SUPPORTING DOCUMENT

SEPARATING TIME-FREQUENCY SOURCES FROM TIME-DOMAIN CONVOLUTIVE MIXTURES USING NON-NEGATIVE MATRIX FACTORIZATION

Simon Leglaive, Roland Badeau, Gaël Richard

LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France

This document provides additional calculation details for the variational expectation-maximization (VEM) algorithm presented in [1].

1. REMINDER OF THE MODEL

Let us define:

- ▷ $s_j(t)$, $t = 0, ..., L_s - 1$, $j = 1, ..., J$, the j-th source signal;
- ▷ $a_{ij}(t)$, $t = 0, ..., L_a - 1$, $i = 1, ..., I$, the mixing filter between source j and microphone i;
- ▷ $y_{ij}(t) = [a_{ij} \star s_j](t)$, $t = 0, ..., T - 1$ with $T = L_s + L_a - 1$, the j-th source image seen by the i-th microphone.

The signal $x_i(t)$ recorded by the i-th microphone is modeled as:

$$x_i(t) = \sum_{j=1}^{J} y_{ij}(t) + b_i(t), \tag{1}$$

where $b_i(t) \sim \mathcal{N}_{\mathbb{R}}(0, \sigma_i^2)$. The probability density function (pdf) of $\mathcal{N}_{\mathbb{R}}$ is defined in Appendix A.

A source signal $s_j(t)$ is represented by a set of time-frequency (TF) synthesis coefficients $\{s_{j,fn} \in \mathbb{K} = \mathbb{C} \text{ or } \mathbb{R}\}_{f,n}$ for $(f, n) \in \mathcal{B}$ with $\mathcal{B} = \{0, ..., F - 1\} \times \{0, ..., N - 1\}$:

$$s_j(t) = \frac{2}{\phi} \Re \left(\sum\nolimits_{(f,n) \in \mathcal{B}} s_{j,fn} \psi_{fn}(t) \right), \tag{2}$$

where $\Re(\cdot)$ denotes the real part and $\psi_{fn}(t)$ is a TF synthesis atom that corresponds to the modified discrete cosine transform (MDCT) if $\mathbb{K} = \mathbb{R}$ (in this case $\phi = 2$) and to the odd-frequency short-time Fourier transform (OFSTFT) if $\mathbb{K} = \mathbb{C}$ (in this case $\phi = 1$). From this model a source image can be further written as follows:

$$y_{ij}(t) = \frac{2}{\phi} \Re \left(\sum\nolimits_{(f,n) \in \mathcal{B}} s_{j,fn} g_{ij,fn}(t) \right), \tag{3}$$

where $g_{ij,fn}(t) = [a_{ij} \star \psi_{fn}](t)$.

The synthesis coefficients $s_{j,fn}$ are then modeled as centered and real Gaussian random variables if $\mathbb{K} = \mathbb{R}$ ($\phi = 2$) or complex circularly symmetric Gaussian random variables if $\mathbb{K} = \mathbb{C}$ ($\phi = 1$):

$$s_{j,fn} \sim \begin{cases} \mathcal{N}_{\mathbb{R}}(0, v_{j,fn}) & \text{if } \mathbb{K} = \mathbb{R}; \\ \mathcal{N}_{\mathbb{C}}^{p}(0, v_{j,fn}) & \text{if } \mathbb{K} = \mathbb{C}, \end{cases} \tag{4}$$

where the pdfs of these distributions are provided in Appendix A. The variances $v_{j,fn} \in \mathbb{R}_+$ are finally structured by means of an NMF model:

$$v_{j,fn} = [\mathbf{W}_j \mathbf{H}_j]_{fn}, \tag{5}$$

with $\mathbf{W}_j \in \mathbb{R}_+^{F \times K_j}$, $\mathbf{H}_j \in \mathbb{R}_+^{K_j \times N}$ and K_j is the rank of the factorization.

