Supporting Technical Report for the Article ”Variational Bayesian Inference for Source Separation and Robust Feature Extraction”

Kamil Adiloglu, Emmanuel Vincent

Supporting Technical Report for the Article "Variational Bayesian Inference for Source Separation and Robust Feature Extraction"

Kamil Adiloğlu, Emmanuel Vincent

TECHNICAL REPORT
N° 423
February 2014
Project-Team Metiss

ISSN 0249-0803   ISRN INRIA/RT-423--FR+ENG

Supporting Technical Report for the Article "Variational Bayesian Inference for Source Separation and Robust Feature Extraction"

Kamil Adiloğlu, Emmanuel Vincent

Project-Team Metiss

Technical Report n° 423 — February 2014 — 23 pages

Abstract: This technical report presents the details of the derivation of the variational Bayesian source separation algorithm in [1]. For the motivation behind this algorithm and for experimental results, see [1].

Key-words: Audio source separation, local Gaussian modeling, non-negative matrix factorization, variational Bayesian inference

This work is a part of the QUAERO project funded by OSEO.

RESEARCH CENTRE
RENNES – BRETAGNE ATLANTIQUE

Campus universitaire de Beaulieu
35042 Rennes Cedex

Rapport Technique pour l'Article "Variational Bayesian Inference for Source Separation and Robust Feature Extraction"

Résumé : Ce rapport présente les détails de conception de l'algorithme variationel bayesien pour la séparation de sources dans [1]. La motivation sous-jacente á cet algorithme et les résultats sont donnés dans [1].

Mots-clés : Séparation de sources audio, modèle gaussien local, factorisation matricielle positive, inférence variationelle bayésienne

Contents

1 Model

1.1 Spatial Model

We adopt the model proposed in [5]. This model operates in the STFT domain. The mixing equation in time frequency bin (n, f) can be formulated by using the source images $\mathbf{y}_{j,fn}$ as follows:

$$\mathbf{x}_{fn} = \sum_{j=1}^{J} \mathbf{y}_{j,fn} + \boldsymbol{\epsilon}_{fn}. \tag{1}$$

In this domain, $\mathbf{x}_{fn}$ represents the $I \times 1$ vector containing the mixture STFT coefficients in time-frequency bin (n, f), , where I is the number of channels. $\mathbf{y}_{j,fn}$ represents the $I \times 1$ vector consisting of the spatial image of source j on all mixture channels in the same time-frequency bin. Finally $\boldsymbol{\epsilon}_{fn}$ is the noise in the STFT domain. We assume that each source image follows a zero-mean complex-valued Gaussian distribution

$$\mathbf{y}_{j,fn} \sim \mathcal{N}(\mathbf{0}, v_{j,fn}\mathbf{R}_{j,f}), \tag{2}$$

where $\mathbf{R}_{j,f}$ is an $I \times I$ matrix of rank R_j which is called the spatial covariance matrix and represents the spatial characteristcs of source j and of the mixing system, and $v_{j,fn}$ is a scalar spectral power which represents the spectral characteristics of source j.

This model can alternatively be represented as follows [5]. $\mathbf{R}_{j,f}$ can be written as $\mathbf{R}_{j,f} = \mathbf{A}_{j,f}\mathbf{A}_{j,f}^H$, where $\mathbf{A}_{j,f}$ is a $I \times R_j$ dimensional complex-valued mixing matrix of rank R_j for source j. For each source j, we define R_j source components $s_{jr,fn}$ distributed as

$$s_{jr,fn} \sim \mathcal{N}(0, v_{j,fn}). \tag{3}$$

Denoting by $R = \sum_{j=1}^{J} R_j$ the total number of source components, the source coefficients can be formulated as

$$\mathbf{s}_{fn} = [\mathbf{s}_{1,fn}^T, \ldots, \mathbf{s}_{j,fn}^T, \ldots, \mathbf{s}_{J,fn}^T]^T, \tag{4}$$

where $\mathbf{s}_{fn}$ is an $R \times 1$ vector of source coefficients with

$$\mathbf{s}_{j,fn} = [s_{j1,fn}, \ldots, s_{jr,fn}, \ldots, s_{jR_j,fn}]^T. \tag{5}$$

Hence, the prior distribution of the sources in time-frequency bin (f, n) is given by

$$\mathbf{s}_{fn} = \mathcal{N}(0, \boldsymbol{\Sigma}_{s,fn}), \tag{6}$$

where $\boldsymbol{\Sigma}_{s,fn}$ is an $R \times R$ diagonal covariance matrix consisting of $v_{j,fn}$ repeated R_j times for each source component $s_{jr,fn}$ of source j. Finally, the mixing equation using the source components instead of the source spatial images as in (1) is given by

$$\mathbf{x}_{fn} = \mathbf{A}_f \mathbf{s}_{fn} + \boldsymbol{\epsilon}_{fn}. \tag{7}$$

We assume that the mixing system $\mathbf{A}$ is stationary. So, the mixing system is expressed as

$$\mathbf{A}_f = [\mathbf{A}_{1,f}, \ldots, \mathbf{A}_{j,f}, \ldots, \mathbf{A}_{J,f}]. \tag{8}$$

Finally, we assume a Gaussian zero mean noise with a constant noise variance $\boldsymbol{\epsilon}_{fn} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_b)$, where $\boldsymbol{\Sigma}_b = \sigma_b^2 \mathbf{I}$. This gives us the possibility to formulate the likelihood of the mixture coefficients as follows:

$$p(\mathbf{X}|\mathbf{S}, \mathbf{A}) = \prod_{n=1}^{N} \prod_{f=1}^{F} \mathcal{N}(\mathbf{x}_{fn}|\mathbf{A}_f \mathbf{s}_{fn}, \sigma_b^2 \mathbf{I}), \tag{9}$$

where $\mathbf{X} = \{\mathbf{x}_{fn}\}_{n=1\cdots N, f=1\cdots F}$ and $\mathbf{S} = \{\mathbf{s}_{fn}\}_{n=1\cdots N, f=1\cdots F}$.

1.2 Nonnegative Tensor Factorization as a Spectral Model

We assume a three-level decomposition of the source variances $v_{j,fn}$ in a nonnegative tensor factorization (NTF) fashion [5]. In the first level we assume that the variances are the product of an excitation and a filter

$$v_{j,fn} = v_{j,fn}^{\mathrm{ex}} v_{j,fn}^{\mathrm{ft}}. \tag{10}$$

At the second level, the excitation spectral power $v_{j,fn}^{\mathrm{ex}}$ is expressed as the sum of basis spectra scaled by time activation coefficients. Finally, at the third level, the basis spectra are defined as the sum of narrowband spectral patterns $w_{j,fl}^{\mathrm{ex}}$ weighted by spectral envelope coefficients $u_{j,lk}^{\mathrm{ex}}$. Similarly, the time activation coefficients are represented as the sum of time-localized patterns $h_{j,mn}^{\mathrm{ex}}$ weighted by temporal envelope coefficients $g_{j,km}^{\mathrm{ex}}$. The same decomposition applies to the filter spectral power $v_{j,fn}^{\mathrm{ft}}$. Overall, the complete factorization scheme is as follows:

$$v_{j,fn}^{\mathrm{ex}} = \sum_{k=1}^{K_j^{\mathrm{ex}}} \sum_{m=1}^{M_j^{\mathrm{ex}}} \sum_{l=1}^{L_j^{\mathrm{ex}}} h_{j,mn}^{\mathrm{ex}} g_{j,km}^{\mathrm{ex}} u_{j,lk}^{\mathrm{ex}} w_{j,fl}^{\mathrm{ex}}, \tag{11}$$

$$v_{j,fn}^{\mathrm{ft}} = \sum_{k'=1}^{K_j^{\mathrm{ft}}} \sum_{m'=1}^{M_j^{\mathrm{ft}}} \sum_{l'=1}^{L_j^{\mathrm{ft}}} h_{j,m'n}^{\mathrm{ft}} g_{j,k'm'}^{\mathrm{ft}} u_{j,l'k'}^{\mathrm{ft}} w_{j,fl'}^{\mathrm{ft}}. \tag{12}$$

This framework makes it possible to exploit a wide range of prior information about the sources. For instance, harmonicity can be enforced by fixing $w_{j,fl}^{\mathrm{ex}}$ as narrowband harmonic spectra and letting the spectral envelope and the active pitches be inferred from the data via $u_{j,lk}^{\mathrm{ex}}$ and $g_{j,km}^{\mathrm{ex}}$, respectively [5]. For more details and examples of possible spectral and temporal constraints, see [5].

This three-level NTF decomposition can be shown in matrix form as follows

$$\mathbf{V}_j = (\mathbf{W}_j^{\mathrm{ex}} \mathbf{U}_j^{\mathrm{ex}} \mathbf{G}_j^{\mathrm{ex}} \mathbf{H}_j^{\mathrm{ex}}) \odot (\mathbf{W}_j^{\mathrm{ft}} \mathbf{U}_j^{\mathrm{ft}} \mathbf{G}_j^{\mathrm{ft}} \mathbf{H}_j^{\mathrm{ft}}), \tag{13}$$

where $\odot$ means element-wise matrix multiplication.

