VARIATIONAL BAYESIAN EM ALGORITHM FOR MODELING MIXTURES OF NON-STATIONARY SIGNALS IN THE TIME-FREQUENCY DOMAIN (HR-NMF)

Roland Badeau, Angélique Drémeau

Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI

ABSTRACT

We recently introduced the high-resolution nonnegative matrix factorization (HR-NMF) model for analyzing mixtures of non-stationary signals in the time-frequency domain, and highlighted its capability to both reach high spectral resolution and reconstruct high quality audio signals. In order to estimate the model parameters and the latent components, we proposed to resort to an expectation-maximization (EM) algorithm based on a Kalman filter/smoother. The approach proved to be appropriate for modeling audio signals in applications such as source separation and audio inpainting. However, its computational cost is high, dominated by the Kalman filter/smoother, and may be prohibitive when dealing with high-dimensional signals. In this paper, we consider two different alternatives, using the variational Bayesian EM algorithm and two mean-field approximations. We show that, while significantly reducing the complexity of the estimation, these novel approaches do not alter its quality.

***Index Terms*—** Nonnegative Matrix Factorization, High Resolution methods, Expectation-Maximization algorithm, Variational inference.

1. INTRODUCTION

Nonnegative matrix factorization (NMF) [1] is a powerful tool for decomposing mixtures of non-stationary signals in the time-frequency (TF) domain. However, unlike the high resolution (HR) methods [2] dedicated to mixtures of complex exponentials, its spectral resolution is limited by that of the underlying TF representation. Following previous works which aimed at providing a probabilistic framework for NMF [3–6], we introduced in [7, 8] a unified probabilistic model called HR-NMF, that permits to overcome this limit by taking both phases and local correlations in each frequency band into account. It can be used with both complex-valued and real-valued TF representations (like the short-time Fourier transform or the modified discrete cosine transform). Moreover, we showed that HR-NMF generalizes some very popular models: the Itakura-Saito NMF model (IS-NMF) [6], autoregressive (AR) processes, and the exponential sinusoidal model (ESM), commonly used in HR spectral analysis of time series [2]. In [7, 8], HR-NMF was estimated with the expectation-maximization (EM) algorithm, which involves time-demanding Kalman filtering and smoothing. In this paper, we introduce two faster algorithms based on variational inference, and compare the performance of the three algorithms.

This paper is organized as follows. In Section 2, we present the HR-NMF model, as introduced in [7]. We recall the basics of the variational Bayesian EM algorithm in Section 3, before particularizing it to the HR-NMF model in Section 4. Section 5 is devoted to experimental results, and conclusions are drawn in Section 6.

The following notation will be used throughout the paper:

- $\boldsymbol{M}^*$: conjugate of matrix (or vector) $\boldsymbol{M}$,
- $\boldsymbol{M}^\top$: transpose of matrix (or vector) $\boldsymbol{M}$,
- $\boldsymbol{M}^H$: conjugate transpose of matrix (or vector) $\boldsymbol{M}$,
- $[\boldsymbol{M};\boldsymbol{N}]$: vertical concatenation of $\boldsymbol{M}$ and $\boldsymbol{N}$,
- $\overset{\text{c}}{=}$: equality up to an additive constant,
- $h * m$: discrete convolution of times series h and m,
- $\mathcal{N}_{\mathbb{F}}(\boldsymbol{\mu}, \boldsymbol{R})$: real (if $\mathbb{F} = \mathbb{R}$) or circular complex (if $\mathbb{F} = \mathbb{C}$) multivariate normal distribution of mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{R}$.

2. HR-NMF TIME-FREQUENCY MIXTURE MODEL

The HR-NMF mixture model of TF data $x(f,t) \in \mathbb{F}$ (where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) is defined for all discrete frequencies $1 \leq f \leq F$ and times $1 \leq t \leq T$ as the sum of K latent components $c_k(f,t) \in \mathbb{F}$ plus a white noise $n(f,t) \sim \mathcal{N}_{\mathbb{F}}(0, \sigma^2)$:

