Circuits Syst Signal Process (2014) 33:1841–1857
DOI 10.1007/s00034-013-9719-8

PARAFAC-Based Blind Identification of Underdetermined Mixtures Using Gaussian Mixture Model

Fanglin Gu · Hang Zhang · Wenwu Wang · Desheng Zhu

Received: 4 March 2013 / Revised: 20 November 2013 / Published online: 7 December 2013

Abstract This paper presents a novel algorithm, named GMM-PARAFAC, for blind identification of underdetermined instantaneous linear mixtures. The GMM-PARAFAC algorithm uses Gaussian mixture model (GMM) to model non-Gaussianity of the independent sources. We show that the distribution of the observations can also be modeled by a GMM, and derive a maximum-likelihood function with regard to the mixing matrix by estimating the GMM parameters of the observations via the expectation-maximization algorithm. In order to reduce the computation complexity, the mixing matrix is estimated by maximizing a tight upper bound of the likelihood instead of the log-likelihood itself. The maximum of the tight upper bound is obtained by decomposition of a three-way tensor which is obtained by stacking the covariance matrices of the GMM of the observations. Simulation results validate the superiority of the GMM-PARAFAC algorithm.

Keywords Blind identification · Gaussian mixture model · Expectation-maximization · Underdetermined mixture · Parallel factor decomposition

F. Gu (✉) · H. Zhang · D. Zhu
Institute of Communication Engineering, PLA University of Science & Technology, Nanjing 210007, P.R. China
e-mail: gu.fanglin@gmail.com

H. Zhang
e-mail: hangzh_2002@163.com

D. Zhu
e-mail: dszhu@vip.sina.com

W. Wang
Department of Electronic Engineering, University of Surrey, Guildford GU2 7XK, UK
e-mail: w.wang@surrey.ac.uk

1 Introduction

Blind identification (BI), which aims to estimate the mixing system from observed signals without prior information about the sources and the mixing system, has attracted a lot of research interests due to its wide range of applications in signal processing, including blind source separation (BSS), and direction of arrival (DOA) estimation. There exist a large number of publications dealing with determined/overdetermined mixtures [1, 3, 9, 19, 25], which are unlikely to be reviewed exhaustively here due to the lack of space. However, in practical situations, for example, in telecommunications, it becomes increasingly likely for more sources to be received by fewer sensors due to the increase of the reception bandwidth. Therefore, BI methods that are able to process underdetermined mixtures, where the number of sources P is greater than that of sensors Q, are desired.

To this aim, many methods for BI of underdetermined mixtures turn to the use of various decomposition methods based on different data structures such as the second-order (SO), the fourth-order (FO), or the higher-order (HO) cumulants of the data [6, 10, 13, 15, 23]. The main idea of these algorithms is to construct a tensor based on the cumulants of the observations and then to estimate the mixing matrix by the decomposition of such a tensor. It has been shown in these methods that underdetermined mixtures can be identified up to a trivial matrix, i.e., a diagonal matrix and a permutation matrix, provided that all sources are non-Gaussian. However, these methods use restrictive assumptions on the source distribution, which makes them inapplicable in some scenarios. For example, the second-order blind identification of underdetermined mixtures (SOBIUM) algorithm [13] cannot be applied when the sources follow identical and independent distributions (i.i.d.) due to the fact that the spatial covariance of the i.i.d. sources equals zero.

In [16] it is shown that many probability density functions (PDFs) can be closely approximated, in terms of Kullback–Leibler (KL) divergence [12], by a finite-order Gaussian mixture model (GMM). Therefore, in this paper, GMM is used for modeling the PDF of each source and for exploiting its non-Gaussianity. Several researchers have utilized the GMM in solving the BSS problem. For example, Moulines et al. [17] developed an approximate maximum-likelihood (ML) method for blind separation and deconvolution of noisy linear mixtures, where the PDF of each source was modeled by a GMM. According to this approach, an expectation-maximization (EM) algorithm [5], which jointly estimates the mixing matrix, the source distribution parameters, and the noise covariance matrix, was developed. Some related works have been presented in [8, 22, 26]. For example, instead of using the EM method, a variational Bayesian method is used to obtain the estimates of the mixing matrix and model parameters in [26], and the conjugate prior densities are introduced in [8] to incorporate the prior information to improve the separation performance. However, there are still several limitations associated with these methods. First, the computational complexity of these methods is very high even for identifying the mixing matrix only (i.e., without considering the recovery of the sources). Second, an accurate initialization is vital to ensure the algorithms to converge to the desired minima/maxima, whereas it is usually difficult to obtain a good initialization, especially for the underdetermined case.

On the other hand, an alternative strategy, in which the GMM is fitted to the observed data, rather than the sources, is developed in [20, 24]. In this method, the GMM parameters are estimated via the EM algorithm and the separation matrix is estimated by applying a joint diagonalization (JD) technique to the covariance matrices of the GMM of the observations. An important feature of this method is that it is easy to initialize the parameters when implementing the EM algorithm since the EM algorithm is applied to estimate the GMM parameters of the observations rather than those of the sources. However, only the determined/overdetermined mixtures are considered. In this paper, we generalize the GMM-JD method to the underdetermined mixtures. The main contributions of our work are two-fold.