2. VARIATIONAL INFERENCE

Let $\mathbf{x} = \{x_i(t)\}_{i,t}$ denote the set of observed mixture variables, $\mathbf{s} = \{s_{j,fn}\}_{j,f,n}$ the latent source variables and $\boldsymbol{\theta} = \{\{\sigma_i^2\}_i, \{\mathbf{W}_j, \mathbf{H}_j\}_j\}$ the unknown model parameters. The mixing filters $\{a_{ij}(t)\}_{i,j,t}$ are assumed to be known. Let $q \in \mathcal{F}$ be a pdf over $\mathbf{s}$, where $\mathcal{F}$ is a variational family. Variational inference consists in optimizing a criterion called the variational free energy and defined as [2]:

$$\mathcal{L}(q; \boldsymbol{\theta}) = \left\langle \ln \left(p(\mathbf{x}, \mathbf{s}; \boldsymbol{\theta}) \,/\, q(\mathbf{s}) \right) \right\rangle_q, \tag{6}$$

where $< \cdot >_q$ denotes the mathematical expectation taken with respect to q. We will use the VEM algorithm that consists in iterating two steps until convergence: the E-step where we compute $q^\star = \arg\max_{q\in\mathcal{F}} \mathcal{L}(q;\boldsymbol{\theta}^\star)$ and the M-step where we compute $\boldsymbol{\theta}^\star = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(q^\star;\boldsymbol{\theta})$. In practice we will use the mean-field approximation by constraining the variational family $\mathcal{F}$ to the set of densities that factorize as $q(\mathbf{s}) = \prod_{j,f,n} q_{jfn}(s_{j,fn})$. Under this approximation we can show that the pdf over $s \in \mathbf{s}$ that maximizes the variational free energy satisfies [2]:

$$\ln q^\star(s) \stackrel{c}{=} \langle \ln p(\mathbf{x},\mathbf{s};\boldsymbol{\theta})\rangle_{q(\mathbf{s}\backslash s)}, \tag{7}$$

where $\stackrel{c}{=}$ represents equality up to an additive constant and $\mathbf{s}\backslash s$ denotes the set of all latent variables but s.

2.1. Source estimate

Under the variational mean-field approximation, the estimate of the j-th source in the TF domain is given by:

$$\hat{s}_{j,fn} = \langle s_{j,fn}\rangle_q. \tag{8}$$

The time-domain signal $\hat{s}_j(t)$ is then reconstructed by inverse TF transform and the source image $\hat{y}_{ij}(t)$ is obtained by convolution with the corresponding mixing filter: $\hat{y}_{ij}(t) = [a_{ij} \star \hat{s}_j](t)$.

2.2. Complete-data log-likelihood

According to the model defined in the previous section, the complete-data log-likelihood $\ln p(\mathbf{x},\mathbf{s};\boldsymbol{\theta}) = \ln p(\mathbf{x}|\mathbf{s};\boldsymbol{\theta}) + \ln p(\mathbf{s};\boldsymbol{\theta})$ can be expressed as:

$$\ln p(\mathbf{x},\mathbf{s};\boldsymbol{\theta}) \stackrel{c}{=} -\frac{1}{2}\sum_{i=1}^{I}\sum_{t=0}^{T-1}\left[\ln(\sigma_i^2) + \frac{1}{\sigma_i^2}\left(x_i(t) - \sum_{j=1}^{J} y_{ij}(t)\right)^2\right] - \frac{1}{\phi}\sum_{j=1}^{J}\sum_{(f,n)\in\mathcal{B}}\left[\ln(v_{j,fn}) + \frac{|s_{j,fn}|^2}{v_{j,fn}}\right]. \tag{9}$$

2.3. E-step

Under the mean-field approximation, we can show that the densities $q_{jfn}(s_{j,fn})$ which maximize the variational free energy satisfy:

$$\begin{aligned}
\ln q^\star_{jfn}(s_{j,fn}) &= \langle \ln p(\mathbf{x},\mathbf{s};\boldsymbol{\theta})\rangle_{q(\mathbf{s}\backslash s_{j,fn})} \\
&\stackrel{c}{=} -\frac{1}{2}\sum_{i=1}^{I}\sum_{t=0}^{T-1}\left[\ln(\sigma_i^2) + \left\langle \frac{1}{\sigma_i^2}\left(x_i(t) - \sum_{j'=1}^{J} y_{ij'}(t)\right)^2\right\rangle_{q(\mathbf{s}\backslash s_{j,fn})}\right] - \frac{1}{\phi}\frac{|s_{j,fn}|^2}{v_{j,fn}} \\
&\stackrel{c}{=} -\frac{1}{2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\left\langle\left[x_i(t) - \frac{2}{\phi}\Re\big(s_{j,fn}g_{ij,fn}(t)\big) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(s_{j',f'n'}g_{ij',f'n'}(t)\big)\right]^2\right\rangle_{q(\mathbf{s}\backslash s_{j,fn})} - \frac{1}{\phi}\frac{|s_{j,fn}|^2}{v_{j,fn}} \\
&\stackrel{c}{=} -\frac{1}{2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\left[x_i(t) - \frac{2}{\phi}\Re\big(s_{j,fn}g_{ij,fn}(t)\big) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\big)\right]^2 - \frac{1}{\phi}\frac{|s_{j,fn}|^2}{v_{j,fn}} \\
&\stackrel{c}{=} -\frac{1}{2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\left[\frac{4}{\phi^2}\Re\big(s_{j,fn}g_{ij,fn}(t)\big)^2 - \frac{4}{\phi}\Re\big(s_{j,fn}g_{ij,fn}(t)\big)\Big(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\big)\Big)\right] \\
&\quad - \frac{1}{\phi}\frac{|s_{j,fn}|^2}{v_{j,fn}} \\
&\stackrel{c}{=} \Re(s_{j,fn})^2\left[-\frac{2}{\phi^2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))^2 - \frac{1}{\phi v_{j,fn}}\right] + \Im(s_{j,fn})^2\left[-\frac{2}{\phi^2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Im(g_{ij,fn}(t))^2 - \frac{1}{\phi v_{j,fn}}\right] \\
&\quad - \Re(s_{j,fn})\left[-\frac{2}{\phi}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))\left(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\right)\right] \\
&\quad - \Im(s_{j,fn})\left[\frac{2}{\phi}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Im(g_{ij,fn}(t))\left(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\right)\right] \\
&\quad + \frac{4}{\phi^2}\Re(s_{j,fn})\Im(s_{j,fn})\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))\Im(g_{ij,fn}(t)),
\end{aligned} \tag{10}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts respectively.

In the complex case ($\mathbb{K} = \mathbb{C}$) we see that $q^{\star}_{jfn}(s_{j,fn})$ is the pdf of a complex Gaussian distribution which is not proper (see equation (29) in Appendix A for the pdf of this distribution). It means that the real and imaginary parts of the source coefficients are a posteriori correlated and have different variances. It the real case ($\mathbb{K} = \mathbb{R}$), as $\Im(s_{j,fn}) = 0$ we see that $q^{\star}_{jfn}(s_{j,fn})$ corresponds to a real Gaussian distribution. We now need to identify the parameters of the variational distribution:

- ▷ $\hat{s}^{r}_{j,fn} = \langle \Re(s_{j,fn}) \rangle_{q^\star}$;
- ▷ $\hat{s}^{\imath}_{j,fn} = \langle \Im(s_{j,fn}) \rangle_{q^\star}$;
- ▷ $\gamma^{r}_{j,fn} = \langle (\Re(s_{j,fn}) - \hat{s}^{r}_{j,fn})^2 \rangle_{q^\star}$;
- ▷ $\gamma^{\imath}_{j,fn} = \langle (\Im(s_{j,fn}) - \hat{s}^{\imath}_{j,fn})^2 \rangle_{q^\star}$;
- ▷ $\rho_{j,fn} = \dfrac{\langle (\Re(s_{j,fn}) - \hat{s}^{r}_{j,fn})(\Im(s_{j,fn}) - \hat{s}^{\imath}_{j,fn}) \rangle_{q^\star}}{\sqrt{\gamma^{r}_{j,fn}\gamma^{\imath}_{j,fn}}} \in [-1, 1]$.

Note that in the real case $\mathbb{K} = \mathbb{R}$, the optimal variational distribution is a real Gaussian which is only parametrized by $\hat{s}^{r}_{j,fn}$ and $\gamma^{r}_{j,fn}$. From equations (10) and (29), identifying the parameters of the variational distribution consists in solving the following system of equations:

▷ $$\frac{1}{2(1-\rho^2_{j,fn})\gamma^{r}_{j,fn}} = \frac{2}{\phi^2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))^2 + \frac{1}{\phi v_{j,fn}};$$

▷ $$\frac{1}{2(1-\rho^2_{j,fn})}\left(\frac{2}{\gamma^{r}_{j,fn}}\hat{s}^{r}_{j,fn} - \frac{2\rho_{j,fn}}{\sqrt{\gamma^{r}_{j,fn}\gamma^{\imath}_{j,fn}}}\hat{s}^{\imath}_{j,fn}\right) = \frac{2}{\phi}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))\left(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\big)\right);$$