$$x(f,t) = n(f,t) + \sum_{k=1}^{K} c_k(f,t) \tag{1}$$

where

- $c_k(f,t) = \sum_{p=1}^{P(k,f)} a(p,k,f)\, c_k(f,t-p) + b_k(f,t)$ is obtained by autoregressive filtering of a non-stationary signal $b_k(f,t) \in \mathbb{F}$ (where $a(p,k,f) \in \mathbb{F}$ and $P(k,f) \in \mathbb{N}$ is such that $a(P(k,f),k,f) \neq 0$),
- $b_k(f,t) \sim \mathcal{N}_{\mathbb{F}}(0, v_k(f,t))$ where $v_k(f,t)$ is defined as

$$v_k(f,t) = w(k,f)\, h(k,t), \tag{2}$$

  with $w(k,f) \geq 0$ and $h(k,t) \geq 0$,
- processes n and $b_1 \dots b_K$ are mutually independent.

Moreover, $\forall (k,f) \in \{1 \dots K\} \times \{1 \dots F\}$, the random vectors $\boldsymbol{c}_k(f,0) = [c_k(f,0); \dots; c_k(f,-P(k,f)+1)]$ are assumed to be independent and distributed according to the prior distribution $\boldsymbol{c}_k(f,0) \sim \mathcal{N}_{\mathbb{F}}(\boldsymbol{\mu}_k(f), \boldsymbol{Q}_k(f)^{-1})$, where the mean $\boldsymbol{\mu}_k(f)$ and the precision matrix $\boldsymbol{Q}_k(f)$ are fixed parameters[1]. Lastly, we assume that $\forall f \in \{1 \dots F\}, \forall t \leq 0, x(f,t)$ is unobserved.

Let c denote the set $\{c_k(f,t)\}_{(k,f,t)}$, x denote the set $\{x(f,t)\}_{(f,t)}$ and θ the set of model parameters $\sigma^2, \{a(p,k,f)\}_{(p,k,f)}, \{w(k,f)\}_{(k,f)}$

This work is supported by the French National Research Agency (ANR) as a part of the DReaM project (ANR-09-CORD-006-03) and partly supported by the Quaero Program, funded by OSEO.

[1] In practice we choose $\boldsymbol{\mu}_k(f) = [0; \dots; 0]^\top$ and $\boldsymbol{Q}_k(f)^{-1} = \xi \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix and ξ is small relative to 1, in order to both enforce the causality of the latent components and avoid singular matrices.

and $\{h(k,t)\}_{(k,t)}$. Considering model (1), we focus on the maximum a posteriori (MAP) estimation of the latent components

$$c^\star = \underset{c}{\operatorname{argmax}}\, p(c|x;\theta^\star), \tag{3}$$

where the model parameters are estimated according to a maximum likelihood (ML) criterion

$$\theta^\star = \underset{\theta}{\operatorname{argmax}}\, p(x;\theta). \tag{4}$$

The solution of (3)-(4) can be found by means of an EM algorithm. We proposed in [7, 8] an efficient implementation, using a Kalman filter/smoother in the E-step. However, the computational cost remains high, dominated by the complexity of the Kalman filter/smoother, and may be prohibitive when dealing with large dimensions. We propose here an alternative, based on the *variational Bayesian EM* (VB-EM) algorithm, which uses a mean-field approximation of the posterior $p(c|x;\theta^\star)$ to reach a good compromise between quality and complexity of the MAP estimation (3).

3. VARIATIONAL BAYESIAN EM ALGORITHM

Variational inference [9, 10] is now a classical approach for estimating a probabilistic model involving both observed variables x and latent variables c, parametrized by θ. Let $\mathcal{F}$ be a set of probability density functions (PDF) over the latent variables c. For any PDF $q \in \mathcal{F}$ and any function $f(c)$, we note $\langle f \rangle_q = \int f(c)q(c)dc$. Then for any PDF $q \in \mathcal{F}$ and any parameter θ, the log-likelihood $L(\theta) = \ln(p(x;\theta))$ can be decomposed as

$$L(\theta) = D_{\text{KL}}(q||p(c|x;\theta)) + \mathcal{L}(q;\theta) \tag{5}$$

$$\text{where} \qquad D_{\text{KL}}(q||p(c|x;\theta)) = \left\langle \ln\left(\frac{q(c)}{p(c|x;\theta)}\right)\right\rangle_q \tag{6}$$

