi] \times [0, \pi]$).

3.1 Oracle test

We first investigate the algorithm with the wsj0-mix dataset. $N = 2$ speakers in an anechoic environment and $M = 2$ microphones. We assume the microphones and sources positions to be known. Only two steering vectors corresponding to p_1 and p_2 the positions of speaker 1 and speaker 2 respectively are computed. The algorithm is set as follow:

- $\beta = 0$ (Itakura-Saito divergence), $\Delta_t = 4$ for the moving average of Levy exponent parameter along the time axis, $K = 32$ NMF bases and $\alpha = 1.4$.
- The NMF parameters are randomly initialized as the absolute value of a gaussian sampling.
- 500 iterations for the M-Step.
- The spatial measures are initialized as oracle ($\Gamma_{1f} = [1, 0]^\top$, $\Gamma_{2f} = [0, 1]^\top$) i.e. as a dirac measure (make sense for an anechoic model).
- we normalize the NMF coefficients + spatial measure
- scalar ambiguity in the estimated logPSD

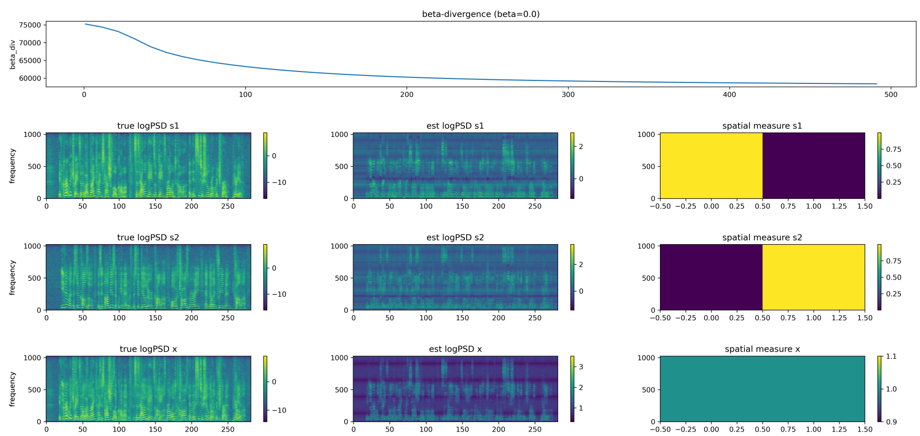

Figure 1: log-spectrogram of the target **(left heatmap)** and estimated **(middle heatmap)** sources. The spatial measure are displayed on the right column.