▷ $$\frac{\rho_{j,fn}}{(1-\rho^2_{j,fn})}\frac{1}{\sqrt{\gamma^{r}_{j,fn}\gamma^{\imath}_{j,fn}}} = \frac{4}{\phi^2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))\Im(g_{ij,fn}(t));$$

▷ $$\frac{1}{2(1-\rho^2_{j,fn})\gamma^{\imath}_{j,fn}} = \frac{2}{\phi^2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Im(g_{ij,fn}(t))^2 + \frac{1}{\phi v_{j,fn}};$$

▷ $$-\frac{1}{2(1-\rho^2_{j,fn})}\left(\frac{2}{\gamma^{\imath}_{j,fn}}\hat{s}^{\imath}_{j,fn} - \frac{2\rho_{j,fn}}{\sqrt{\gamma^{r}_{j,fn}\gamma^{\imath}_{j,fn}}}\hat{s}^{r}_{j,fn}\right) = \frac{2}{\phi}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Im(g_{ij,fn}(t))\left(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\big)\right).$$

After calculation we obtain:

$$\rho_{j,fn} = \frac{\frac{2}{\phi^2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))\Im(g_{ij,fn}(t))}{\left(\frac{2}{\phi^2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))^2 + \frac{1}{\phi v_{j,fn}}\right)^{0.5}\left(\frac{2}{\phi^2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Im(g_{ij,fn}(t))^2 + \frac{1}{\phi v_{j,fn}}\right)^{0.5}}; \tag{11}$$

$$\gamma^{r}_{j,fn} = \left[2(1-\rho^2_{j,fn})\left(\frac{2}{\phi^2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))^2 + \frac{1}{\phi v_{j,fn}}\right)\right]^{-1}; \tag{12}$$

$$\gamma^{\imath}_{j,fn} = \left[2(1-\rho^2_{j,fn})\left(\frac{2}{\phi^2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Im(g_{ij,fn}(t))^2 + \frac{1}{\phi v_{j,fn}}\right)\right]^{-1}; \tag{13}$$

$$\begin{aligned}\hat{s}^{r}_{j,fn} =& \frac{2}{\phi}\gamma^{r}_{j,fn}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))\left(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\big)\right)\\ &- \frac{2}{\phi}\rho_{j,fn}\sqrt{\gamma^{r}_{j,fn}\gamma^{\imath}_{j,fn}}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Im(g_{ij,fn}(t))\left(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\big)\right);\end{aligned} \tag{14}$$

$$\hat{s}^{\imath}_{j,fn} = -\frac{2}{\phi}\gamma^{\imath}_{j,fn}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Im(g_{ij,fn}(t))\left(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\big)\right)$$
$$+\frac{2}{\phi}\rho_{j,fn}\sqrt{\gamma^r_{j,fn}\gamma^{\imath}_{j,fn}}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))\left(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\big)\right). \tag{15}$$

We will now simplify equations (14) and (15). Let us define:

$$d^r_{j,fn} = \frac{2}{\phi}\left[\frac{\hat{s}^r_{j,fn}}{v_{j,fn}} - \sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))\left(x_i(t) - \sum_{j'=1}^{J}\hat{y}_{ij'}(t)\right)\right]; \tag{16}$$

$$d^{\imath}_{j,fn} = \frac{2}{\phi}\left[\frac{\hat{s}^{\imath}_{j,fn}}{v_{j,fn}} + \sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Im(g_{ij,fn}(t))\left(x_i(t) - \sum_{j'=1}^{J}\hat{y}_{ij'}(t)\right)\right]. \tag{17}$$