is the *Kullback-Leibler divergence* between q and $p(c|x;\theta)$, and

$$\mathcal{L}(q;\theta) = \left\langle \ln\left(\frac{p(c,x;\theta)}{q(c)}\right)\right\rangle_q \tag{7}$$

is called the *variational free energy*. Moreover, $\mathcal{L}(q;\theta)$ can be further decomposed as $\mathcal{L}(q;\theta) = E(q;\theta) + H(q)$, where

$$E(q;\theta) = \left\langle \ln\left(p(c,x;\theta)\right)\right\rangle_q, \tag{8}$$

and $H(q) = -\left\langle \ln\left(q(c)\right)\right\rangle_q$ is the entropy of distribution q. Since $D_{\text{KL}}(q||p(c|x;\theta)) \geq 0$, $\mathcal{L}(q;\theta)$ is a lower bound of $L(\theta)$. The variational Bayesian EM algorithm is a recursive algorithm for estimating θ. It consists of the two following steps at each iteration i:

- E-step (update q):

$$q^\star = \underset{q\in\mathcal{F}}{\operatorname{argmin}}\, D_{\text{KL}}(q||p(c|x;\theta_{i-1})) = \underset{q\in\mathcal{F}}{\operatorname{argmax}}\, \mathcal{L}(q;\theta_{i-1}) \tag{9}$$

- M-step (update θ):

$$\theta_i = \underset{\theta}{\operatorname{argmax}}\, \mathcal{L}(q^\star;\theta) = \underset{\theta}{\operatorname{argmax}}\, E(q^\star;\theta). \tag{10}$$

$\mathcal{F}$ defines a set of constraints leading to a particular approximation of the posterior distribution $p(c|x;\theta_{i-1})$. We note that:

- In the standard EM algorithm, q is not constrained, thus $q^\star = p(c|x;\theta_{i-1})$ and $D_{\text{KL}}(q^\star||p(c|x;\theta_{i-1})) = 0$. Therefore $L(\theta_i) \geq \mathcal{L}(q^\star;\theta_i) \geq \mathcal{L}(q^\star;\theta_{i-1}) = L(\theta_{i-1})$, which proves that the log-likelihood is non-decreasing.
- In the general case, $L(\theta)$ is no longer guaranteed to be non-decreasing, but its lower bound $\mathcal{L}(q;\theta)$ is still non-decreasing.

4. VARIATIONAL BAYESIAN EM FOR HR-NMF

Considering the HR-NMF model defined in (1), $\forall (k,f) \in \{1\dots K\}\times\{1\dots F\}$, let c_{kf} denote the set $\{c_k(f,t)\}_{t\in\{-P(k,f)+1\dots T\}}$. Moreover, let $\alpha = 1$ if $\mathbb{F} = \mathbb{C}$, and $\alpha = 2$ if $\mathbb{F} = \mathbb{R}$. Then

$$\begin{aligned}
&\alpha \ln(p(c,x)) = \alpha \sum_{k=1}^{K}\sum_{f=1}^{F} \ln(p(c_{kf}))\\
&+\alpha \sum_{f=1}^{F}\sum_{t=1}^{T} \delta(f,t)\ln(p(x(f,t)|c_1(f,t)\dots c_K(f,t)))\\
&= -\left(KFT + \sum_{k=1}^{K}\sum_{f=1}^{F} P(k,f)\right)\ln(\alpha\pi)\\
&- \sum_{k=1}^{K}\sum_{f=1}^{F} (\boldsymbol{c}_k(f,0) - \boldsymbol{\mu}_k(f))^H \boldsymbol{Q}_k(f)(\boldsymbol{c}_k(f,0) - \boldsymbol{\mu}_k(f))\\
&+ \sum_{k=1}^{K}\sum_{f=1}^{F}\left(\ln(\det(\boldsymbol{Q}_k(f))) + \sum_{t=1}^{T}\ln(\rho_k(f,t))\right)\\
&- \sum_{k=1}^{K}\sum_{f=1}^{F}\sum_{t=1}^{T} \rho_k(f,t)\left|c_k(f,t) - \sum_{p=1}^{P(k,f)} a(p,k,f)c_k(f,t-p)\right|^2\\
&- \sum_{f=1}^{F}\sum_{t=1}^{T}\delta(f,t)\left(\ln(\alpha\pi\sigma^2) + \tfrac{1}{\sigma^2}\left|x(f,t) - \sum_{k=1}^{K} c_k(f,t)\right|^2\right)
\end{aligned} \tag{11}$$