- We derive an ML function of the mixing matrix by estimating the GMM parameters of the observed data via the EM algorithm, and show that a tight upper bound of the log-likelihood, instead of the log-likelihood itself, can be utilized to estimate the mixing matrix, which alleviates the cumbersome computational loads for the direct maximization of the log-likelihood function of the mixing matrix.
- We propose a parallel factor (PARAFAC)-based approach to estimate the mixing matrix by decomposing the tensor which is formed from the covariance matrices of the GMM of the observations in a way similar to our previous work [7], except in which the tensor is formed from the Hessian matrices of the second generalized generating function of the observations. The proposed algorithm is shown to work well in underdetermined cases, thanks to the powerful uniqueness properties of the PARAFAC decomposition.

The remainder of the paper is organized as follows. In Sect. 2, the mathematical models for the PDFs of the sources and observation signals are introduced. In Sect. 3, an ML objective function with regard to the mixing matrix is derived by estimating the parameters of the GMM of the observations via the EM algorithm. Furthermore, a tight upper bound is proposed to approximate the ML objective function to reduce the computation complexity. In Sect. 4, the mixing matrix is estimated by applying the PARAFAC decomposition to the tensor formed from the estimated covariance matrices of the GMM of the observations, and the application of the proposed method to the underdetermined case is also discussed. In Sect. 5, the performance of the proposed method is evaluated and compared with other existing methods. Finally, Sect. 6 summarizes the main findings of this paper.

Notation Scalars are denoted by lower-case italic letters $(a, b, \ldots)$, vectors by lower-case boldface letters $(\mathbf{a}, \mathbf{b}, \ldots)$, matrices by boldface capitals $(\mathbf{A}, \mathbf{B}, \ldots)$ and tensors by calligraphic letters $(\mathcal{A}, \mathcal{B}, \ldots)$. The entry with row index i and column index j in a matrix $\mathbf{A}$, i.e., $(\mathbf{A})_{ij}$, is symbolized by a_{ij}. Likewise, we have $(\mathcal{A})_{ijk} = a_{ijk}$. The columns of $\mathbf{A}$ are denoted by $\mathbf{a}_1, \mathbf{a}_2, \ldots$. Conversely, the matrix with columns $\mathbf{a}_1, \mathbf{a}_2, \ldots$ is denoted by $\mathbf{A}$. The superscript $(\cdot)^{\mathrm{T}}$ denotes the transpose operator.

2 The Distribution Models for the Sources and Observations

Considering the following instantaneous linear mixture model:

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t + \mathbf{w}_t, \quad t = 1, 2, \ldots, T \tag{1}$$

The random vector $\mathbf{s}_t = [s_{1,t}, \ldots, s_{P,t}]^{\mathrm{T}}$, representing P statistically independent sources at time instance t, is mixed by a fixed unknown mixing matrix $\mathbf{A}$. The observation vector $\mathbf{x}_t = [x_{1,t}, \ldots, x_{Q,t}]^{\mathrm{T}}$ is obtained from an array of Q sensors, and contaminated by the noise vector $\mathbf{w}_t = [w_{1,t}, \ldots, w_{Q,t}]^{\mathrm{T}}$. The noises are assumed to be Gaussian white with zero-mean and unknown variance matrix $\mathbf{R}_{\mathbf{w}} = \mathrm{Diag}(\sigma_1^2, \ldots, \sigma_Q^2)$, and independent of $\mathbf{s}_t = [s_{1,t}, \ldots, s_{P,t}]^{\mathrm{T}}$. Under the assumption of stationary and non-Gaussian source signals, the sources and observations distribution models are discussed next.

2.1 Sources Distribution Model

The PDF of the ith source signal at time instance t is modeled as in [8] by a GMM in the following manner:

$$f_s(s_{i,t}|\boldsymbol{\theta}_i) = \sum_{l_i=1}^{N_i} \alpha_{i,l_i} \mathcal{N}\big(s_{i,t}; \mu_{i,l_i}, \sigma_{i,l_i}^2\big), \quad i = 1, \ldots, P \tag{2}$$

where $\mathcal{N}(\cdot;\cdot,\cdot)$ represents a Gaussian density function and N_i denotes the number of Gaussian components. The mixing weights are represented by $\{\alpha_{i,l_i}\}_{l_i=1}^{N_i}$, such that $\sum_{l_i=1}^{N_i} \alpha_{i,l_i} = 1$. The means and variances of the Gaussians are represented by $\{\mu_{i,l_i}\}_{l_i=1}^{N_i}$ and $\{\sigma_{i,l_i}^2\}_{l_i=1}^{N_i}$, respectively. Owing to the independence between the source signals, their joint PDF is a multivariate GMM with diagonal covariance matrices, which can be formulated as follows [8]:

$$\begin{aligned}
f_{\mathbf{s}}(\mathbf{s}_t|\Theta_{\mathbf{s}}) &= \prod_{i=1}^{P} f_s(s_{i,t}|\boldsymbol{\theta}_i) \\
&= \sum_{l_1}^{N_1} \alpha_{1,l_1} \mathcal{N}\big(s_{1,t}|\mu_{1,l_1}, \sigma_{1,l_1}^2\big) \sum_{l_2}^{N_2} \alpha_{2,l_2} \mathcal{N}\big(s_{2,t}|\mu_{2,l_2}, \sigma_{2,l_2}^2\big) \cdots \\
&\quad \times \sum_{l_P}^{N_P} \alpha_{P,l_P} \mathcal{N}\big(s_{P,t}|\mu_{P,l_P}, \sigma_{P,l_P}^2\big) \\
&= \sum_{l_1}^{N_1} \sum_{l_2}^{N_2} \cdots \sum_{l_P}^{N_P} \alpha_{1,l_1} \alpha_{2,l_2} \cdots \alpha_{P,l_P} \\
&\quad \times \mathcal{N}\big([s_{1,t}, s_{2,t}, \ldots, s_{P,t}]^{\mathrm{T}}|[\mu_{1,l_1}, \mu_{2,l_2}, \ldots, \mu_{P,l_P}]^{\mathrm{T}}, \\
&\qquad \mathrm{Diag}\big(\sigma_{1,l_1}^2, \sigma_{2,l_2}^2, \ldots, \sigma_{P,l_P}^2\big)\big) \\
&= \sum_{m=1}^{M} \omega_m \mathcal{N}(\mathbf{s}_t|\boldsymbol{\mu}_m, \mathbf{C}_m)
\end{aligned} \tag{3}$$

where $M = \prod_{i=1}^{P} N_i$ is the total number of Gaussian components used to model the joint PDF and $\omega_m = \prod_{i=1}^{P} \alpha_{i,l_i}; m = 1, \ldots, M$ are the mixing weights for each Gaus-

sian component such that $\sum_{m=1}^{M} \omega_m = 1$. The index m represents a unique combination of Gaussians from each source, i.e., $l_1, \ldots, l_P \rightarrow m$, where $l_i \in \{1, \ldots, N_i\}$ represents a Gaussian index of the ith source. The mean vector and covariance matrix of the mth Gaussian are represented by $\boldsymbol{\mu}_m = [\mu_{1,l_1}, \mu_{2,l_2}, \ldots, \mu_{P,l_P}]^{\mathrm{T}}$ and $\mathbf{C}_m = \mathrm{Diag}(\sigma^2_{1,l_1}, \sigma^2_{2,l_2}, \ldots, \sigma^2_{P,l_P})$, respectively.

2.2 Observations Distribution Model

In this subsection, the generative model of the observation signals at time instance t, similar to that in [8], is utilized to derive an expression for their joint PDF.

In this model, a hidden indication variable y_t, is used to indicate the Gaussian component of $\mathbf{s}_t$ at every instance, which is randomized by the following PDF:

$$f_y(y_t) = \sum_{m=1}^{M} \omega_m \delta(y_t - m) \tag{4}$$

where $\delta(\cdot)$ denotes Dirac's delta function and

$$y_t = \begin{cases} m & \text{if } \mathbf{s}_t \text{ is generated by the } m\text{th Gaussian component} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Note that the value of y_t determines the distribution parameters $\boldsymbol{\mu}_m$ and $\mathbf{C}_m$, the mixing process depicted in (1) is a linear mixing process, and $\mathbf{w}(t)$ is assumed to be zero-mean and white Gaussian noise. Thus, we get

$$f_{\mathbf{x}|y}(\mathbf{x}_t | y_t = m, \Theta_{\mathbf{x}}) = \mathcal{N}\big(\mathbf{x}_t | \mathbf{A}\boldsymbol{\mu}_m, \mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}\big) \tag{6}$$

where $\Theta_{\mathbf{x}} = \{\omega_m, \mathbf{A}\boldsymbol{\mu}_m, \mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}\}_{m=1}^{M}$ denotes the unknown distribution parameters of the observation signals. According to Bayes' theorem, the PDF of $\mathbf{x}_t$ is given by

$$f_{\mathbf{x}}(\mathbf{x}_t | \Theta_{\mathbf{x}}) = \sum_{m=1}^{M} \omega_m f_{\mathbf{x}|y}(\mathbf{x}_t | y_t = m, \Theta_{\mathbf{x}}) \tag{7}$$

Thus, the joint PDF of the observation signals can also be represented by the GMM with non-diagonal covariance matrices, as follows:

$$f_{\mathbf{x}}(\mathbf{x}_t | \Theta_{\mathbf{x}}) = \sum_{m=1}^{M} \omega_m \mathcal{N}(\mathbf{x}_t | \boldsymbol{\eta}_m, \mathbf{R}_m) \tag{8}$$