1.3 Prior Distributions for the Parameters

Each parameter may be fixed or adapted to the data. In a fully Bayesian treatment, we need to define the prior distributions for those parameters which are adapted to the data. We assume the NTF parameters of the source variances follow the non-informative Jeffreys prior $\mathcal{J}(x) \propto \frac{1}{x}$

$$\begin{aligned} w^{\mathrm{ex}}_{j,fl} &\sim \mathcal{J}, & (14)\\ u^{\mathrm{ex}}_{j,lk} &\sim \mathcal{J}, & (15)\\ g^{\mathrm{ex}}_{j,km} &\sim \mathcal{J}, & (16)\\ h^{\mathrm{ex}}_{j,mn} &\sim \mathcal{J}, & (17)\\ w^{\mathrm{ft}}_{j,fl'} &\sim \mathcal{J}, & (18)\\ u^{\mathrm{ft}}_{j,l'k'} &\sim \mathcal{J}, & (19)\\ g^{\mathrm{ft}}_{j,k'm'} &\sim \mathcal{J}, & (20)\\ h^{\mathrm{ft}}_{j,m'n} &\sim \mathcal{J}. & (21) \end{aligned}$$

For the mixing system, we take the dependencies between the channels and between the source components into account. Therefore, we consider the mixing matrix $\mathbf{A}_f$ as a whole and define the prior distribution accordingly. First, we reshape the mixing matrix $\mathbf{A}_f$ into a vector $\underline{\mathbf{A}}_f$. For this, we concatenate the row vectors of $\mathbf{A}_f$ into the column vector $\underline{\mathbf{A}}_f$. Then, we define the prior distribution of $\underline{\mathbf{A}}_f$ to be a complex multivariate Gaussian distribution as follows

$$\underline{\mathbf{A}}_f \sim \mathcal{N}(\mu_{\underline{A},f}, \mathbf{\Sigma}_{\underline{A},f}). \tag{22}$$

In the following, we assume that $\mathbf{\Sigma}_{\underline{A},f} \to +\infty$ so that this prior is actually flat.

1.4 Joint Distribution

With this prior information assumed, we can formulate the joint distribution $p(\mathbf{X}, \mathbf{Z})$, where $\mathbf{Z}$ is the set of all model parameters

$$\mathbf{Z} = \{\mathbf{S}, \mathbf{A}, \mathbf{W}^{\mathrm{ex}}, \mathbf{U}^{\mathrm{ex}}, \mathbf{G}^{\mathrm{ex}}, \mathbf{H}^{\mathrm{ex}}, \mathbf{W}^{\mathrm{ft}}, \mathbf{U}^{\mathrm{ft}}, \mathbf{G}^{\mathrm{ft}}, \mathbf{H}^{\mathrm{ft}}\}. \tag{23}$$

as

$$\begin{aligned} p(\mathbf{X}, \mathbf{Z}) \;=\;& p(\mathbf{X}|\mathbf{S}, \mathbf{A}) p(\mathbf{S}|\mathbf{W}^{\mathrm{ex}}, \mathbf{U}^{\mathrm{ex}}, \mathbf{G}^{\mathrm{ex}}, \mathbf{H}^{\mathrm{ex}}, \mathbf{W}^{\mathrm{ft}}, \mathbf{U}^{\mathrm{ft}}, \mathbf{G}^{\mathrm{ft}}, \mathbf{H}^{\mathrm{ft}}) p(\mathbf{A})\\ & p(\mathbf{W}^{\mathrm{ex}}) p(\mathbf{U}^{\mathrm{ex}}) p(\mathbf{G}^{\mathrm{ex}}) p(\mathbf{H}^{\mathrm{ex}}) p(\mathbf{W}^{\mathrm{ft}}) p(\mathbf{U}^{\mathrm{ft}}) p(\mathbf{G}^{\mathrm{ft}}) p(\mathbf{H}^{\mathrm{ft}}). \qquad (24) \end{aligned}$$

The log-distribution is then given by

$$
\begin{aligned}
\log p(\mathbf{X},\mathbf{Z}) &= \log p(\mathbf{X}|\mathbf{S},\mathbf{A}) + \log p(\mathbf{S}|\mathbf{W}^{\text{ex}},\mathbf{U}^{\text{ex}},\mathbf{G}^{\text{ex}},\mathbf{H}^{\text{ex}},\mathbf{W}^{\text{ft}},\mathbf{U}^{\text{ft}},\mathbf{G}^{\text{ft}},\mathbf{H}^{\text{ft}}) \\
&\quad + \log p(\mathbf{A}) + \log p(\mathbf{W}^{\text{ex}}) + \log p(\mathbf{U}^{\text{ex}}) + \log p(\mathbf{G}^{\text{ex}}) + \log p(\mathbf{H}^{\text{ex}}) \\
&\quad + \log p(\mathbf{W}^{\text{ft}}) + \log p(\mathbf{U}^{\text{ft}}) + \log p(\mathbf{G}^{\text{ft}}) + \log p(\mathbf{H}^{\text{ft}}), \\
&= \sum_{f,n} \log \mathcal{N}(\mathbf{x}_{fn}|\mathbf{A}_f\mathbf{s}_{fn},\sigma_b^2\mathbf{I}) \\
&\quad + \sum_{f,n} \log \mathcal{N}(\mathbf{s}_{fn}|0,\mathbf{\Sigma}_{\mathbf{s},fn}) \\
&\quad + \sum_{j,f,l} \log \mathcal{J}(w^{\text{ex}}_{j,fl}) \\
&\quad + \sum_{j,l,k} \log \mathcal{J}(u^{\text{ex}}_{j,lk}) \\
&\quad + \sum_{j,k,m} \log \mathcal{J}(g^{\text{ex}}_{j,km}) \\
&\quad + \sum_{j,m,n} \log \mathcal{J}(h^{\text{ex}}_{j,mn}) \\
&\quad + \sum_{j,f,l'} \log \mathcal{J}(w^{\text{ft}}_{j,fl'}) \\
&\quad + \sum_{j,l',k'} \log \mathcal{J}(u^{\text{ft}}_{j,l'k'}) \\
&\quad + \sum_{j,k',m'} \log \mathcal{J}(g^{\text{ft}}_{j,k'm'}) \\
&\quad + \sum_{j,m',n} \log \mathcal{J}(h^{\text{ft}}_{j,m'n}) \\
&\quad + \sum_{f} \log \mathcal{N}(\underline{\mathbf{A}}_f|\mu_{\underline{\mathbf{A}},f},\mathbf{\Sigma}_{\underline{\mathbf{A}},f}).
\end{aligned} \tag{25}
$$

2 Variational Inference

We aim to obtain the posterior distribution of the model parameters $p(\mathbf{Z}|\mathbf{X})$. However exact Bayesian inference is intractable. Therefore we resort to a variational Bayesian approximation [2].

2.1 General Approach

Variational Bayesian inference aims to obtain an approximation $q(\mathbf{Z})$ of the true posterior distribution $p(\mathbf{Z}|\mathbf{X})$ that minimizes the Kullback-Leibler (KL) divergence

$$KL(q||p) = -\int q(\mathbf{Z}) \log \frac{p(\mathbf{Z}|\mathbf{X})}{q(\mathbf{Z})} d\mathbf{Z}. \tag{26}$$

Marginalizing out the model parameters $\mathbf{Z}$ from (24) gives the marginal likelihood $p(\mathbf{X})$ or evidence. We can show that the following equation holds

$$\log p(\mathbf{X}) = \mathscr{L}(q) + KL(q||p), \tag{27}$$

where the free energy $\mathscr{L}(q)$ is given by

$$\mathscr{L}(q) \quad = \quad \int q(\mathbf{Z}) \log \frac{p(\mathbf{X}, \mathbf{Z})}{q(\mathbf{Z})} d\mathbf{Z}, \tag{28}$$

$$\tag{29}$$

In this formulation, $q(\mathbf{Z})$ is the joint distribution of the model parameters which we use to approximate the true posterior distribution $p(\mathbf{Z}|\mathbf{X})$. $\mathscr{L}(q)$ is called the free energy and it is a lower bound of the marginal likelihood. Maximizing this free energy wrt. $q(\mathbf{Z})$ minimizes the KL-divergence between the approximating distribution $q(\mathbf{Z})$ and the true posterior [2]. Note that the KL-divergence vanishes when $q(\mathbf{Z})$ is equal to the true posterior.

From now on, we assume the following factorization of the variational distribution $q(\mathbf{Z})$:

$$q(\mathbf{Z}) = \prod_{i=1}^{M} q_i(\mathbf{Z}_i). \tag{30}$$

If we embed this factorization into the free energy given in (28) and dissect out the dependency on one of the factors $q_i(\mathbf{Z}_i)$, we obtain

$$\mathscr{L}(q) = \int q_i \Big\{ \int \log p(\mathbf{X}, \mathbf{Z}) \prod_{i' \neq i} q_{i'} d\mathbf{Z}_{i'} \Big\} d\mathbf{Z}_i - \int q_i \log q_i d\mathbf{Z}_i + \text{const}, \tag{31}$$

$$= \int q_i \log \tilde{p}(\mathbf{X}, \mathbf{Z}_i) d\mathbf{Z}_i - \int q_i \log q_i d\mathbf{Z}_i + \text{const}, \tag{32}$$

$$= \int q_i \log \frac{\tilde{p}(\mathbf{X}, \mathbf{Z}_i)}{q_i} d\mathbf{Z}_i + \text{const}, \tag{33}$$

where

$$\log \tilde{p}(\mathbf{X}, \mathbf{Z}_i) = \int \log p(\mathbf{X}, \mathbf{Z}) \prod_{i' \neq i} q_{i'} d\mathbf{Z}_{i'} + \text{const}, \tag{34}$$

$$= \mathbb{E}_{i' \neq i}[\log p(\mathbf{X}, \mathbf{Z})] + \text{const} \tag{35}$$

and the normalizing constant is such that $\tilde{p}(\mathbf{X}, \mathbf{Z}_i)$ is a proper distribution.

Now, suppose that we keep $q_{i' \neq i}$ constant and maximize the free energy shown in (28) wrt. q_i. The minimum occurs when $q_i(\mathbf{Z}_i) = \tilde{p}(\mathbf{X}, \mathbf{Z}_i)$. As a result, we obtain a general equation for the solution by maximizing the free energy as follows

$$q_i^*(\mathbf{Z}_i) = \tilde{p}(\mathbf{X}, \mathbf{Z}_i). \tag{36}$$

Note that in this equation, the update equation of the optimal approximating distribution $q_i^*(\mathbf{Z}_i)$ depends on the expectation of the log of the joint distribution wrt. all other variational distributions. In that sense, (36) indicates a set of equations for $i = \{1, \ldots, M\}$. Therefore an iterative update procedure is needed. After proper initialization of all the variational distributions, each distribution is updated in an iterative cycle.