where

- $\delta(f,t) = 1$ if $x(f,t)$ is observed, and $\delta(f,t) = 0$ else (in particular $\delta(f,t) = 0\ \forall t < 1$ and $\forall t > T$),
- $\rho_k(f,t) = \frac{1}{v_k(f,t)}$ if $t \in \{1\dots T\}$, and $\rho_k(f,t) = 0$ else.

In the following subsections, we will first recall the EM-based algorithm presented in [7, 8] as a particular case of the variational procedure (9)-(10) (Sections 4.1 and 4.2) and then propose two different alternatives to this costly approach, based on two mean-field approximations, *i.e.* two different definitions of $\mathcal{F}$ (Sections 4.3 and 4.4). These three algorithms only differ in the E-step, but they share the same implementation of the M-step.

4.1. M-step

The M-step defined in equation (10) consists in maximizing $E(q^\star;\theta)$ w.r.t. the model parameters θ. First, equations (8) and (11) yield

$$\begin{aligned}
&\alpha E(q^\star;\theta) \overset{\text{c}}{=} -\sum_{f=1}^{F}\sum_{t=1}^{T}\delta(f,t)\ln(\alpha\pi\sigma^2) + e(f,t)/\sigma^2\\
&-\sum_{k=1}^{K}\sum_{f=1}^{F}\sum_{t=1}^{T}\ln(w(k,f)h(k,t)) + \frac{\boldsymbol{a}(k,f)^H \boldsymbol{S}(k,f,t)\boldsymbol{a}(k,f)}{w(k,f)h(k,t)},
\end{aligned} \tag{12}$$

where

- $e(f,t) = \delta(f,t)\left\langle\left|x(f,t) - \sum_{k=1}^{K} c_k(f,t)\right|^2\right\rangle_{q^\star}$,
- $\boldsymbol{S}(k,f,t) = \langle \overline{\boldsymbol{c}}_k(f,t)^*\, \overline{\boldsymbol{c}}_k(f,t)^\top\rangle_{q^\star}$,
- $\overline{\boldsymbol{c}}_k(f,t) = [c_k(f,t);\dots;c_k(f,t-P(k,f))]$,
- $\boldsymbol{a}(k,f) = [1;-a(1,k,f);\dots;-a(P(k,f),k,f)]$.

We note that $e(f,t)$ and $\boldsymbol{S}(k,f,t)$ can be computed as

- $e(f,t) = \delta(f,t)\left(\left|x(f,t) - \sum_{k=1}^{K} m_k(f,t)\right|^2 + \sum_{k=1}^{K}\Gamma_k(f,t)\right)$,
- $\boldsymbol{S}(k,f,t) = \overline{\boldsymbol{R}}_k(f,t)^* + \overline{\boldsymbol{m}}_k(f,t)^*\,\overline{\boldsymbol{m}}_k(f,t)^\top$,

where we have defined:

- $m_k(f,t) = \langle c_k(f,t) \rangle_{q^\star}$,
- $\Gamma_k(f,t) = \langle |c_k(f,t) - m_k(f,t)|^2 \rangle_{q^\star}$,
- $\overline{\boldsymbol{m}}_k(f,t) = \langle \overline{\boldsymbol{c}}_k(f,t) \rangle_{q^\star}$,
- $\overline{\boldsymbol{R}}_k(f,t) = \langle (\overline{\boldsymbol{c}}_k(f,t) - \overline{\boldsymbol{m}}_k(f,t))\,(\overline{\boldsymbol{c}}_k(f,t) - \overline{\boldsymbol{m}}_k(f,t))^H \rangle_{q^\star}$.

The maximization of $E(q^\star;\theta)$ in equation (12), w.r.t. $\sigma^2, a(p,k,f)$, $w(k,f)$, and $h(k,t)$, can then be performed as in the M-step presented in [8], using the current estimations of $\overline{\boldsymbol{m}}_k(f,t)$ and $\overline{\boldsymbol{R}}_k(f,t)$ derived from the E-steps as presented in the next sections.