3 Derivation of an ML-Based Objective Function

According to the above analysis, the observation signal matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$ can be viewed as incomplete data, and the hidden data is the indication variable vector

$\mathbf{y} = [y_1, \ldots, y_T]$. We assume that a complete data set exists, $Z = (\mathbf{X}, \mathbf{y})$, and also we assume a joint density function

$$f(Z|\Theta_{\mathbf{x}}) = f(\mathbf{y}, \mathbf{X}|\Theta_{\mathbf{x}}) = f(\mathbf{y}|\mathbf{X}, \Theta_{\mathbf{x}}) f(\mathbf{X}|\Theta_{\mathbf{x}}) \tag{9}$$

With this new density function, we can define the complete data likelihood $\mathcal{L}(\Theta_{\mathbf{x}}|\mathbf{X}, \mathbf{y}) = f(\mathbf{X}, \mathbf{y}|\Theta_{\mathbf{x}})$. In order to get the maximum-likelihood estimate of the parameters, we define the following auxiliary function:

$$\mathcal{F}(\Theta_{\mathbf{x}}, \hat{\Theta}_{\mathbf{x}}) = \mathrm{E}\big[\log f(\mathbf{X}, \mathbf{y}|\Theta_{\mathbf{x}})|\mathbf{X}, \hat{\Theta}_{\mathbf{x}}\big] \tag{10}$$

where $\hat{\Theta}_{\mathbf{x}}$ are the current parameter estimates which may be an initial guess or derived from the former iteration, and used to evaluate the expectation E[.], and $\Theta_{\mathbf{x}}$ are the new parameters that we optimize to increase $\mathcal{F}$. Note that, in (10), $\mathbf{X}$ and $\hat{\Theta}_{\mathbf{x}}$ are constants, $\Theta_{\mathbf{x}}$ is the variable that we wish to adjust, and $\mathbf{y}$ is a random variable governed by the distribution $f(\mathbf{y}|\mathbf{X}, \hat{\Theta}_{\mathbf{x}})$. Therefore, the right side of (10) can be rewritten as

$$\mathrm{E}\big[\log f(\mathbf{X}, \mathbf{y}|\Theta_{\mathbf{x}})|\mathbf{X}, \hat{\Theta}_{\mathbf{x}}\big] = \int_{\Upsilon} \log f(\mathbf{X}, \mathbf{y}|\Theta_{\mathbf{x}}) f(\mathbf{y}|\mathbf{X}, \hat{\Theta}_{\mathbf{x}})\, d\mathbf{y} \tag{11}$$

Note that $f(\mathbf{y}|\mathbf{X}, \hat{\Theta}_{\mathbf{x}})$ is the marginal distribution of the unobserved data and is dependent on both the observation signals $\mathbf{X}$ and the current parameters $\hat{\Theta}_{\mathbf{x}}$, and Υ is the range of values that $\mathbf{y}$ can take. The values of the hidden variables $\mathbf{y} = \{y_t\}_{t=1}^{T}$, indicates which component density is used to generate each observation signal. Hence, if we know them, the complete data likelihood expression can be significantly simplified. In this way, the log-likelihood becomes

$$\log\big(f(\mathbf{X}, \mathbf{y}|\Theta_{\mathbf{x}})\big) = \sum_{t=1}^{T} \log\big(f(\mathbf{x}_t|y_t) f(y_t)\big) = \sum_{t=1}^{T} \log\big(\omega_{y_t} f(\mathbf{x}_t|y_t)\big) \tag{12}$$

which can be optimized using a variety of techniques. The problem, of course, is that we do not know the values of $\mathbf{y}$. However, if we assume $\mathbf{y}$ to be a random vector, then we can proceed. We first must derive an expression for the distribution of the hidden variable. Given $\hat{\Theta}_{\mathbf{x}}$, we can easily compute $f_m(\mathbf{x}_t|\hat{\vartheta}_m)$ for each t and m, where $\hat{\vartheta}_m$ is the current parameter of mth Gaussian component corresponding to the observation signals. Therefore, using Bayes' rule, we can compute

$$f(y_t|\mathbf{x}_t, \hat{\Theta}_{\mathbf{x}}) = \frac{\hat{\omega}_{y_t} f(\mathbf{x}_t|\hat{\vartheta}_{y_t})}{\sum_{m'=1}^{M} \hat{\omega}_{m'} f(\mathbf{x}_t|\hat{\vartheta}_{m'})} \tag{13}$$

and

$$f(\mathbf{y}|\mathbf{X}, \hat{\Theta}_{\mathbf{x}}) = \prod_{t=1}^{T} f(y_t|\mathbf{x}_t, \hat{\Theta}_{\mathbf{x}}) \tag{14}$$