We can show from equations (11)-(15) that the following three equalities hold:

$$\hat{s}^{\imath}_{j,fn}\rho_{j,fn}\sqrt{\frac{\gamma^r_{j,fn}}{\gamma^{\imath}_{j,fn}}} = -\frac{2}{\phi}\rho_{j,fn}\sqrt{\gamma^r_{j,fn}\gamma^{\imath}_{j,fn}}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Im(g_{ij,fn}(t))\left(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\big)\right)$$
$$+\frac{2}{\phi}\rho^2_{j,fn}\gamma^r_{j,fn}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))\left(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\big)\right); \tag{18}$$

$$\hat{s}^r_{j,fn}\rho_{j,fn}\sqrt{\frac{\gamma^{\imath}_{j,fn}}{\gamma^r_{j,fn}}} = \frac{2}{\phi}\rho_{j,fn}\sqrt{\gamma^r_{j,fn}\gamma^{\imath}_{j,fn}}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))\left(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\big)\right)$$
$$-\frac{2}{\phi}\rho^2_{j,fn}\gamma^{\imath}_{j,fn}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Im(g_{ij,fn}(t))\left(x_i(t) - \frac{2}{\phi}\sum_{j'f'n'\neq jfn}\Re\big(\hat{s}_{j',f'n'}g_{ij',f'n'}(t)\big)\right); \tag{19}$$

$$\frac{2}{\phi^2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}\Re(g_{ij,fn}(t))\Im(g_{ij,fn}(t)) = \frac{\rho_{j,fn}}{2\left(1-\rho^2_{j,fn}\right)\sqrt{\gamma^r_{j,fn}\gamma^{\imath}_{j,fn}}}. \tag{20}$$

Using (18) and (20) for rewriting (14) and recognizing $d^r_{j,fn}$ defined in (16) we obtain:

$$\hat{s}^r_{j,fn} = \hat{s}^r_{j,fn} - \gamma^r_{j,fn}(1-\rho^2_{j,fn})d^r_{j,fn}. \tag{21}$$

Similarly, using (19) and (20) for rewriting (15) and recognizing $d^{\imath}_{j,fn}$ defined in (17) we obtain:

$$\hat{s}^{\imath}_{j,fn} = \hat{s}^{\imath}_{j,fn} - \gamma^{\imath}_{j,fn}(1-\rho^2_{j,fn})d^{\imath}_{j,fn}. \tag{22}$$

We have to mention that updates (21) and (22) hold if the parameters are updated in turn. Finally, the first and second-order moments of this variational distribution are given as follows:

- ▷ Mean: $\hat{s}_{j,fn} = \langle s_{j,fn}\rangle_q = \hat{s}^r_{j,fn} + \imath\hat{s}^{\imath}_{j,fn}$;
- ▷ Variance: $\gamma_{j,fn} = \langle|s_{j,fn} - \hat{s}_{j,fn}|^2\rangle_q = \gamma^r_{j,fn} + \gamma^{\imath}_{j,fn}$;
- ▷ Pseudo-variance: $\tilde{\gamma}_{j,fn} = \langle(s_{j,fn} - \hat{s}_{j,fn})^2\rangle_q = \gamma^r_{j,fn} - \gamma^{\imath}_{j,fn} + 2\imath\rho_{j,fn}\sqrt{\gamma^r_{j,fn}\gamma^{\imath}_{j,fn}}$.

2.4. Variational free energy

From (6), (9) and the E-step, the variational free energy can be written as follows:

$$\mathcal{L}(q^\star;\boldsymbol{\theta}) \stackrel{c}{=} -\frac{1}{2}\sum_{i=1}^{I}\sum_{t=0}^{T-1}\left[\ln(\sigma_i^2) + \frac{e_i(t)}{\sigma_i^2}\right] - \frac{1}{\phi}\sum_{j=1}^{J}\sum_{(f,n)\in\mathcal{B}}\left[\ln\left(v_{j,fn}\right) + \frac{|\hat{s}_{j,fn}|^2 + \gamma_{j,fn}}{v_{j,fn}}\right]$$
$$+\frac{1}{2}\sum_{j=1}^{J}\sum_{(f,n)\in\mathcal{B}}\begin{cases}\ln\left(\gamma^r_{j,fn}\right) & \text{if } \mathbb{K}=\mathbb{R};\\ \ln\left(\gamma^r_{j,fn}\gamma^{\imath}_{j,fn}(1-\rho^2_{j,fn})\right) & \text{if } \mathbb{K}=\mathbb{C},\end{cases} \tag{23}$$

where $e_i(t) = \left\langle\left(x_i(t) - \sum_{j=1}^{J}y_{ij}(t)\right)^2\right\rangle_{q^\star}$ can be expressed from the mean-field approximation and (3) as:

$$e_i(t) = \left(x_i(t) - \sum_{j=1}^{J}\hat{y}_{ij}(t)\right)^2 + \frac{2}{\phi^2}\sum_{j=1}^{J}\sum_{(f,n)\in\mathcal{B}}\left[\Re\left(\tilde{\gamma}_{j,fn}g^2_{ij,fn}(t)\right) + \gamma_{j,fn}|g_{ij,fn}(t)|^2\right]. \tag{24}$$