4.2. E-step in the exact EM algorithm

As mentioned in Section 3, in the exact EM algorithm q is not constrained, thus the solution of (9) is given by $q^\star = p(c|x;\theta)$, and the variational free energy $\mathcal{L}(q^\star,\theta_i)$ is equal to the log-likelihood $L(\theta_i)$. In [7,8], we showed that the posterior distribution $p(c|x;\theta)$ is Gaussian, and that its first and second order moments, as well as the value of $L(\theta_i)$, can be computed by means of Kalman filtering and smoothing. The resulting E-step can symbolically be written as:

for $1 \leq f \leq F$ **do**
$\{\overline{\boldsymbol{m}}_k(f,t), \overline{\boldsymbol{R}}_k(f,t)\}_{\substack{1\leq k\leq K\\ 1\leq t\leq T}} = \textbf{Kalman}\left(\{x(f,t)\}_{1\leq t\leq T}\right)$
end for

Its computational complexity was shown to be $O(FTK^3(1+P)^3)$, where $P = \max\limits_{k,f} P(k,f)$.

4.3. E-step with structured mean field approximation

If $K > 1$, we assume that $\mathcal{F}$, introduced in Section 3, is the set of PDFs which can be factorized in the form

$$q(c) = \prod_{k=1}^{K}\prod_{f=1}^{F} q_{kf}(c_{kf}). \tag{13}$$

Using this particular factorization for $q(c)$, the solution of (9) satisfies (see [9]): $\forall (k,f) \in \{1\dots K\}\times\{1\dots F\}$,

$$\ln(q_{kf}(c_{kf})) \stackrel{c}{=} \langle \ln(p(c,x)) \rangle_{\left(\prod_{(l,g)\neq(k,f)} q_{lg}\right)}. \tag{14}$$

Then, reformulating equation (11) and using equation (14), we get

$$\alpha \ln(p(c,x)) \stackrel{c}{=} \alpha \ln(p(c_{kf})) - \sum_{t=1}^{T} \frac{\delta(f,t)}{\sigma^2} \left|c_k(f,t) - \hat{c}_k(f,t)\right|^2, \tag{15}$$

with $\hat{c}_k(f,t) = x(f,t) - \sum\limits_{l\neq k} m_l(f,t)$. We observe that q_{kf} is the posterior distribution of a HR-NMF model of order $K=1$, where the posterior means of all components other than k have been subtracted to the observed data $x(f,t)$. Hence q_{kf} is Gaussian, and its first and second moments can be computed by applying the Kalman filter/smoother presented in [7,8] to $\hat{c}_k(f,t)$ instead of $x(f,t)$. The resulting E-step can symbolically be written as:

for $1 \leq f \leq F$ **do**
for $1 \leq k \leq K$ **do**
$\forall 1 \leq t \leq T, \hat{c}_k(f,t) = x(f,t) - \sum\limits_{l\neq k} m_l(f,t)$
$\{\overline{\boldsymbol{m}}_k(f,t), \overline{\boldsymbol{R}}_k(f,t)\}_{1\leq t\leq T} = \textbf{Kalman}(\{\hat{c}_k(f,t)\}_{1\leq t\leq T})$
end for
end for

The complexity of this procedure is $O(FTK(1+P)^3)$ instead of $O(FTK^3(1+P)^3)$ for the "classical" E-step.

In order to evaluate this algorithm, we are also interested in computing the variational free energy $\mathcal{L}$. After some straightforward calculations, we note that the entropy $H(q_{kf})$ satisfies

$$\begin{aligned}\alpha H(q_{kf}) &= (T + P(k,f))\,(\ln(\alpha\pi) + 1)\\ &+ \sum_{t=1}^{T} \ln(\det(\overline{\boldsymbol{R}}_k(f,t))) - \sum_{t=1}^{T-1} \ln(\det(\boldsymbol{R}_k(f,t))),\end{aligned} \tag{16}$$

where $\boldsymbol{R}_k(f,t)$ is the $P(k,f)\times P(k,f)$ top-left submatrix of $\overline{\boldsymbol{R}}_k(f,t)$. Thus equations (7), (11) and (16) yield