In this case, Eq. (11) takes the form

$$\begin{aligned}\mathcal{F}(\Theta_{\mathbf{x}}, \hat{\Theta}_{\mathbf{x}}) &= \sum_{\Upsilon} \log\big(f(\mathbf{X}, \mathbf{y}|\Theta_{\mathbf{x}})\big) f(\mathbf{y}|\mathbf{X}, \hat{\Theta}_{\mathbf{x}}) \\ &= \sum_{\Upsilon} \sum_{t=1}^{T} \log\big(\omega_{y_t} f(\mathbf{x}_t|\vartheta_{y_t})\big) \prod_{t'=1}^{T} f(y_{t'}|\mathbf{x}_{t'}, \hat{\Theta}_{\mathbf{x}}) \\ &= \sum_{y_1=1}^{M} \sum_{y_2=1}^{M} \cdots \sum_{y_T=1}^{M} \sum_{t=1}^{T} \log\big(\omega_{y_t} f(\mathbf{x}_t|\vartheta_{y_t})\big) \prod_{t'=1}^{T} f(y_{t'}|\mathbf{x}_{t'}, \hat{\Theta}_{\mathbf{x}}) \\ &= \sum_{y_1=1}^{M} \sum_{y_2=1}^{M} \cdots \sum_{y_T=1}^{M} \sum_{t=1}^{T} \sum_{m=1}^{M} \delta_{m,y_t} \log\big(\omega_m f(\mathbf{x}_t|\vartheta_m)\big) \prod_{t'=1}^{T} f(y_{t'}|\mathbf{x}_{t'}, \hat{\Theta}_{\mathbf{x}}) \\ &= \sum_{m=1}^{M} \sum_{t=1}^{T} \log\big(\omega_m f(\mathbf{x}_t|\vartheta_m)\big) \sum_{y_1=1}^{M} \sum_{y_2=1}^{M} \cdots \sum_{y_T=1}^{M} \delta_{m,y_t} \prod_{t'=1}^{T} f(y_{t'}|\mathbf{x}_{t'}, \hat{\Theta}_{\mathbf{x}})\end{aligned} \tag{15}$$

In this form, $\mathcal{F}(\Theta_{\mathbf{x}}, \hat{\Theta}_{\mathbf{x}})$ looks fairly daunting, yet it can be greatly simplified. We first note that for $m \in 1, \ldots, M$

$$\begin{aligned}&\sum_{y_1=1}^{M} \sum_{y_2=1}^{M} \cdots \sum_{y_T=1}^{M} \delta_{m,y_t} \prod_{t'=1}^{T} f(y_{t'}|\mathbf{x}_{t'}, \hat{\Theta}_{\mathbf{x}}) \\ &\quad= \left(\sum_{y_1=1}^{M} \cdots \sum_{y_{t-1}=1}^{M} \sum_{y_{t+1}=1}^{M} \cdots \sum_{y_T=1}^{M} \prod_{t'=1, t'\neq t}^{T} f(y_{t'}|\mathbf{x}_{t'}, \hat{\Theta}_{\mathbf{x}})\right) f(y_t = m|\mathbf{x}_t, \hat{\Theta}_{\mathbf{x}}) \\ &\quad= \prod_{t'=1, t'\neq t}^{T} \left(\sum_{y_{t'}=1}^{M} f(y_{t'}|\mathbf{x}_{t'}, \hat{\Theta}_{\mathbf{x}})\right) f(y_t = m|\mathbf{x}_t, \hat{\Theta}_{\mathbf{x}}) = f(y_t = m|\mathbf{x}_t, \hat{\Theta}_{\mathbf{x}})\end{aligned} \tag{16}$$

since $\sum_{m=1}^{M} f(m|\mathbf{x}_{t'}, \hat{\Theta}_{\mathbf{x}}) = 1$. Substituting (16) into (15) and normalizing (15) by a factor of $1/T$, then, we can write (15) as

$$\mathcal{F}(\Theta_{\mathbf{x}}, \hat{\Theta}_{\mathbf{x}}) = \frac{1}{T} \sum_{m=1}^{M} \sum_{t=1}^{T} \gamma_{t,m} \log\big(\omega_m \mathcal{N}(\mathbf{x}_t|\boldsymbol{\eta}_m, \mathbf{R}_m)\big) \tag{17}$$

where

$$\gamma_{t,m} = f(y_t = m|\mathbf{x}_t, \hat{\Theta}_{\mathbf{x}}) = \frac{\hat{\omega}_m \mathcal{N}(\mathbf{x}_t|\hat{\boldsymbol{\eta}}_m, \hat{\mathbf{R}}_m)}{\sum_{m'=1}^{M} \hat{\omega}_{m'} \mathcal{N}(\mathbf{x}_t|\hat{\boldsymbol{\eta}}_{m'}, \hat{\mathbf{R}}_{m'})} \tag{18}$$

Therefore, the estimation of $\mathbf{A}$ can be performed as follows:

$$\hat{\mathbf{A}} = \arg\max_{\mathbf{A}} \mathcal{F}(\Theta_{\mathbf{x}}, \hat{\Theta}_{\mathbf{x}}) \tag{19}$$

Note that direct maximization of $\mathcal{F}(\Theta_{\mathbf{x}}, \hat{\Theta}_{\mathbf{x}})$ with regard to $\mathbf{A}$ is analytically cumbersome. Hence, $\mathbf{A}$ is estimated in two separate steps. In the first step, a tight upper bound on the log-likelihood of a function of $\mathbf{A}$ with the observations is obtained by applying the EM algorithm for the GMM parameter estimation. In the second step, the obtained tight upper bound is maximized over the subspace of $\mathbf{A}$.