In practice, (36) is applicable only when $\mathbb{E}_{i' \neq i}[\log p(\mathbf{X}, \mathbf{Z})]$ is computable in closed form and corresponds to a known parametric distribution for which the normalizing constant is computable in closed form. When this is not the case, $p(\mathbf{X}, \mathbf{Z})$ must be replaced by a lower bound for which the resulting approximating distribution becomes tractable.

Let us consider a parametric the lower bound $f(\mathbf{X}, \mathbf{Z}, \mathbf{\Omega})$ of $p(\mathbf{X}, \mathbf{Z})$ such that

$$p(\mathbf{X}, \mathbf{Z}) \geq f(\mathbf{X}, \mathbf{Z}, \mathbf{\Omega}), \tag{37}$$

where $\mathbf{\Omega}$ is a set of auxiliary variables. Using this definition, we define $\mathscr{B}$, which further lower bounds $\mathscr{L}$ as in the following

$$\mathscr{L}(q) \geq \mathscr{B}(q, \mathbf{\Omega}) = \int q(\mathbf{Z}) \log \frac{f(\mathbf{X}, \mathbf{Z}, \mathbf{\Omega})}{q(\mathbf{Z})} d\mathbf{Z}. \tag{38}$$

Using $\mathscr{B}(q, \mathbf{\Omega})$, we rewrite the marginal distribution given in (27) as follows

$$\log p(\mathbf{X}) = \max_{\mathbf{\Omega}} \mathscr{B}(q, \mathbf{\Omega}) + (\mathscr{L}(q) - \max_{\mathbf{\Omega}} \mathscr{B}(q, \mathbf{\Omega})) + KL(q||p). \tag{39}$$

Here we maximize the lower bound $\mathscr{B}(q, \mathbf{\Omega})$ wrt. $\mathbf{\Omega}$, which tightens the lower bound to the free energy $\mathscr{L}(q)$. Using the factorization we introduced in (30) and dissecting out the dependency on one of the factors $q_i(\mathbf{Z}_i)$, we obtain

$$\mathscr{B}(q, \mathbf{\Omega}) = \int q_i \log \frac{\tilde{f}(\mathbf{X}, \mathbf{Z}_i, \mathbf{\Omega})}{q_i} d\mathbf{Z}_i + \text{const}, \tag{40}$$

where $\log \tilde{f}(\mathbf{X}, \mathbf{Z}_i, \mathbf{\Omega})$ is defined as

$$\log \tilde{f}(\mathbf{X}, \mathbf{Z}_i, \mathbf{\Omega}) = \int \log f(\mathbf{X}, \mathbf{Z}, \mathbf{\Omega}) \prod_{i' \neq i} q_{i'} d\mathbf{Z}_{i'} + \text{const},$$

$$= \mathbb{E}_{i' \neq i}[\log f(\mathbf{X}, \mathbf{Z}, \mathbf{\Omega})] + \text{const}, \tag{41}$$

where the normalizing constant is such that $\tilde{f}(\mathbf{X}, \mathbf{Z}_i, \mathbf{\Omega})$ is a proper probability distribution.

Hence, maximizing the lower bound $\mathscr{B}(q, \mathbf{\Omega})$ wrt. $q_i(\mathbf{Z}_i)$ minimizes an approximation to the KL-divergence between $q(\mathbf{Z})$ and the true posterior distribution $p(\mathbf{Z}|\mathbf{X})$.

Finally, we conclude that two alternative steps yields the general equation for the optimal approximating distributions q_i^*. In the first step, we maximize the lower bound $\mathscr{B}(q, \mathbf{\Omega})$ to the free energy $\mathscr{L}(q)$ wrt. the auxiliary variables $\mathbf{\Omega}$, which tightens the lower bound, and in the second step, we maximize $\mathscr{B}(q, \mathbf{\Omega})$ wrt. the approximating distributions q_i.

The optimal approximating distribution $q_i^*(\mathbf{Z}_i)$ is given by

$$q_i^*(\mathbf{Z}_i) = \tilde{f}(\mathbf{X}, \mathbf{Z}_i, \mathbf{\Omega}). \tag{42}$$

In the following, we adopt this strategy for the derivation of the approximating distributions of the multilevel NMF parameters and the source components since $\mathbb{E}[\log p(\mathbf{S}|\mathbf{V})]$ is not tractable in closed form as we will show in Section 2.2.1.

2.2 Application to the Proposed Model

Pursuing the Variational Bayesian approach, we consider the following mean field factorization for $q(\mathbf{Z})$:

$$\begin{aligned} q(\mathbf{Z}) \quad = \quad & \Big(\prod_{f,n} q(\mathbf{s}_{fn})\Big) \\ & \Big(\prod_{f} q(\mathbf{A}_f))\Big) \\ & \Big(\prod_{j,m,n} q(h^{\mathrm{ex}}_{j,mn})\Big) \\ & \Big(\prod_{j,k,m} q(g^{\mathrm{ex}}_{j,km})\Big) \\ & \Big(\prod_{j,k,l} q(u^{\mathrm{ex}}_{j,lk})\Big) \\ & \Big(\prod_{j,f,l} q(w^{\mathrm{ex}}_{j,fl})\Big) \\ & \Big(\prod_{j,m',n} q(h^{\mathrm{ft}}_{j,m'n})\Big) \\ & \Big(\prod_{j,k',m'} q(g^{\mathrm{ft}}_{j,k'm'})\Big) \\ & \Big(\prod_{j,k',l'} q(u^{\mathrm{ft}}_{j,l'k'})\Big) \\ & \Big(\prod_{j,f,l'} q(w^{\mathrm{ft}}_{j,fl'})\Big). \end{aligned} \tag{43}$$

2.2.1 Auxiliary Variables

For the sake of readibility, let us define $\eta = \{k, m, l\}$ the joint index of the excitation NTF parameters and $\eta' = \{k', m', l'\}$ the joint index of the filter NTF parameters. With these joint indices let us further define $v_{j,fn,\eta,\eta'}$ as the product of the NTF parameters

$$v_{j,fn,\eta,\eta'} = h^{\text{ex}}_{j,mn} g^{\text{ex}}_{j,km} u^{\text{ex}}_{j,lk} w^{\text{ex}}_{j,fl} h^{\text{ft}}_{j,m'n} g^{\text{ft}}_{j,k'm'} u^{\text{ft}}_{j,l'k'} w^{\text{ft}}_{j,fl'}. \tag{44}$$

Let us further define $v^{\text{ex}}_{j,fn,\eta}$ as the product of the excitation coefficients of the NTF components and $v^{\text{ft}}_{j,fn,\eta'}$ as the product of the filter coefficients

$$v^{\text{ex}}_{j,fn,\eta} = h^{\text{ex}}_{j,mn} g^{\text{ex}}_{j,km} u^{\text{ex}}_{j,lk} w^{\text{ex}}_{j,fl}, \tag{45}$$
$$v^{\text{ft}}_{j,fn,\eta'} = h^{\text{ft}}_{j,m'n} g^{\text{ft}}_{j,k'm'} u^{\text{ft}}_{j,l'k'} w^{\text{ft}}_{j,fl'}. \tag{46}$$

Having defined these notations, let us have a look at $\mathbb{E}[\log p(\mathbf{S}|\mathbf{V})]$ more closely:

$$\begin{aligned}
\mathbb{E}[\log p(\mathbf{S}|\mathbf{V})] &= \mathbb{E}\Big[\log \prod_{f,n} \mathcal{N}(\mathbf{s}_{fn}|0, \Sigma_{\mathbf{s},fn})\Big] \\
&= \mathbb{E}\Big[\sum_{f,n} \log \mathcal{N}\Big(\mathbf{s}_{fn}|0, \operatorname{diag}\big(\sum_{\eta}\sum_{\eta'} v_{j,fn,\eta,\eta'}\big)\Big)\Big], \\
&= \mathbb{E}\Big[\sum_{f,n} -R\log\pi - \log\det\Big(\operatorname{diag}\big(\sum_{\eta}\sum_{\eta'} v_{j,fn,\eta,\eta'}\big)\Big) \\
&\quad - \mathbf{s}^H_{fn}\operatorname{diag}\big(\sum_{\eta}\sum_{\eta'} v_{j,fn,\eta,\eta'}\big)^{-1}\mathbf{s}_{fn}\Big], \\
&= \sum_{f,n} -R\log\pi + \sum_{j} -R_j\mathbb{E}\Big[\log\sum_{\eta}\sum_{\eta'} v_{j,fn,\eta,\eta'}\Big] \\
&\quad + \sum_{j,r} -\mathbb{E}[|s_{jr,fn}|^2]\mathbb{E}\Big[\frac{1}{\sum_{\eta}\sum_{\eta'} v_{j,fn,\eta,\eta'}}\Big].
\end{aligned} \tag{47}$$

As one can easily see, none of the two expectations in this equation above is tractable. So, we resort to the alternative method, which we introduced in Section 2.1 and lower bound $p(\mathbf{S}|\mathbf{V})$ as proposed in [3].