$$\begin{aligned}\alpha\mathcal{L}(q;\theta) &= KFT + \sum_{k=1}^{K}\sum_{f=1}^{F} P(k,f) - \operatorname{trace}(\boldsymbol{Q}_k(f)\boldsymbol{R}_k(f,0))\\ &- \sum_{k=1}^{K}\sum_{f=1}^{F} (\boldsymbol{m}_k(f,0) - \boldsymbol{\mu}_k(f))^H \boldsymbol{Q}_k(f)(\boldsymbol{m}_k(f,0) - \boldsymbol{\mu}_k(f))\\ &+ \sum_{k=1}^{K}\sum_{f=1}^{F} \ln(\det(\boldsymbol{Q}_k(f))) + \sum_{t=1}^{T} \ln(\rho_k(f,t))\\ &- \sum_{k=1}^{K}\sum_{f=1}^{F}\sum_{t=1}^{T} \rho_k(f,t)\,\boldsymbol{a}(k,f)^H \boldsymbol{S}(k,f,t)\boldsymbol{a}(k,f)\\ &- \sum_{f=1}^{F}\sum_{t=1}^{T} \delta(f,t)\ln(\alpha\pi\sigma^2) + e(f,t)/\sigma^2\\ &+ \sum_{k=1}^{K}\sum_{f=1}^{F}\left(\sum_{t=1}^{T} \ln(\det(\overline{\boldsymbol{R}}_k(f,t))) - \sum_{t=1}^{T-1} \ln(\det(\boldsymbol{R}_k(f,t)))\right)\end{aligned} \tag{17}$$

where $\boldsymbol{m}_k(f,0)$ is the $P(k,f)\times 1$ top subvector of $\overline{\boldsymbol{m}}_k(f,0)$.

4.4. E-step with mean field approximation

If $P > 0$, we further assume that $\mathcal{F}$ is the set of PDFs which can be factorized in the form

$$q(c) = \prod_{k=1}^{K}\prod_{f=1}^{F}\prod_{t=-(P(k,f)-1)}^{T} q_{kft}(c_k(f,t)). \tag{18}$$

With this particular factorization of $q(c)$, the solution of (9) satisfies (see [9]): $\forall (k,f,t)\in\{1\dots K\}\times\{1\dots F\}\times\{-P(k,f)+1\dots T\}$,

$$\ln(q_{kft}(c_k(f,t))) \stackrel{c}{=} \langle \ln(p(c,x)) \rangle_{\left(\prod_{(l,g,u)\neq(k,f,t)} q_{lgu}\right)}. \tag{19}$$

Let us define the filter of impulse response h_{kf}, such that $h_{kf}(0) = 1$, $h_{kf}(p) = -a(p,k,f)\ \forall p \in \{1\dots P(k,f)\}$, and $h_{kf}(p) = 0$ everywhere else, and the filter $\tilde{h}_{kf}(p) = h_{kf}(-p)^*$. After some straightforward calculations, equations (11) and (19) yield $\forall (k,f,t) \in \{1\dots K\}\times\{1\dots F\}\times\{-P(k,f)+1\dots T\}$, $q_{kft}(c_k(f,t)) \sim \mathcal{N}_{\mathbb{F}}\left(m_k(f,t), \Gamma_k(f,t)\right)$, where[2]

$$\Gamma_k(f,t) = \left(\frac{\delta(f,t)}{\sigma^2} + q_k(f,t) + |\tilde{h}_{kf}|^2 * \rho_k(f,t)\right)^{-1}$$

(with $q_k(f,t) = Q_k(f)_{(1-t,1-t)}$ if $-P(k,f)+1 \leq t \leq 0$ and $q_k(f,t) = 0$ else), and

$$\begin{aligned}m_k(f,t) = m_k(f,t) + \Gamma_k(f,t)\Big(&- \boldsymbol{q}_k(f,t)^H(\boldsymbol{m}_k(f,0) - \boldsymbol{\mu}_k(f))\\ &+ \tfrac{\delta(f,t)}{\sigma^2}(x(f,t) - \sum_{l=1}^{K} m_l(f,t)) - \tilde{h}_{kf} * (\rho_k(f,t)(h_{kf} * m_k(f,t)))\Big)\end{aligned}$$

[2] $|\tilde{h}_{kf}|^2$ denotes the filter whose coefficients are the square magnitude of the corresponding coefficients of $\tilde{h}_{kf}$.