The GMM parameters of the observation signals, obtained in the final step of the EM algorithm [2] are denoted by $\hat{\Theta}^*_{\mathbf{x}} = \{\hat{\omega}^*_m, \hat{\boldsymbol{\eta}}^*_m, \hat{\mathbf{R}}^*_m\}_{m=1}^M$. In the Appendix, it is shown that [24]

$$\mathcal{F}\big(\Theta_{\mathbf{x}}, \hat{\Theta}^*_{\mathbf{x}}\big) = -\sum_{m=1}^{M} \hat{\omega}^*_m \Big\{ \mathrm{KL}_{\mathrm{norm}}\big[\hat{\mathbf{R}}^*_m | \big(\mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}\big)\big] + \frac{1}{2}\big[\big(\hat{\boldsymbol{\eta}}^*_m - \mathbf{A}\boldsymbol{\mu}_m\big)^{\mathrm{T}}\big(\mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}\big)^{-1}\big(\hat{\boldsymbol{\eta}}^*_m - \mathbf{A}\boldsymbol{\mu}_m\big)\big]\Big\} + \mathrm{const} \quad (20)$$

where $\mathrm{KL}_{\mathrm{norm}}(\Sigma_1|\Sigma_2)$ denotes the KL divergence [12] of a zero-mean Q-variate normal density with covariance matrix Σ_2 from a zero-mean Q-variate normal density with covariance matrix Σ_1.

According to (17), one can notice that in order to estimate $\mathbf{A}$, a structure on $\boldsymbol{\eta}_m = \mathbf{A}\boldsymbol{\mu}_m$ and $\mathbf{R}_m = \mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}$ is constrained. Nevertheless, $\{\hat{\boldsymbol{\eta}}^*_m\}_{m=1}^M$ and $\{\hat{\mathbf{R}}^*_m\}_{m=1}^M$ in $\hat{\Theta}^*_{\mathbf{x}}$ are estimated by applying the EM algorithm in which this constraint is not enforced, and the maximum of $\mathcal{F}(\Theta_{\mathbf{x}}, \hat{\Theta}_{\mathbf{x}})$ with regard to $\{\mathbf{A}, \{\boldsymbol{\eta}_m\}_{m=1}^M, \{\mathbf{R}_m\}_{m=1}^M\}$ is not strictly attained. The middle term of (20) denotes the divergence of $\hat{\boldsymbol{\eta}}^*_m$ and $\mathbf{A}\boldsymbol{\mu}_m$, which is always nonnegative. Thus, by removing the middle term of (20), the following upper bound is obtained:

$$\mathcal{F}\big(\mathbf{A}, \hat{\Theta}^*_{\mathbf{x}}\big) \le \mathcal{F}^*\big(\mathbf{A}, \hat{\Theta}^*_{\mathbf{x}}\big) \quad (21)$$

where

$$\mathcal{F}^*\big(\mathbf{A}, \hat{\Theta}^*_{\mathbf{x}}\big) = -\sum_{m=1}^{M} \hat{\omega}^*_m \big\{ \mathrm{KL}_{\mathrm{norm}}\big[\hat{\mathbf{R}}^*_m | \big(\mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}\big)\big]\big\} \quad (22)$$

The following simulated example is used to examine in a graphical manner the relation between $\mathcal{F}(\mathbf{A}, \hat{\Theta}^*_{\mathbf{x}})$ and $\mathcal{F}^*(\mathbf{A}, \hat{\Theta}^*_{\mathbf{x}})$. Two thousand samples of two source signals are synthesized by the following GMM PDF:

$$f_{\mathbf{s}}(\mathbf{s}_t|\Theta_{\mathbf{s}}) = \sum_{m=1}^{4} \omega_m \mathcal{N}(\mathbf{s}_t|\boldsymbol{\mu}_m, \mathbf{C}_m) \quad (23)$$

where the GMM order of each source is 2. The values of the mixing weights, mean vectors and covariance matrices are $\omega_1 = 0.06$, $\omega_2 = 0.14$, $\omega_3 = 0.24$, $\omega_4 = 0.56$, $\boldsymbol{\mu}_1 = [-5, -5]^{\mathrm{T}}$, $\boldsymbol{\mu}_2 = [-5, 5]^{\mathrm{T}}$, $\boldsymbol{\mu}_3 = [5, -5]^{\mathrm{T}}$, $\boldsymbol{\mu}_4 = [5, 5]^{\mathrm{T}}$, $\mathbf{C}_1 = \mathrm{diag}([10, 1])$, $\mathbf{C}_2 = \mathrm{diag}([10, 12])$, $\mathbf{C}_3 = \mathrm{diag}([3, 1])$, $\mathbf{C}_4 = \mathrm{diag}([3, 12])$, respectively. The source signals are mixed by a unitary mixing matrix