For the first expectation, given that $x \to -\log x$ is convex, we can lower bound it by its first-order Taylor series expansion around an arbitrary positive point $\omega_{j,fn}$ as follows

$$\begin{aligned}
-\log\sum_{\eta}\sum_{\eta'} v_{j,fn,\eta,\eta'} &\geq -\log\omega_{j,fn} - \frac{1}{\omega_{j,fn}}\big(\sum_{\eta}\sum_{\eta'} v_{j,fn,\eta,\eta'} - \omega_{j,fn}\big), \\
&= -\log\omega_{j,fn} + 1 - \frac{1}{\omega_{j,fn}}\sum_{\eta}\sum_{\eta'} v_{j,fn,\eta,\eta'}.
\end{aligned} \tag{48}$$

For the second expectation, given that $x \to \frac{-1}{x}$ is concave, for any positive $\phi_{j,fn,\eta,\eta'}$ such that $\sum_{\eta}\sum_{\eta'}\phi_{j,fn,\eta,\eta'} = 1$, using Jensen's inequality:

$$-\frac{1}{\sum_\eta \sum_{\eta'} v_{j,fn,\eta,\eta'}} = -\frac{1}{\sum_k \phi_{j,fn,\eta,\eta'} \frac{v_{j,fn,\eta,\eta'}}{\phi_{j,fn,\eta,\eta'}}}$$
$$\geq -\sum_k \phi_{j,fn,\eta,\eta'} \frac{1}{\frac{v_{j,fn,\eta,\eta'}}{\phi_{j,fn,\eta,\eta'}}}$$
$$= -\sum_\eta \sum_{\eta'} \phi^2_{j,fn,\eta,\eta'} \frac{1}{v_{j,fn,\eta,\eta'}}. \tag{49}$$

With these two inequalities, we can lower bound $\log p(\mathbf{S}|\mathbf{V})$ using the auxiliary variables $\mathbf{\Omega} = \{\{\omega_{j,fn}\}_{j,fn}, \{\phi_{j,fn,\eta,\eta'}\}_{j,fn,\eta,\eta'}\}$ as follows

$$\begin{aligned}\log p(\mathbf{S}|\mathbf{V}) \geq & -F \cdot N \cdot R \cdot \log \pi \\ & + \sum_{j,fn} R_j \big(-\log \omega_{j,fn} + 1 - \frac{1}{\omega_{j,fn}} \sum_\eta \sum_{\eta'} v_{j,fn,\eta,\eta'} \big) \\ & + \sum_{j,fn} \sum_r -|s_{jr,fn}|^2 \sum_\eta \sum_{\eta'} \phi^2_{j,fn,\eta,\eta'} \frac{1}{v_{j,fn,\eta,\eta'}}.\end{aligned} \tag{50}$$

2.2.2 Tightening the Bound wrt. the Auxiliary Variables

The update equations for Ω are obtained by maximizing the bound $\mathscr{B}(q, \Omega)$. Concerning $\omega_{j,fn}$, we simply take the partial derivative of the bound $\mathscr{B}(q, \Omega)$ w.r.t. $\omega_{j,fn}$ which is given by

$$\begin{aligned}\mathscr{B}(q, \Omega) = & - F \cdot N \cdot R \cdot \log \pi \\ & + \sum_{j,fn} R_j \big(-\log \omega_{j,fn} + 1 - \frac{1}{\omega_{j,fn}} \sum_\eta \sum_{\eta'} \mathbb{E}[v_{j,fn,\eta,\eta'}] \big) \\ & + \sum_{j,fn} \sum_r -|s_{jr,fn}|^2 \sum_\eta \sum_{\eta'} \phi^2_{j,fn,\eta,\eta'} \mathbb{E}\Big[\frac{1}{v_{j,fn,\eta,\eta'}}\Big] + \text{const.}\end{aligned} \tag{51}$$

So, the derivative is given as

$$\frac{\partial \mathscr{B}(q, \Omega)}{\partial \omega_{j,fn}} = -\frac{R_j}{\omega_{j,fn}} + \frac{R_j}{\omega^2_{j,fn}} \sum_\eta \sum_{\eta'} \mathbb{E}[v_{j,fn,\eta,\eta'}] \tag{52}$$

and make it equal to zero, which yields

$$\omega_{j,fn} = \sum_\eta \sum_{\eta'} \mathbb{E}[v_{j,fn,\eta,\eta'}]. \tag{53}$$

For $\phi_{j,fn,\eta,\eta'}$, we use Lagrange multipliers $\delta_{j,fn}$, because of the constraint. The Lagrangian is given by

$$\begin{aligned}\mathscr{B}(q, \Omega) = & \sum_{j,fn} \big(\sum_r -\mathbb{E}[|s_{jr,fn}|^2] \big) \sum_\eta \sum_{\eta'} \phi^2_{j,fn,\eta,\eta'} \mathbb{E}\Big[\frac{1}{v_{j,fn,\eta,\eta'}}\Big] \\ & + \delta_{j,fn} \Big(\big(\sum_\eta \sum_{\eta'} \phi_{j,fn,\eta,\eta'} \big) - 1 \Big) + \text{const.}\end{aligned} \tag{54}$$

Taking the partial derivative with respect to $\phi_{j,fn,\eta,\eta'}$ and $\delta_{j,fn}$ yields the following system of equations:

$$\frac{\partial \mathscr{B}(q,\Omega)}{\partial \phi_{j,fn,c,\eta,\eta'}} = \Big(\sum_r -\mathbb{E}[|s_{jr,fn}|^2]\Big)\Big(\sum_\eta \sum_{\eta'} \phi^2_{j,fn,\eta,\eta'} \mathbb{E}\Big[\frac{1}{v_{j,fn,\eta,\eta'}}\Big]\Big) \tag{55}$$

$$\frac{\partial \mathscr{B}(q,\Omega)}{\partial \delta_{j,fn}} = \Big(\sum_\eta \sum_{\eta'} \phi_{j,fn,\eta,\eta'}\Big) - 1. \tag{56}$$

Solving the system for $\phi_{j,fn,\eta,\eta'}$ yields

$$\phi_{j,fn,\eta,\eta'} = \frac{1}{C_{j,fn}} \mathbb{E}\Big[\frac{1}{v_{j,fn,\eta,\eta'}}\Big]^{-1}, \tag{57}$$

where $C_{j,fn}$ is the normalization constant given by

$$C_{j,fn} = \sum_\eta \sum_{\eta'} \mathbb{E}\Big[\frac{1}{v_{j,fn,\eta,\eta'}}\Big]^{-1}. \tag{58}$$

2.2.3 Variational Updates for the NTF Parameters

In this section, we will determine the approximating q distributions of the NTF parameters and derive the corresponding update equations. For this we will use the template given in (41, 42).

Update Equations for $w^{\mathrm{ex}}_{j,fl}$: The probability distribution $\tilde{f}(\mathbf{X}, w^{\mathrm{ex}}_{j,fl}, \mathbf{\Omega})$ defined in (41) is given by

$$\begin{aligned}\log q^*(w^{\mathrm{ex}}_{j,fl}) = {} & w^{\mathrm{ex}}_{j,fl} \sum_n \Big(-\frac{R_j}{\omega_{j,fn}} \sum_{k,m} \sum_{\eta'} \mathbb{E}[h^{\mathrm{ex}}_{j,mn} g^{\mathrm{ex}}_{j,km} u^{\mathrm{ex}}_{j,lk} v^{\mathrm{ft}}_{j,fn,\eta'}]\Big) \\ & - \sum_n \sum_r \frac{\mathbb{E}[|s_{jr,fn}|^2]}{w^{\mathrm{ex}}_{j,fl}} \sum_{k,m} \sum_{\eta'} \phi^2_{j,fn,\eta,\eta'} \mathbb{E}\Big[\frac{1}{h^{\mathrm{ex}}_{j,mn} g^{\mathrm{ex}}_{j,km} u^{\mathrm{ex}}_{j,lk} v^{\mathrm{ft}}_{j,fn,\eta'}}\Big] \\ & - \log w^{\mathrm{ex}}_{j,fl} + \text{const.}\end{aligned} \tag{59}$$

Observing this distribution, one can see that it involves a linear term in $\log w^{\mathrm{ex}}_{j,fl}$, a linear term in $w^{\mathrm{ex}}_{j,fl}$ and a linear term in $\frac{1}{w^{\mathrm{ex}}_{j,fl}}$. The optimal approximating distribution $q^*(w^{\mathrm{ex}}_{j,fl})$ is hence an instance of the generalized inverse Gaussian (GIG) distribution, whose probability distribution function (PDF) is given by

$$GIG(y;\gamma,\rho,\tau) = \frac{\exp\{(\gamma-1)\log y - \rho y - \frac{\tau}{y}\}\rho^{\frac{\gamma}{2}}}{2\tau^{\frac{\gamma}{2}} K_\gamma(2\sqrt{\rho\tau})}, \tag{60}$$

for $y \geq 0$, $\rho \geq 0$ and $\tau \geq 0$, where $K_\gamma(\cdot)$ is the modified Bessel function of the second kind. The gamma distribution is a special case of the GIG distribution [4] when $\tau = 0$ and $\gamma > 0$. Similarly, the inverse gamma distribution is another special case [4] when $\rho = 0$ and $\gamma < 0$.

In the exponent of the PDF of the GIG distribution, we can match $\frac{\tau}{y}$ to the first line of (59) in a "completing the square" fashion [2]. Similarly, ρy is matched to the second line of (59). Finally, the last line of (59) is matched to $(\gamma-1)\log y$. By doing this we obtain the following update equations for the parameter of $q^*(w^{\mathrm{ex}}_{j,fl})$ $\tau^{\mathrm{ex}}_{w,j,fl}$, $\rho^{\mathrm{ex}}_{w,j,fl}$ and $\gamma^{\mathrm{ex}}_{w,j,fl}$

$$\tau^{\mathrm{ex}}_{w,j,fl} = \sum_n \sum_r \mathbb{E}[|s_{jr,fn}|^2] \sum_{k,m} \sum_{\eta'} \phi^2_{j,fn,\eta,\eta'} \mathbb{E}\Big[\frac{1}{h^{\mathrm{ex}}_{j,mn} g^{\mathrm{ex}}_{j,km} u^{\mathrm{ex}}_{j,lk} v^{\mathrm{ft}}_{j,fn,\eta'}}\Big], \tag{61}$$

$$\rho^{\mathrm{ex}}_{w,j,fl} = \sum_n \frac{R_j}{\omega_{j,fn}} \sum_{k,m} \sum_{\eta'} \mathbb{E}[h^{\mathrm{ex}}_{j,mn} g^{\mathrm{ex}}_{j,km} u^{\mathrm{ex}}_{j,lk} v^{\mathrm{ft}}_{j,fn,\eta'}], \tag{62}$$

$$\gamma^{\mathrm{ex}}_{w,j,fl} = 0. \tag{63}$$

Embedding the update equation for $\phi_{j,fn,\eta,\eta'}$ given in (57) into (61), we obtain the following:

$$\begin{aligned}
\tau^{\mathrm{ex}}_{w,j,fl} &= \sum_n \Bigg[\Big(\sum_r \mathbb{E}[|s_{jr,fn}|^2] \Big) \\
&\quad \sum_{k,m} \sum_{\eta'} \Bigg(\bigg(\frac{1}{C_{j,fn}} \mathbb{E}\Big[\frac{1}{h^{\mathrm{ex}}_{j,mn} g^{\mathrm{ex}}_{j,km} u^{\mathrm{ex}}_{j,lk} w^{\mathrm{ex}}_{j,fl} v^{\mathrm{ft}}_{j,fn,\eta'}}\Big]^{-1} \bigg)^2 \\
&\quad \mathbb{E}\Big[\frac{1}{h^{\mathrm{ex}}_{j,c,mn} g^{\mathrm{ex}}_{j,km} u^{\mathrm{ex}}_{j,lk} v^{\mathrm{ft}}_{j,fn,\eta'}}\Big] \Bigg) \Bigg] \\
&= \mathbb{E}\Big[\frac{1}{w^{\mathrm{ex}}_{j,c,fl}}\Big]^{-2} \sum_n \Bigg(\frac{1}{{C_{j,fn}}^2} \Big(\sum_r \mathbb{E}[|s_{jr,fn}|^2] \Big) \\
&\quad \sum_{k,m} \sum_{\eta'} \mathbb{E}\Big[\frac{1}{h^{\mathrm{ex}}_{j,mn} g^{\mathrm{ex}}_{j,km} u^{\mathrm{ex}}_{j,lk} v^{\mathrm{ft}}_{j,fn,\eta'}}\Big]^{-1} \Bigg).
\end{aligned} \tag{64}$$

Finally, we can write this update equation in matrix form as shown below:

$$\begin{aligned}
\boldsymbol{\tau}^{\mathrm{ex}}_{w,j} &= \mathbb{E}\Big[\frac{1}{\mathbf{W}^{\mathrm{ex}}_j}\Big]^{.-2} \odot \Bigg(\quad \Big(\mathbb{E}[|\mathbf{S}_j|^{.2}] \odot \mathbf{C}_j^{.-2} \odot \mathbb{E}\Big[\frac{1}{\mathbf{V}^{\mathrm{ft}}_j}\Big]^{.-1} \Big) \\
&\qquad \Big(\mathbb{E}\Big[\frac{1}{\mathbf{U}^{\mathrm{ex}}_j}\Big]^{.-1} \mathbb{E}\Big[\frac{1}{\mathbf{G}^{\mathrm{ex}}_j}\Big]^{.-1} \mathbb{E}\Big[\frac{1}{\mathbf{H}^{\mathrm{ex}}_j}\Big]^{.-1} \Big)^T \Bigg).
\end{aligned} \tag{65}$$

Note that in (65), the power operations like $\mathbf{X}^{.-a}$ are element-wise operations. Furthermore, the symbol $\odot$ means element-wise matrix multiplication.

Similarly, replacing $\omega_{j,fn}$ in (66) with its update equation shown in (53) yields the following:

$$\rho^{\mathrm{ex}}_{w,j,fl} = \sum_n \Big(\frac{R_j}{\sum_\eta \sum_{\eta'} \mathbb{E}[v_{j,fn,\eta,\eta'}]} \sum_{k,m} \sum_{\eta'} \mathbb{E}[h^{\mathrm{ex}}_{j,mn} g^{\mathrm{ex}}_{j,km} u^{\mathrm{ex}}_{j,lk} v^{\mathrm{ft}}_{j,fn,\eta'}] \Big). \tag{66}$$

The update equation of $\boldsymbol{\rho}^{\mathrm{ex}}_{w,j}$ is written in the matrix form as follows:

$$\boldsymbol{\rho}^{\mathrm{ex}}_{w,j} = R_j \mathbb{E}[\mathbf{V}^{\mathrm{ex}}_j]^{.-1} \big(\mathbb{E}[\mathbf{U}^{\mathrm{ex}}_j] \mathbb{E}[\mathbf{G}^{\mathrm{ex}}_j] \mathbb{E}[\mathbf{H}^{\mathrm{ex}}_j] \big)^T. \tag{67}$$

For the other NTF parameters, the derivations are performed by following the same steps. Therefore, in the following, we skip the details of the derivations and give only the final update equations.

Update Equations for $u^{\text{ex}}_{j,lk}$: The update equation for $\tau^{\text{ex}}_{u,j,lk}$ is given in the following

$$\begin{aligned}\tau^{\text{ex}}_{u,j,lk} =&\mathbb{E}\Big[\frac{1}{u^{\text{ex}}_{j,lk}}\Big]^{-2}\sum_{f,n}\Big(\frac{1}{C_{j,fn}{}^2}\sum_r \mathbb{E}[|s_{jr,fn}|^2]\\&\sum_m\sum_{\eta'}\mathbb{E}\Big[\frac{1}{h^{\text{ex}}_{j,mn}g^{\text{ex}}_{j,km}w^{\text{ex}}_{j,fl}v^{\text{ft}}_{j,fn,\eta'}}\Big]^{-1}\Big).\end{aligned}\tag{68}$$

The matrix version of this update equation is written as

$$\begin{aligned}\boldsymbol{\tau}^{\text{ex}}_{u,j} = \mathbb{E}_q\Big[\frac{1}{\mathbf{U}^{\text{ex}}_j}\Big]^{.-2}\odot\Bigg[&\Big(\mathbb{E}_q\Big[\frac{1}{\mathbf{W}^{\text{ex}}_{j,c}}\Big]^{.-1}\Big)^T\Big(\mathbb{E}[|\mathbf{S}_j|^{.2}]\odot\mathbf{C}_j^{.-2}\odot\mathbb{E}\Big[\frac{1}{\mathbf{V}^{\text{ft}}_{j,c}}\Big]^{.-1}\Big)\\&\Big(\mathbb{E}\Big[\frac{1}{\mathbf{G}^{\text{ex}}_{j,c}}\Big]^{.-1}\mathbb{E}\Big[\frac{1}{\mathbf{H}^{\text{ex}}_{j,c}}\Big]^{.-1}\Big)^T\Bigg].\end{aligned}\tag{69}$$

The update equation for $\rho^{\text{ex}}_{u,j,lk}$ is given in the following

$$\rho^{\text{ex}}_{u,j,lk} = \sum_{f,n}\Big(\frac{R_j}{\sum_\eta\sum_{\eta'}\mathbb{E}[v_{j,fn,\eta,\eta'}]}\sum_m\sum_{\eta'}\mathbb{E}[h^{\text{ex}}_{j,mn}g^{\text{ex}}_{j,km}w^{\text{ex}}_{j,fl}v^{\text{ft}}_{j,fn,\eta'}]\Big).\tag{70}$$

The matrix form if given by

$$\boldsymbol{\rho}^{\text{ex}}_{u,j} = R_j\mathbb{E}[\mathbf{W}_j]^T\mathbb{E}[\mathbf{V}^{\text{ex}}_j]^{.-1}\big(\mathbb{E}[\mathbf{G}^{\text{ex}}_j]\mathbb{E}[\mathbf{H}^{\text{ex}}_j]\big)^T.\tag{71}$$

Finally,

$$\gamma^{\text{ex}}_{u,j,lk} = 0.\tag{72}$$

Update Equations for $g^{\text{ex}}_{j,km}$: The update equation for $\tau^{\text{ex}}_{g,j,km}$ is written as follows

$$\begin{aligned}\tau^{\text{ex}}_{g,j,km} = \mathbb{E}\Big[\frac{1}{g^{\text{ex}}_{j,km}}\Big]^{-2}\sum_{f,n}\Big(\frac{1}{C_{j,fn}{}^2}\big(\sum_r\mathbb{E}[|s_{jr,fn}|^2]\big)\\\sum_l\sum_{\eta'}\mathbb{E}\Big[\frac{1}{h^{\text{ex}}_{j,mn}u^{\text{ex}}_{j,lk}w^{\text{ex}}_{j,fl}v^{\text{ft}}_{j,fn,\eta'}}\Big]^{-1}\Big).\end{aligned}\tag{73}$$

The matrix version is

$$\boldsymbol{\tau}^{\text{ex}}_{g,j} = \mathbb{E}\Big[\frac{1}{\mathbf{G}^{\text{ex}}_j}\Big]^{.-2} \odot \Bigg(\Big(\mathbb{E}\Big[\frac{1}{\mathbf{W}^{\text{ex}}_j}\Big]^{.-1}\mathbb{E}\Big[\frac{1}{\mathbf{U}^{\text{ex}}_j}\Big]^{.-1}\Big)^T \Big(\mathbb{E}[|\mathbf{S}_j|^{.2}] \odot \mathbf{C}_j^{.-2} \odot \mathbb{E}\Big[\frac{1}{\mathbf{V}^{\text{ft}}_j}\Big]^{.-1}\Big)\Big(\mathbb{E}\Big[\frac{1}{\mathbf{H}^{\text{ex}}_j}\Big]^{.-1}\Big)^T\Bigg). \tag{74}$$

The update equation for $\rho^{\text{ex}}_{g,j,km}$ is as follows

$$\rho^{\text{ex}}_{g,j,km} = \sum_{f,n}\Big(\frac{R_j}{\sum_\eta\sum_{\eta'}\mathbb{E}[v_{j,fn,\eta,\eta'}]}\sum_l\sum_{\eta'}\mathbb{E}_q[h^{\text{ex}}_{j,mn}u^{\text{ex}}_{j,lk}w^{\text{ex}}_{j,fl}v^{\text{ft}}_{j,fn,\eta'}]\Big). \tag{75}$$