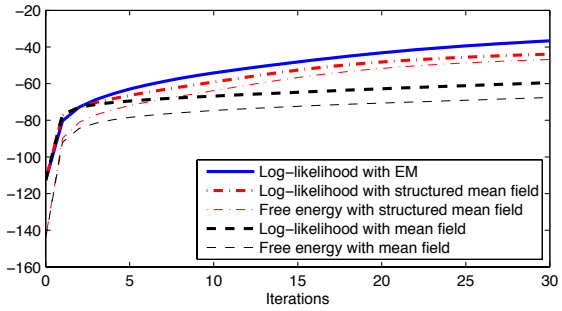

Fig. 1. Maximization of the log-likelihood and the variational free energy over the iterations.

(where $\boldsymbol{q}_k(f,t)$ is the $(1-t)^{th}$ column of $\boldsymbol{Q}_k(f)$ if $-P(k,f)+1 \leq t \leq 0$ and $\boldsymbol{q}_k(f,t) = [0;\ldots;0]$ else)[3]. The computational complexity of the E-step is thus further reduced from $O(KFT(1+P)^3)$ to $O(KFT(1+P))$, which is linear w.r.t. all dimensions.

Finally, note that the variational free energy $\mathcal{L}(q;\theta)$ can be calculated as in equation (17), where $\overline{\boldsymbol{R}}_k(f,t)$ becomes a diagonal matrix of diagonal coefficients $\Gamma_k(f,t)\ldots\Gamma_k(f,t-P(k,f))$.

5. SIMULATION RESULTS

The VB-EM algorithm aims to maximize the free energy. As we mentioned in Section 3, the log-likelihood is thus no longer guaranteed to increase, while remaining an indicator of the estimation quality. It can then be interesting to evaluate the influence of the approximations (13) and (18) on the maximization of the log-likelihood. To this end, we consider a fully observed TF data $x(f,t)$ generated according to model (1) with $T=20$, $F=3$, $P(k,f)=3 \ \ \forall(k,f)$ and $K=2$ (and random parameters θ), and compare the performance of the three algorithms described respectively in Subsections 4.2, 4.3 and 4.4 with regard to the maximization of the log-likelihood. Figure 1 presents the value of the log-likelihood at each iteration of the three algorithms. Interestingly, we can observe that although focusing on the maximization of the free energy, the VB-EM algorithm permits here to increase the log-likelihood, whatever the considered approximation (mean-field or structured mean-field). In addition, as intuitively expected, the most constrained factorization (18) leads to a lesser increase of the log-likelihood. In practice however, this expected quality loss is not tangible. As an example of the good behavior of the VB-EM approach, we focus here on a simple case of source separation, where the observation is the whole STFT $x(f,t)$ (of dimensions $F=400$ and $T=44$) of a 1.05 s-long piano sound sampled at 11025 Hz, containing three notes, C3, C4 and C5, starting respectively at 0 ms, 260 ms and 525 ms, and lasting until the end of the sound. Within this scenario, we aim at separating $K=3$ components $c_k(f,t)$ of order $P(k,f)=2$ in the frequency band f which corresponds to the first harmonic of C5, the second harmonic of C4 and the fourth harmonic of C3 (around 540 Hz). These three sinusoidal components (whose real parts are represented as red solid lines in Figure 2) have very close frequencies, making them hardly separable. We compare then three different approaches, namely, the HR-NMF model estimated by means of the EM algorithm, the HR-NMF model estimated by means of the VB-EM algorithm using the mean-field approximation (18) and the IS-NMF model [6].

[3]In this equation, although the term $m_k(t)$ appears several times in the right-hand side, it can be easily verified that its contributions add up to zero.