2.5. Preconditioned conjugate gradient method

We can easily show from the expression of the variational free energy in (23) that $d_{j,fn}^{(\cdot)}$ as defined in (16) or (17) further satisfies the following equality: $d_{j,fn}^{(\cdot)} = \partial(-\mathcal{L}(q^\star;\boldsymbol{\theta}))/(\partial \hat{s}_{j,fn}^{(\cdot)})$. Therefore we clearly see from (21) and (22) that the update of $\hat{s}_{j,fn}^{(\cdot)}$ corresponds to going in the opposite direction of the derivative $d_{j,fn}^{(\cdot)}$ with a step size $\gamma_{j,fn}^{(\cdot)}(1-\rho_{j,fn}^2)$. When the derivative is zero, it is clear that we achieve a fixed point of the algorithm. Therefore, for the sake of computational efficiency, we will use the Preconditioned Conjugate Gradient (PCG) method [3] instead of the coordinate-wise updates (21) and (22).

For the sake of conciseness, we will work with complex-valued vectors. More precisely we rely on Wirtinger calculus for computing the generalized complex derivatives defined as follows [4]:

$$\frac{1}{2} d_{j,fn} = \frac{\partial - \mathcal{L}(q^\star;\boldsymbol{\theta})}{\partial \hat{s}_{j,fn}^*} = \frac{1}{2}\left(d_{j,fn}^r + \imath d_{j,fn}^\imath\right). \tag{25}$$

From equations (16) and (17) we have:

$$d_{j,fn} = \frac{2}{\phi}\left[\frac{\hat{s}_{j,fn}}{v_{j,fn}} - \sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1} g_{ij,fn}^*(t)\Big(x_i(t) - \sum_{j'=1}^{J}\hat{y}_{ij'}(t)\Big)\right]. \tag{26}$$

Let us introduce the following notations:

- ▷ $\hat{\mathbf{s}} = [\hat{\mathbf{s}}^r + \imath\hat{\mathbf{s}}^\imath] \in \mathbb{C}^{JFN}$ with $\hat{\mathbf{s}}^{(\cdot)} \in \mathbb{R}^{JFN}$ the column vector of entries $\hat{s}_{j,fn}^{(\cdot)}$;
- ▷ $\mathbf{d} = [\mathbf{d}^r + \imath\mathbf{d}^\imath] \in \mathbb{C}^{JFN}$ with $\mathbf{d}^{(\cdot)} \in \mathbb{R}^{JFN}$ the column vector of entries $d_{j,fn}^{(\cdot)}$;
- ▷ $\mathbf{g}_i(t) \in \mathbb{C}^{JFN}$ the column vector of entries $g_{ij,fn}(t)$;
- ▷ $\mathbf{D}$ the diagonal preconditioning matrix of size $JFN \times JFN$ with entries:

$$\frac{(\gamma_{j,fn}^r)^{-1} + (\gamma_{j,fn}^\imath)^{-1}}{1-\rho_{j,fn}^2} = \left(\frac{4}{\phi^2}\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1}|g_{ij,fn}(t)|^2 + \frac{2}{\phi v_{j,fn}}\right).$$

The order of the coefficients indexed by j, f, n for constructing these vectors and this diagonal matrix does not matter as long as it is kept identical. The PCG method is summarized in Algorithm 1.

Algorithm 1: PCG method for the E-step (update of $\hat{s}_{j,fn}$)