The matrix version is

$$\boldsymbol{\rho}^{\text{ex}}_{g,j} = R_j(\mathbb{E}[\mathbf{W}^{\text{ex}}_j]\mathbb{E}[\mathbf{U}^{\text{ex}}_j])^T\mathbb{E}[\mathbf{V}^{\text{ex}}_j]^{.-1}\mathbb{E}[\mathbf{H}^{\text{ex}}_j]^T. \tag{76}$$

Similar to the other NTF parameters above,

$$\gamma^{\text{ex}}_{g,j,km} = 0. \tag{77}$$

Update Equations for $h^{\text{ex}}_{j,mn}$: The update equation for $\tau^{\text{ex}}_{h,j,mn}$ is expressed as follows

$$\tau^{\text{ex}}_{h,j,mn} = \mathbb{E}\Big[\frac{1}{h^{\text{ex}}_{j,mn}}\Big]^{-2}\sum_f\Big(\frac{1}{C_{j,fn}{}^2}(\sum_r\mathbb{E}[|s_{jr,fn}|^2]) \sum_{k,l}\sum_{\eta'}\mathbb{E}\Big[\frac{1}{g^{\text{ex}}_{j,km}u^{\text{ex}}_{j,lk}w^{\text{ex}}_{j,fl}v^{\text{ft}}_{j,fn,\eta'}}\Big]^{-1}\Big). \tag{78}$$

The matrix version

$$\boldsymbol{\tau}^{\text{ex}}_{h,j} = \mathbb{E}\Big[\frac{1}{\mathbf{H}^{\text{ex}}_j}\Big]^{.-2} \odot \Bigg(\Big(\mathbb{E}\Big[\frac{1}{\mathbf{W}^{\text{ex}}_j}\Big]^{.-1}\mathbb{E}\Big[\frac{1}{\mathbf{U}^{\text{ex}}_j}\Big]^{.-1}\mathbb{E}\Big[\frac{1}{\mathbf{G}^{\text{ex}}_j}\Big]^{.-1}\Big)^T \Big(\mathbb{E}[|\mathbf{S}_j|^{.2}] \odot \mathbf{C}_j^{.-2} \odot \mathbb{E}\Big[\frac{1}{\mathbf{V}^{\text{ft}}_j}\Big]^{.-1}\Big)\Bigg). \tag{79}$$

The update equation for $\rho^{\text{ex}}_{h,j,mn}$ is as follows

$$\rho^{\text{ex}}_{h,j,mn} = \sum_{f,n}\Big(\frac{R_j}{\sum_\eta\sum_{\eta'}\mathbb{E}[v_{j,fn,\eta,\eta'}]}\sum_{k,l}\sum_{\eta'}\mathbb{E}[g^{\text{ex}}_{j,km}u^{\text{ex}}_{j,lk}w^{\text{ex}}_{j,fl}v^{\text{ft}}_{j,fn,\eta'}]\Big). \tag{80}$$

The matrix version is

$$\boldsymbol{\rho}^{\text{ex}}_{h,j} = R_j\big(\mathbb{E}[\mathbf{W}^{\text{ex}}_j]\mathbb{E}[\mathbf{U}^{\text{ex}}_j]\mathbb{E}[\mathbf{G}^{\text{ex}}_j]\big)^T\mathbb{E}[\mathbf{V}^{\text{ex}}_j]^{.-1}. \tag{81}$$

As with all the other NTF parameters,

$$\gamma^{\text{ex}}_{h,j,km} = 0. \tag{82}$$

Expectations to be computed in the update equations of the NTF parameters The expectation $\mathbb{E}[|s_{jr,fn}|^{.2}]$ in (61), (68), (73) and (78) is calculated as

$$\mathbb{E}[|s_{jr,fn}|^2] = |(\boldsymbol{\mu}_{s,fn})_{jr}|^2 + (\mathbf{R}_{\mathbf{ss},fn})_{jr,jr} \tag{83}$$

where $\boldsymbol{\mu}_{\mathbf{s},fn}$ and $\mathbf{R}_{\mathbf{ss},fn}$ are the first and second order posterior moments of $\mathbf{s}_{fn}$ derived below in (91) and (92). The expectations involving the NTF parameters in (62), (70), (75), (80) and (61), (68), (73) and (78) are computed via the following formulae for the GIG distribution [4]:

$$\mathbb{E}[y] = \frac{\mathcal{K}_{\gamma+1}(2\sqrt{\rho\tau})\sqrt{\tau}}{\mathcal{K}_{\gamma}(2\sqrt{\rho\tau})\sqrt{\rho}}, \tag{84}$$

$$\mathbb{E}\Big[\frac{1}{y}\Big] = \frac{\mathcal{K}_{\gamma-1}(2\sqrt{\rho\tau})\sqrt{\rho}}{\mathcal{K}_{\gamma}(2\sqrt{\rho\tau})\sqrt{\tau}}. \tag{85}$$

$\mathbb{E}[\mathbf{V}_j^{\mathrm{ex}}]$ is obtained as

$$\mathbb{E}[\mathbf{V}_j^{\mathrm{ex}}] = \mathbb{E}[\mathbf{W}_j^{\mathrm{ex}}]\mathbb{E}[\mathbf{U}_j^{\mathrm{ex}}]\mathbb{E}[\mathbf{G}_j^{\mathrm{ex}}]\mathbb{E}[\mathbf{H}_j^{\mathrm{ex}}] \tag{86}$$

and $\mathbb{E}[1/\mathbf{V}_j^{\mathrm{ft}}]^{.-1}$ is a shorthand notation for

$$\mathbb{E}\Big[\frac{1}{\mathbf{W}_j^{\mathrm{ft}}}\Big]^{.-1}\mathbb{E}\Big[\frac{1}{\mathbf{U}_j^{\mathrm{ft}}}\Big]^{.-1}\mathbb{E}\Big[\frac{1}{\mathbf{G}_j^{\mathrm{ft}}}\Big]^{.-1}\mathbb{E}\Big[\frac{1}{\mathbf{H}_j^{\mathrm{ft}}}\Big]^{.-1}. \tag{87}$$

For the filter NTF parameters, the derivation of the update equations is similar to the derivation of the excitation NTF parameters and so are the final update equations. Therefore, we skip them here.

2.2.4 Variational Updates for the Source Components

The distribution $\tilde{f}(\mathbf{X}, \mathbf{s}_{fn}, \boldsymbol{\Omega})$ of the source components $\mathbf{s}_{fn}$ is given by

$$\begin{aligned}\log q^*(\mathbf{s}_{fn}) = &\ \mathbf{s}_{fn}^H \boldsymbol{\mu}_{\mathbf{A},f}^H (\sigma_b^2\mathbf{I})^{-1}\mathbf{x}_{fn} + \mathbf{x}_{fn}^H(\sigma_b^2\mathbf{I})^{-1}\boldsymbol{\mu}_{\mathbf{A},f}\mathbf{s}_{fn} \\ &- \frac{1}{\sigma_b^2}(\mathbf{s}_{fn}^H\mathbf{R}_{\mathbf{A},f}\mathbf{s}_{fn}) \\ &- \sum_j\sum_r |s_{jr,fn}|^2 \sum_\eta\sum_{\eta'} \phi_{j,fn,\eta,\eta'}^2 \mathbb{E}\Big[\frac{1}{v_{j,fn,\eta,\eta'}}\Big] + \text{const.}\end{aligned} \tag{88}$$

Replacing $\phi_{j,fn,\eta,\eta'}$ with its value shown in (57) yields

$$\begin{aligned}\log q^*(\mathbf{s}_{fn}) = &\ \mathbf{s}_{fn}^H \boldsymbol{\mu}_{\mathbf{A},f}^H (\sigma_b^2\mathbf{I})^{-1}\mathbf{x}_{fn} + \mathbf{x}_{fn}^H(\sigma_b^2\mathbf{I})^{-1}\boldsymbol{\mu}_{\mathbf{A},f}\mathbf{s}_{fn} \\ &- \frac{1}{\sigma_b^2}(\mathbf{s}_{fn}^H\mathbf{R}_{\mathbf{A},f}\mathbf{s}_{fn}) \\ &- \mathbf{s}_{fn}^H {\mathbf{C}_{fn}}^{-1}\mathbf{s}_{fn} + \text{const},\end{aligned} \tag{89}$$

where $\boldsymbol{\mu}_{\mathbf{A},f}$ and $\mathbf{R}_{\mathbf{A},f}$ are the first and second order posterior raw moments of $\mathbf{A}_f$ derived below in (101) and (99), and ${\mathbf{C}_{fn}}^{-1} = \mathrm{diag}({C_{j,fn}}^{-1})_{r=1}^R$ is a diagonal matrix with the main diagonal containing the normalization factor ${C_{j,fn}}^{-1}$ repeated R_j times for each j.