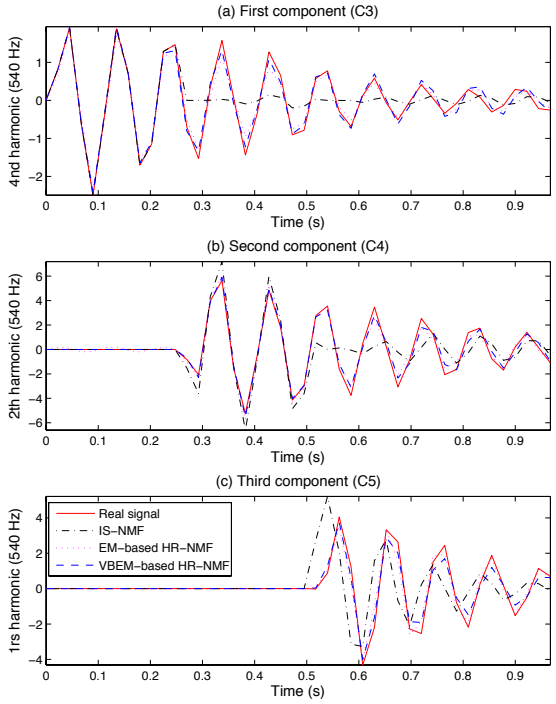

Fig. 2. Separation of three sinusoidal components.

Two important observations can be made about Figure 2. As previously noticed in [7,8], IS-NMF (in black dash-dotted lines), which involves Wiener filtering, is not able to properly separate the components when they overlap. As a comparison, the components estimated by HR-NMF (blue dashed lines and magenda dotted lines) better fit the ground truth. We see on this example that the EM-based and VB-EM-based approaches lead to very similar results: the separated components are often merged. More precisely, we measured an averaged mean squared error of 0.0161 for IS-NMF, 0.0016 for the VBEM-based HR-NMF and 0.0006 for the EM-based HR-NMF on the whole set of frequencies and components reconstructed within this experiment. The slight quality loss due to the mean-field approximation is largely compensated by a significant computation time saving: with a 2.20GHz CPU processor and 8Go RAM, the CPU time required to run the E-step in the exact EM approach with a Matlab implementation is 19.5s, while 1.9s is enough for the E-step with mean-field approximation.

6. CONCLUSIONS

This paper introduces two novel methods as alternatives to estimate the HR-NMF model introduced in [7,8]. These methods are based on the variational Bayesian EM algorithm and two different mean-field approximations. Their low complexities allow using the HR-NMF model in high-dimensional problems without altering the good quality of the estimation. We illustrated these good properties with a simple example of source separation. In future work, we will investigate other kinds of structured and unstructured mean field approximations, as well as a fully Bayesian approach involving uninformative or informative priors for the various model parameters. We will also apply variational inference to the future extensions of the HR-NMF model (*e.g.* involving convolutive and multichannel mixtures).

7. REFERENCES

[1] Daniel D. Lee and H. Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, Oct. 1999.

[2] Roland Badeau, Bertrand David, and Gaël Richard, "High resolution spectral analysis of mixtures of complex exponentials modulated by polynomials," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1341–1350, Apr. 2006.

[3] Mikkel N. Schmidt and Hans Laurberg, "Non-negative matrix factorization with Gaussian process priors," *Computational Intelligence and Neuroscience*, vol. 2008, pp. 1–10, 2008, Article ID 361705.

[4] Paris Smaragdis, *Blind Speech Separation*, chapter Probabilistic decompositions of spectra for sound separation, pp. 365–386, Springer, 2007.

[5] T. Virtanen, A.T. Cemgil, and S. Godsill, "Bayesian extensions to non-negative matrix factorisation for audio signal modelling," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, Nevada, USA, Apr. 2008, pp. 1825–1828.

[6] Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, Mar. 2009.

[7] Roland Badeau, "Gaussian modeling of mixtures of non-stationary signals in the time-frequency domain (HR-NMF)," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA, Oct. 2011, pp. 253–256.

[8] Roland Badeau, "High resolution NMF for modeling mixtures of non-stationary signals in the time-frequency domain," Tech. Rep. 2012D004, Télécom ParisTech, Paris, France, July 2012.

[9] M. Beal, *Variational Algorithms for Approximate Bayesian Inference*, Ph.D. thesis, Univ. College of London, London, U.K., May 2003.

[10] Martin J. Wainwright and Michael I. Jordan, *Graphical Models, Exponential Families, and Variational Inference*, vol. 1 of *Foundations and Trends® in Machine Learning*, Now Publishers, 2008.