$$\mathbf{A} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \quad (24)$$

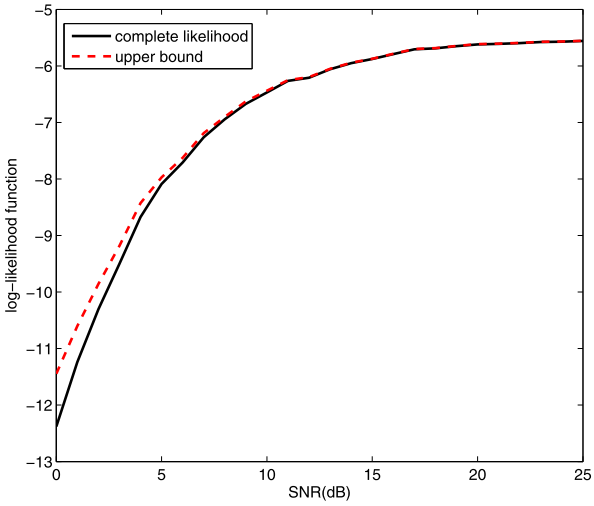

Fig. 1 The complete log-likelihood function of $\mathcal{F}(\mathbf{A}, \hat{\Theta}_{\mathbf{x}}^*)$ (*solid curve*) and its tight upper bound $\mathcal{F}^*(\mathbf{A}, \hat{\Theta}_{\mathbf{x}}^*)$ (*dashed curve*) as a function of the SNR of the observations

where $\alpha = 45°$. The observations are contaminated by Gaussian noise, and the signal-to-noise ratio (SNR) ranges from 0 dB to 25 dB. The GMM parameters of the mixed source signals are estimated via the EM algorithm. Hence, $\mathcal{F}(\mathbf{A}, \hat{\Theta}_{\mathbf{x}}^*)$ and $\mathcal{F}^*(\mathbf{A}, \hat{\Theta}_{\mathbf{x}}^*)$ can be sketched as a function of the SNR of the observations, as depicted in Fig. 1. It is shown that $\mathcal{F}^*(\mathbf{A}, \hat{\Theta}_{\mathbf{x}}^*)$ is a tight upper bound for $\mathcal{F}(\mathbf{A}, \hat{\Theta}_{\mathbf{x}}^*)$, especially when the SNR of the observations is high.

4 PARAFAC-Based Estimation of the Mixing Matrix

In this section, we discuss the way to estimate the mixing matrix $\mathbf{A}$ based on the objective function (22). According to the above analysis, maximizing the objective function (22) with regard to the mixing matrix $\mathbf{A}$ results in the following equation array:

$$\begin{cases} \hat{\mathbf{R}}_1^* = \mathbf{A}\mathbf{C}_1\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}} \\ \hat{\mathbf{R}}_2^* = \mathbf{A}\mathbf{C}_2\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}} \\ \vdots \\ \hat{\mathbf{R}}_M^* = \mathbf{A}\mathbf{C}_M\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}} \end{cases} \tag{25}$$

Note that the component $\mathbf{R}_{\mathbf{w}}$ will lead to estimation error for the mixing matrix. Nevertheless, this estimation error can be decreased by transforming the JD problem (25) into a new JD problem in which the new equations are composed by the differences between any two equations in (25). In such a way, the component $\mathbf{R}_{\mathbf{w}}$ is eliminated.

It is obvious that JD methods can be used to estimate the mixing matrix from (25) when the mixtures are determined/overdetermined. However, the JD method does not work for the underdetermined case when $Q < P$. In this paper, the solution will be obtained by interpreting (26) as a tensor decomposition problem, similar to our previous work in [7].

Stack the matrices $\hat{\mathbf{R}}_1^*, \ldots, \hat{\mathbf{R}}_M^*$ in a tensor $\mathcal{R} \in \mathbb{R}^{Q\times Q\times M}$ as follows: $(\mathcal{R})_{ijm} = (\hat{\mathbf{R}}_m^*)_{ij}$, $i = 1, \ldots, Q$, $j = 1, \ldots, Q$, $m = 1, \ldots, M$. Define a matrix $\mathbf{C} \in \mathbb{R}^{M\times P}$ by $(\mathbf{C})_{mk} = (\mathbf{C}_m)_{kk}$, $k = 1, \ldots, P$, $m = 1, \ldots, M$. Then, we have

$$r_{ijm} = \sum_{k=1}^{P} a_{ik} a_{jk} c_{mk} \tag{26}$$

which we write as

$$\mathcal{R} = \sum_{k=1}^{P} \mathbf{a}_k \circ \mathbf{a}_k \circ \mathbf{c}_k \tag{27}$$

where $\circ$ denotes the tensor outer product, $\mathbf{a}_k$ and $\mathbf{c}_k$ are the kth column of $\mathbf{A}$ and $\mathbf{C}$ respectively.