Initialize $\mathbf{d}$ from equation (26) and $\boldsymbol{\omega} = \mathbf{D}^{-1}\mathbf{d}$
while stopping criterion not reached **do**
  Compute $\boldsymbol{\kappa}$ the column vector of size JFN and entries
  $$\kappa_{j,fn} = \frac{2}{\phi}\left[\frac{\omega_{j,fn}}{v_{j,fn}} + \sum_{i=1}^{I}\frac{1}{\sigma_i^2}\sum_{t=0}^{T-1} g_{ij,fn}^*(t)\sum_{j'=1}^{J}\frac{2}{\phi}\Re\left(\sum_{f'=0}^{F}\sum_{n'=0}^{N}\omega_{j',f',n'}g_{ij',f'n'}(t)\right)\right]$$
  $\mu = (\boldsymbol{\omega}^H\mathbf{d})/(\boldsymbol{\omega}^H\boldsymbol{\kappa})$
  $\hat{\mathbf{s}} \leftarrow \hat{\mathbf{s}} - \mu\boldsymbol{\omega}$
  Compute $\mathbf{d}$ from equation (26)
  $\mathbf{d}_p = \mathbf{D}^{-1}\mathbf{d}$
  $\beta = -(\boldsymbol{\kappa}^H\mathbf{d}_p)/(\boldsymbol{\omega}^H\boldsymbol{\kappa})$
  $\boldsymbol{\omega} \leftarrow \mathbf{d}_p + \beta\boldsymbol{\omega}$
end while

2.6. M-step

The M-step consists in maximizing (or only increasing) $\mathcal{L}(q^\star;\boldsymbol{\theta})$ in (23) with respect to $\boldsymbol{\theta}$. Zeroing the derivative of this criterion with respect to σ_i^2 leads to the following update:

$$\sigma_i^2 = \frac{1}{T}\sum_{t=0}^{T-1} e_i(t). \tag{27}$$

Up to an additive term which does not depend on the NMF parameters, we can recognize in equation (23) the Itakura-Saito (IS) divergence [5] between the posterior mean of the source power spectrogram $\langle |s_{j,fn}|^2\rangle_{q^\star} = |\hat{s}_{j,fn}|^2 + \gamma_{j,fn}$ and $v_{j,fn} = [\mathbf{W}_j\mathbf{H}_j]_{fn}$. Therefore the source parameters are updated by computing an NMF on the matrix $\hat{\mathbf{P}}_j \in \mathbb{R}_+^{F\times N}$ with entries $[\hat{\mathbf{P}}_j]_{fn} = |\hat{s}_{j,fn}|^2 + \gamma_{j,fn}$ using the IS divergence. It can be done with the standard multiplicative update rules given in [5].

2.7. Summary of the VEM algorithm

In the complex case ($\mathbb{K} = \mathbb{C}$), the E-step corresponds to first updating $\rho_{j,fn}$, $\gamma^{r}_{j,fn}$ and $\gamma^{\imath}_{j,fn}$ according to equations (11), (12) and (13) respectively and then updating $\hat{s}_{j,fn} = \hat{s}^{r}_{j,fn} + \imath \hat{s}^{\imath}_{j,fn}$ with the PCG method summarized in Algorithm 1. In the real case ($\mathbb{K} = \mathbb{R}$), one only needs to compute $\gamma^{r}_{j,fn}$ and $\hat{s}^{r}_{j,fn}$ with the same updates. The E-step is summarized in Algorithm 2 for $\mathbb{K} = \mathbb{C}$ and Algorithm 3 for $\mathbb{K} = \mathbb{R}$. We clearly see that using the MDCT which is a real-valued TF transform allows us to reduce the computational load compared with the use of the OFSTFT.

Algorithm 2: E-step for $\mathbb{K} = \mathbb{C}$, i.e. $\phi = 1$

for all j, f, n **do**

$$\rho_{j,fn} = \frac{2 \sum_{i=1}^{I} \frac{1}{\sigma_i^2} \sum_{t=0}^{T-1} \Re(g_{ij,fn}(t)) \Im(g_{ij,fn}(t))}{\left(2 \sum_{i=1}^{I} \frac{1}{\sigma_i^2} \sum_{t=0}^{T-1} \Re(g_{ij,fn}(t))^2 + \frac{1}{v_{j,fn}} \right)^{0.5} \left(2 \sum_{i=1}^{I} \frac{1}{\sigma_i^2} \sum_{t=0}^{T-1} \Im(g_{ij,fn}(t))^2 + \frac{1}{v_{j,fn}} \right)^{0.5}}$$

$$\gamma^{r}_{j,fn} = \left[2(1 - \rho^2_{j,fn}) \Big(2 \sum_{i=1}^{I} \frac{1}{\sigma_i^2} \sum_{t=0}^{T-1} \Re(g_{ij,fn}(t))^2 + \frac{1}{v_{j,fn}} \Big) \right]^{-1}$$

$$\gamma^{\imath}_{j,fn} = \left[2(1 - \rho^2_{j,fn}) \Big(2 \sum_{i=1}^{I} \frac{1}{\sigma_i^2} \sum_{t=0}^{T-1} \Im(g_{ij,fn}(t))^2 + \frac{1}{v_{j,fn}} \Big) \right]^{-1}$$

end for

Update $\{\hat{s}_{j,fn} = \hat{s}^{r}_{j,fn} + \imath \hat{s}^{\imath}_{j,fn}\}_{j,f,n}$ with the PCG method (Algorithm 1)