This distribution involves a linear term in $\mathbf{s}_{fn}$, its conjugate, and quadratic terms. The optimal approximating distribution is thus a Gaussian given by

$$\mathbf{s}_{fn} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{s},fn}, \mathbf{R}_{\mathbf{ss},fn}). \tag{90}$$

By "completing the square" w.r.t. $\boldsymbol{\mu}_{\mathbf{s},fn}$ and $\mathbf{R}_{\mathbf{ss},fn}$ in (89) we derive the update equations for these two parameters. First we derive the update equation of the covariance. For this, we rearrange the quadratic terms make them equal to $-\mathbf{s}_{fn}^H \mathbf{R}_{\mathbf{ss},fn}^{-1} \mathbf{s}_{fn}$. By doing this we obtain the covariance given in the following

$$\begin{aligned}\mathbf{R}_{\mathbf{ss},fn} &= ({\mathbf{C}_{fn}}^{-1} + (\sigma_b^2 \mathbf{I})^{-1} \mathbf{R}_{\mathbf{A},f})^{-1} \\ &= \mathbf{C}_{fn} - \mathbf{C}_{fn}(\sigma_b^2 \mathbf{I} + \mathbf{R}_{\mathbf{A},f}\mathbf{C}_{fn})^{-1} \mathbf{R}_{\mathbf{A},f}\mathbf{C}_{fn},\end{aligned} \tag{91}$$

Note that we used the Woodbury matrix identity [2] for the inversion here. Similarly for the update equation of the mean, we arrange the linear terms of the mean and make them equal to $\mathbf{s}_{fn}^H \mathbf{R}_{\mathbf{ss},fn}^{-1} \boldsymbol{\mu}_{\mathbf{s},fn} + \boldsymbol{\mu}_{\mathbf{s},fn}^H \mathbf{R}_{\mathbf{ss},fn}^{-1} \mathbf{s}_{fn}$. This yields the following update equation for the mean

$$\boldsymbol{\mu}_{\mathbf{s},fn} = \mathbf{R}_{\mathbf{ss},fn} \boldsymbol{\mu}_{\mathbf{A},f}^H (\sigma_b^2 \mathbf{I})^{-1} \mathbf{x}_{fn}. \tag{92}$$

2.2.5 Variational Updates for the Mixing Parameters

The distribution $\tilde{f}(\mathbf{X}, \underline{\mathbf{A}}_f, \boldsymbol{\Omega})$ of the reshaped mixing parameters $\underline{\mathbf{A}}_f$ is given by

$$\begin{aligned}\log q^*(\underline{\mathbf{A}}_f) = &-\underline{\mathbf{A}}_f^H \frac{1}{\sigma_b^2} \sum_n \mathrm{diag}(\underbrace{\mathbf{R}_{\mathbf{s},fn}^T, \ldots, \mathbf{R}_{\mathbf{s},fn}^T}_{\text{I times}}) \underline{\mathbf{A}}_f \\ &+ \underline{\mathbf{A}}_f^H \frac{1}{\sigma_b^2} \sum_n \mathbf{R}_{\mathbf{xs},fn} + \frac{1}{\sigma_b^2} \Big(\sum_n \mathbf{R}_{\mathbf{xs},fn}^H \Big) \underline{\mathbf{A}}_f \\ &+ \text{const},\end{aligned} \tag{93}$$

where $\mathbf{R}_{\mathbf{s},fn}$ is the second raw moment of the source coefficients and is given by

$$\mathbf{R}_{\mathbf{s},fn} = \boldsymbol{\mu}_{\mathbf{s},nf} \boldsymbol{\mu}_{\mathbf{s},nf}^H + \mathbf{R}_{\mathbf{ss},fn}. \tag{94}$$

Similarly, the cross moment $\mathbf{R}_{\mathbf{xs},fn}$ is given by

$$\mathbf{R}_{\mathbf{xs},fn} = [x_{1,fn} \boldsymbol{\mu}_{\mathbf{s},fn}^H, \ldots, x_{I,fn} \boldsymbol{\mu}_{\mathbf{s},fn}^H]^T. \tag{95}$$

Note that the contribution of the prior distribution vanishes, because we assumed a flat prior. Also note that due to the reshaping of the mixing matrix $\mathbf{A}_f$ into a vector $\underline{\mathbf{A}}_f$, matrix multiplications *e.g.* $\mathbf{A}_f \mathbf{s}_{fn}$ became multiplications of two vectors. As a result of the dimensionality mismatch the vector multiplication cannot be performed any more. Note that in the given example each row of the mixing matrix, which correspond to each channel of the mixture respectively is multiplied by the source. Therefore, we define the term $\text{diag}(\underbrace{\mathbf{R}^T_{\mathbf{s},fn}, \ldots, \mathbf{R}^T_{\mathbf{s},fn}}_{\text{I times}})$ indicating a block diagonal matrix, where $\mathbf{R}_{\mathbf{s},fn}$ is repeated on the diagonal I times. Similarly, $\mathbf{R}_{\mathbf{xs},fn}$ is composed of multiplying $\boldsymbol{\mu}^H_{\mathbf{s},fn}$ with each channel of the mixture.

The distribution given in (93) involves a linear term in $\underline{\mathbf{A}}_f$, its conjugate, and a quadratic term. Hence, the optimal approximating distribution is a Gaussian given by

$$\underline{\mathbf{A}}_f \sim \mathcal{N}(\boldsymbol{\mu}_{\underline{\mathbf{A}},f}, \mathbf{R}_{\underline{\mathbf{A}}\underline{\mathbf{A}},f}). \tag{96}$$

By "completing the square" wrt. $\boldsymbol{\mu}_{\underline{\mathbf{A}},f}$ and $\mathbf{R}_{\underline{\mathbf{A}}\underline{\mathbf{A}},f}$ in (93) we derive the update equations for these two parameters. First we derive the update equation of the covariance. For this, we rearrange the quadratic terms make them equal to $-\underline{\mathbf{A}}^H_f \mathbf{R}^{-1}_{\underline{\mathbf{A}}\underline{\mathbf{A}},f} \underline{\mathbf{A}}_f$. By doing this we obtain the covariance given in the following

$$\mathbf{R}_{\underline{\mathbf{A}}\underline{\mathbf{A}},f} = \Big(\frac{1}{\sigma_b^2} \sum_n \text{diag}\,\underbrace{(\mathbf{R}^T_{\mathbf{s},fn}, \ldots, \mathbf{R}^T_{\mathbf{s},fn})}_{\text{I times}}\Big)^{-1}, \tag{97}$$

Similarly, for the mean, we rearrange the linear terms and make them equal to $\underline{\mathbf{A}}^H_f {\mathbf{R}_{\underline{\mathbf{A}}\underline{\mathbf{A}},f}}^{-1} \boldsymbol{\mu}_{\underline{\mathbf{A}},f} + \boldsymbol{\mu}^H_{\underline{\mathbf{A}},f} {\mathbf{R}_{\underline{\mathbf{A}}\underline{\mathbf{A}},f}}^{-1} \underline{\mathbf{A}}_f$ which yields

$$\boldsymbol{\mu}_{\underline{\mathbf{A}},f} = \mathbf{R}_{\underline{\mathbf{A}}\underline{\mathbf{A}},f} \frac{1}{\sigma_b^2} \sum_n \mathbf{R}_{\mathbf{xs},fn}, \tag{98}$$

Recall that we reshaped the mixing matrix into a vector and calculated the optimal actor for the distribution we defined for the vector form mixing system. Therefore, we need to reshape the mixing parameters into a matrix and consider the inter-channel relationships within the mixing matrix to calculate the expectations. Hence, from the given equations above, we can derive the second order raw moment $\mathbf{R}_{\mathbf{A},f}$ as in the following

$$\mathbf{R}_{\mathbf{A},f} = \sum_i \left(\Big[\mathbb{E}[\underline{\mathbf{A}}_f \underline{\mathbf{A}}^H_f\Big]_{ii} \right)^T, \tag{99}$$

where $[\cdot]_{ii}$ indicates the $R \times R$ diagonal block corresponding to the channel i and

$$\mathbb{E}[\underline{\mathbf{A}}_f \underline{\mathbf{A}}^H_f] = \boldsymbol{\mu}_{\underline{\mathbf{A}},f} {\boldsymbol{\mu}_{\underline{\mathbf{A}},f}}^H + \mathbf{R}_{\underline{\mathbf{A}}\underline{\mathbf{A}},f}. \tag{100}$$

Similarly, the first order moment $\boldsymbol{\mu}_{\mathbf{A},f}$ is simply obtained by reshaping $\boldsymbol{\mu}^H_{\underline{\mathbf{A}},f}$ back into a matrix

$$\boldsymbol{\mu}_{\mathbf{A},f}(i, 1 \ldots R) = (\boldsymbol{\mu}_{\underline{\mathbf{A}},f})_{((i-1)R+1,\ldots,iR)}. \tag{101}$$

2.3 Lower Bound

We calculate the new lower bound $\mathscr{B}(q, \Omega)$ after each iteration. Here we compare the new lower bound with the previous one and set a termination condition according to this change. The lower bound should never decrease. This fact enables us to check the update equations and their implementation for their correctness. In general, if the increase in the lower bound is insignificant compared to previous iterations, we stop. The lower bound has already been defined in the following

$$
\begin{aligned}
\mathscr{B}(q,\Omega) = \; & \mathbb{E}[\log p(\mathbf{X}|\mathbf{S},\mathbf{A})] \\
& + \mathbb{E}\Big[- F\cdot N\cdot R\cdot \log \pi \\
& + \sum_{j,fn} R_j \big(-\log \omega_{j,fn} + 1 - \frac{1}{\omega_{j,fn}} \sum_{\eta}\sum_{\eta'} v_{j,fn,\eta,\eta'} \big) \\
& + \sum_{j,fn}\sum_{r} -|s_{jr,fn}|^2 \sum_{\eta}\sum_{\eta'} \phi^2_{j,fn,\eta,\eta'} \frac{1}{v_{j,fn,\eta,\eta'}} \Big] \\
& - \mathbb{E}[\log q(\mathbf{S})] \\
& + \mathbb{E}[\log p(\mathbf{A})] - \mathbb{E}[\log q(\mathbf{A})] \\
& + \mathbb{E}[\log p(\mathbf{W}^{\mathrm{ex}})] - \mathbb{E}[\log q(\mathbf{W}^{\mathrm{ex}})] \\
& + \mathbb{E}[\log p(\mathbf{W}^{\mathrm{ft}})] - \mathbb{E}[\log q(\mathbf{W}^{\mathrm{ft}})] \\
& + \mathbb{E}[\log p(\mathbf{U}^{\mathrm{ex}})] - \mathbb{E}[\log q(\mathbf{U}^{\mathrm{ex}})] \\
& + \mathbb{E}[\log p(\mathbf{U}^{\mathrm{ft}})] - \mathbb{E}[\log q(\mathbf{U}^{\mathrm{ft}})] \\
& + \mathbb{E}[\log p(\mathbf{G}^{\mathrm{ex}})] - \mathbb{E}[\log q(\mathbf{G}^{\mathrm{ex}})] \\
& + \mathbb{E}[\log p(\mathbf{G}^{\mathrm{ft}})] - \mathbb{E}[\log q(\mathbf{G}^{\mathrm{ft}})] \\
& + \mathbb{E}[\log p(\mathbf{H}^{\mathrm{ex}})] - \mathbb{E}[\log q(\mathbf{H}^{\mathrm{ex}})] \\
& + \mathbb{E}[\log p(\mathbf{H}^{\mathrm{ft}})] - \mathbb{E}[\log q(\mathbf{H}^{\mathrm{ft}})].
\end{aligned} \tag{102}
$$

Note that for $\mathbb{E}[p(\mathbf{S}|\mathbf{V})]$ we used the approximation we derived which is shown in (51).