Algorithm 3: E-step for $\mathbb{K} = \mathbb{R}$, i.e. $\phi = 2$

for all j, f, n **do**

$$\gamma^{r}_{j,fn} = \left[\sum_{i=1}^{I} \frac{1}{\sigma_i^2} \sum_{t=0}^{T-1} g^2_{ij,fn}(t) + \frac{1}{v_{j,fn}} \right]^{-1}$$

end for

Update $\{\hat{s}_{j,fn} = \hat{s}^{r}_{j,fn}\}_{j,f,n}$ with the PCG method (Algorithm 1)

Algorithm 4: M-step

$\sigma_i^2 = \frac{1}{T} \sum_{t=0}^{T-1} e_i(t)$, with $e_i(t)$ defined in equation (24)

$\mathbf{W}_j, \mathbf{H}_j =$ IS-NMF($\hat{\mathbf{P}}_j$) where $[\hat{\mathbf{P}}_j]_{fn} = |\hat{s}_{j,fn}|^2 + \gamma_{j,fn}$

3. REFERENCES

[1] S. Leglaive, R. Badeau, and G. Richard, "Separating time-frequency sources from time-domain convolutive mixtures using non-negative matrix factorization," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, 2017.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[3] G. H. Golub and C. F. Van Loan, *Matrix computations*. Johns Hopkins University Press, 1996.

[4] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5101–5125, 2011.

[5] 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.

A. GAUSSIAN PROBABILITY DISTRIBUTIONS

A.1. Real Gaussian distribution

Let $\mathcal{N}_{\mathbb{R}}(x;\mu,\sigma^2)$ denote the Gaussian distribution over a real-valued random variable (r.v.) x. Its pdf is given by:

$$N_{\mathbb{R}}(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{28}$$

A.2. Complex Gaussian distribution

Let $\mathcal{N}_{\mathbb{C}}(x;\rho,\mu_{x_r},\mu_{x_i},\sigma_{x_r}^2,\sigma_{x_i}^2)$ denote the Gaussian distribution over a complex-valued r.v. $x = x_r + \imath x_\imath$. Its pdf is given by [4]:

$$N_{\mathbb{C}}(x;\rho,\mu_{x_r},\mu_{x_\imath},\sigma_{x_r}^2,\sigma_{x_\imath}^2) = \frac{1}{2\pi\sigma_{x_r}\sigma_{x_\imath}\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x_r-\mu_{x_r})^2}{\sigma_{x_r}^2} + \frac{(x_\imath-\mu_{x_\imath})^2}{\sigma_{x_\imath}^2} - \frac{2\rho(x_r-\mu_{x_r})(x_\imath-\mu_{x_\imath})}{\sigma_{x_r}\sigma_{x_\imath}}\right)\right], \tag{29}$$

where $\rho = \mathbb{E}[(x_r-\mu_{x_r})(x_\imath-\mu_{x_\imath})]/(\sigma_{x_r}\sigma_{x_\imath}) \in [-1,1]$.

The particular case $\mathcal{N}_{\mathbb{C}}(x;\rho=0,\mu_{x_r},\mu_{x_i},\sigma_{x_r}^2=\sigma^2/2,\sigma_{x_i}^2=\sigma^2/2)$ corresponds to the proper complex Gaussian distribution. It is further written as $\mathcal{N}_{\mathbb{C}}^p(x;\mu,\sigma^2)$ where $\mu = \mu_{x_r} + i\mu_{x_i}$ and $\sigma^2 = 2\sigma_{x_r}^2 = 2\sigma_{x_i}^2$. Its pdf is given by:

$$N_{\mathbb{C}}^p(x;\mu,\sigma^2) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{|x-\mu|^2}{\sigma^2}\right). \tag{30}$$

The complex Gaussian distribution is circularly symmetric if it is proper and $\mu = 0$.