Also note that for all the variational variables, the expectations of the log of the approximating q distributions correspond to their negative entropies. Note also that the lower bound consists of the the expectation of the logarithm of the prior distribution (p distribution) minus the expectation of the logarithm of the variational distribution (q distribution) pairs except for the log-likelihood. For the log-likelihood, we calculate the expectation as follows

$$
\begin{aligned}
\mathbb{E}[\log p(\mathbf{X}|\mathbf{S},\mathbf{A})] &= \mathbb{E}\Big[\sum_{f,n} \log \mathcal{N}(\mathbf{x}_{fn}|\mathbf{A}_f\mathbf{s}_{fn}, \sigma_b^2\mathbf{I})\Big],\\
&= -\sum_{f,n} \big(\log \pi\sigma_b^2 + \frac{1}{\sigma_b^2}\mathbb{E}[(\mathbf{x}_{fn} - \mathbf{A}_f\mathbf{s}_{fn})^H(\mathbf{x}_{fn} - \mathbf{A}_f\mathbf{s}_{fn})]\big),\\
&= -\sum_{f,n} \big(\log \pi\sigma_b^2 + \frac{1}{\sigma_b^2}(\mathbf{x}_{fn}^H\mathbf{x}_{fn} - \mathbf{x}_{fn}^H\mathbb{E}[\mathbf{A}_f]\mathbb{E}[\mathbf{s}_{fn}]\\
&\quad - \mathbb{E}[\mathbf{s}_{fn}^H]\mathbb{E}[\mathbf{A}_f^H]\mathbf{x}_{fn} + \mathbb{E}[\mathbf{s}_{fn}^H(\mathbf{A}_f)^H\mathbf{A}_f\mathbf{s}_{fn}])\big),\\
&= -\sum_{f,n} \log \pi\sigma_b^2 + \frac{1}{\sigma_b^2}\big(\mathbf{x}_{fn}^H\mathbf{x}_{fn} - \mathbf{x}_{fn}^H\boldsymbol{\mu}_{\mathbf{A},f}\boldsymbol{\mu}_{\mathbf{s},fn} - \boldsymbol{\mu}_{\mathbf{s},fn}^H\boldsymbol{\mu}_{\mathbf{A},f}^H\mathbf{x}_{fn}\\
&\quad + \mathrm{tr}(\mathbb{E}[\mathbf{s}_{fn}\mathbf{s}_{fn}^H]\mathbb{E}[\mathbf{A}_f^H\mathbf{A}_f])\big),\\
&= -F\cdot N\cdot I\cdot \log \pi\sigma_b^2 - \sum_{f,n}\frac{1}{\sigma_b^2}\Big(\mathbf{x}_{fn}^H\mathbf{x}_{fn} - \mathbf{x}_{fn}^H\boldsymbol{\mu}_{\mathbf{A},f}\boldsymbol{\mu}_{\mathbf{s},fn}\\
&\quad - \boldsymbol{\mu}_{\mathbf{s},fn}^H\boldsymbol{\mu}_{\mathbf{A},f}^H\mathbf{x}_{fn} + \mathrm{tr}\big((\boldsymbol{\mu}_{s,nf}\boldsymbol{\mu}_{\mathbf{s},nf}^H + \mathbf{R}_{\mathbf{ss},fn})(\mathbf{R}_{\mathbf{A},f})\big)\Big).
\end{aligned}
\tag{103}
$$

For the expectation of the approximation of $p(\mathbf{S}|\mathbf{V})$, –the second, third and fourth lines of (102)– we first replace the auxiliary variables with their values given in (53, 57) as follows

$$
\begin{aligned}
\mathbb{E}[p(\mathbf{S}|\mathbf{V})] \geq & -F\cdot N\cdot R\cdot \log \pi\\
&+ \sum_{j,fn} R_j\Big(-\log\big(\sum_{\eta}\sum_{\eta'}\mathbb{E}[v_{j,fn,\eta,\eta'}]\big)\Big)\\
&- \sum_{j,fn}\sum_{r}\mathbb{E}[|s_{jr,fn}|^2]\frac{1}{\sum_\eta\sum_{\eta'}\mathbb{E}\Big[\frac{1}{v_{j,fn,\eta,\eta'}}\Big]^{-1}}
\end{aligned}
$$

Due to the factorization given in (43), the expectations $\mathbb{E}[v_{j,fn,\eta,\eta'}]$ and $\mathbb{E}\Big[\frac{1}{v_{j,fn,\eta,\eta'}}\Big]$ are computed for each NTF component individually as given in (84, 85). Furthermore, the expectation $\mathbb{E}[|s_{jr,fn}|^2]$ is the second raw moment of source component $s_{jr,fn}$ and is computed using (94).

The expectation of the logarithm of the optimal approximating distribution of the source components $q(\mathbf{S})$ is given by

$$
\begin{aligned}
-\mathbb{E}[\log q(\mathbf{S})] &= -\sum_{fn}\mathbb{E}[\log \mathcal{N}(\mathbf{s}_{fn}|\boldsymbol{\mu}_{\mathbf{s},fn}, \mathbf{R}_{\mathbf{ss},fn})],\\
&= \sum_{fn}\log\{(\pi e)^R \det(\mathbf{R}_{\mathbf{ss},fn})\}.
\end{aligned}
\tag{104}
$$

About the mixing matrix, we remind you that the prior distribution of the mixing coefficients does not contribute to the lower bound, because it is flat. The expectation of the optimal approximating distribution of the mixing coefficients is given by

$$\begin{aligned}\mathbb{E}[\log p(\underline{\mathbf{A}})] - \mathbb{E}[\log q(\underline{\mathbf{A}})] &= -\sum_f \mathbb{E}[\ln \mathcal{N}(\underline{\mathbf{A}}_f | \boldsymbol{\mu}_{\underline{\mathbf{A}},f}, \mathbf{R}_{\underline{\mathbf{A}}\underline{\mathbf{A}},f})], \\ &= \sum_f \log\{(\pi e)^{2R} \det(\mathbf{R}_{\underline{\mathbf{A}}\underline{\mathbf{A}},f})\}. \end{aligned} \tag{105}$$

Finally, for one of the NTF parameters, *e.g.* for $w^{\text{ex}}_{j,c,fl}$, this pair is written as follows:

$$\begin{aligned}\mathbb{E}_q[\log p(w^{\text{ex}}_{j,c,fl})] - \mathbb{E}_q[\log q(w^{\text{ex}}_{j,c,fl})] =& \rho^{\text{ex}}_{w,j,c,fl} \mathbb{E}_q[w^{\text{ex}}_{j,c,fl}] + \tau^{\text{ex}}_{w,j,c,fl} \mathbb{E}_q[\frac{1}{w^{\text{ex}}_{j,c,fl}}] \\ &+ \log \mathcal{K}_{\gamma^{\text{ex}}_{w,j,c,fl}}(2\sqrt{\rho\tau}) + \ln 2 \\ &- \frac{\gamma^{\text{ex}}_{w,j,c,fl}}{2}(\ln \rho^{\text{ex}}_{w,j,c,fl} - \ln \tau^{\text{ex}}_{w,j,c,fl}). \end{aligned} \tag{106}$$

Note that for $\gamma^{\text{ex}}_{w,j,c,fl} = 0$, the third line of this equation disappears. The calculation of the contribution of the other NTF parameters to the lower bound is calculated similarly as given in (106).

2.4 Summary of the Algorithm

The update equations for the approximating distributions require the evaluation of some expectations. These expectations are to be evaluated with respect to the other distributions. Therefore an iterative update procedure similar to the Expectation-Maximization (E-M) algorithm is followed. In the variational E-step, these expectations are calculated. In the variational M-step, the approximating distributions are re-calculated with the new expectation values.

Overall, after proper initialization, each iteration of the algorithm consists of the following steps:

1. compute the statistics of the NTF parameters, the source STFT coefficients and the mixing parameters as in (83)–(87), (94), (95), and (99),
2. update the auxiliary variables in (53) and (58),
3. update the GIG distributions of the NTF parameters according to (63), (65), (67), (69), (71), (72), (74), (76), (77), (79), (81), (82),
4. update the complex-valued Gaussian distribution of the source STFT coefficients according to (91) and (92),
5. update the complex-valued Gaussian distribution of the mixing parameters according to (97) and (98),
6. compute the lower bound using (103)–(106) and terminate the algorithm when the bound increases less than a given threshold or after a certain number of iterations.

Similarly to the original ML algorithm, with two or more channels and for typical signal length, the computational cost is dominated by the updates in (83), (91), (92), (94), and (95), which must be performed for each time-frequency bin. The cost of these updates being similar to the that of the E-step of the original ML algorithm, the overall cost of one iteration of the two algorithms is on the same order.

References

[1] K. Adiloğlu and E. Vincent. Variational Bayesian inference for source separation and robust feature extraction. *IEEE Transactions on Audio Speech and Language Processing*, 2012. submitted.

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

[3] M. D. Hoffman, D. M. Blei, and P. R. Cook. Bayesian nonparametric matrix factorization for recorded music. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2010.

[4] B. Jorgensen. *Statistical Properties of the Generalized Inverse-Gaussian Distribution*. Springer, 1982.

[5] A. Ozerov, E. Vincent, and F. Bimbot. A general flexible framework for the handling of prior information in audio source separation. *IEEE Transactions on Audio, Speech and Language Processing*, 20(4):1118–1133, 2012.

RESEARCH CENTRE
RENNES – BRETAGNE ATLANTIQUE

Campus universitaire de Beaulieu
35042 Rennes Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-0803