In this way, the mixing matrix $\mathbf{A}$ can be obtained by solving the following problem. Given the third-order tensor $\mathcal{R} \in \mathbb{R}^{Q\times Q\times M}$, we can compute its canonical decomposition (CAND) with P components of rank-one tensors that best approximates $\mathcal{R}$, i.e.,

$$\min_{\mathbf{A},\mathbf{C}} \left\| \mathcal{R} - \sum_{k=1}^{P} \mathbf{a}_k \circ \mathbf{a}_k \circ \mathbf{c}_k \right\|^2 \tag{28}$$

where $\| \|$ is the Frobenius norm.

Several algorithms can be used for performing tensor decomposition. The standard way for computing the tensor decomposition, is by using an "alternating least squares" (ALS) algorithm [14]. Several improved versions, such as the enhanced line search (ELS) [18], extrapolating search direction (ESD) [4], are proposed to accelerate the rate of convergence of ALS. Hence, the ALS is chosen here to compute the CAND. To a large extent, the practical importance of PARAFAC stems from its uniqueness properties. It is clear that the PARAFAC can only be unique up to a permutation of the rank-1 terms and scaling of the factors of the rank-1 terms. Therefore, we call the tensor decomposition (27) essentially unique if any other matrix pair $\mathbf{A}'$ and $\mathbf{C}'$ that satisfies (27) related to $\mathbf{A}$ and $\mathbf{C}$ via

$$\mathbf{A} = \mathbf{A}'\mathbf{P}\Delta_1, \qquad \mathbf{C} = \mathbf{C}'\mathbf{P}\Delta_2 \tag{29}$$

with $\Delta_1, \Delta_2 \in \mathbb{R}^{P\times P}$ diagonal matrices, satisfying $\Delta_1\Delta_1\Delta_2 = \mathbf{I}$, and $\mathbf{P} \in \mathbb{R}^{P\times P}$ is a permutation matrix.

Definition The Kruskal rank [11] or k-rank of a matrix $\mathbf{A}$, denoted by $\kappa_{\mathbf{A}}$, is the maximal number λ such that any set of λ columns of $\mathbf{A}$ is linearly independent.

Theorem *The PARAFAC decomposition of* (27) *is essentially unique if* [11]

$$2\kappa_{\mathbf{A}} + \kappa_{\mathbf{C}} \geq 2(P + 1) \tag{30}$$

Generally, a matrix is of full rank and full k-rank. Hence, in practice, $\kappa_{\mathbf{A}} = \min(Q, P)$ and $\kappa_{\mathbf{C}} = \min(M, P)$.

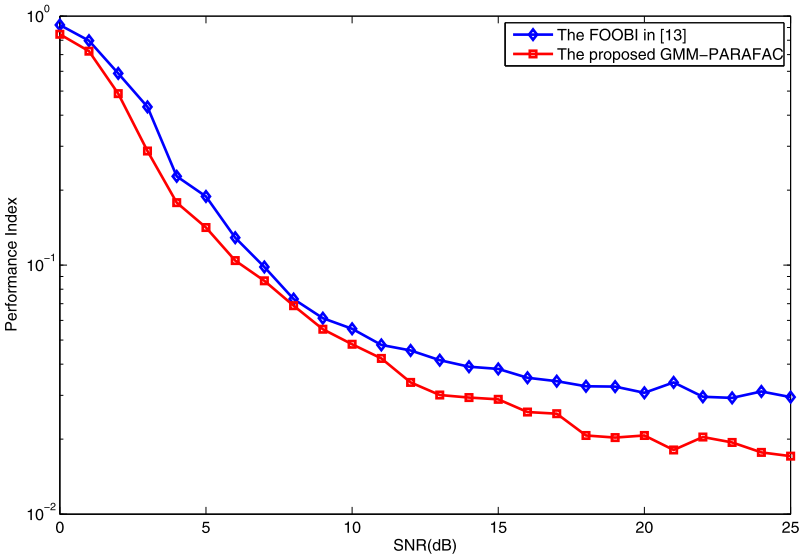

Fig. 2 The average relative error of the tested algorithms versus the SNR of the observations in the 3×4 underdetermined mixture with synthetic sources

Therefore, we can come to the following conclusion: when $Q \geq P$, $P \geq 2$, then the identifiable condition is $M \geq 2$; when $Q < P$, if $M \geq P$, then the identifiable condition is $P \leq 2Q - 2$, if $M < P$, then the identifiable condition is $P < Q - 1 + M/2$.

5 Simulations

In this section, using the experimental protocols and set-ups that are similar to those in [8], simulations with synthesized data and speech signals are used to demonstrate the performance of the proposed GMM-PARAFAC algorithm. The performance is evaluated and compared in terms of the relative error performance index (PI) versus the SNR of the observations. Here the relative error PI is defined as follows: $\mathrm{PI} = \mathrm{E}\{\|\mathbf{A} - \hat{\mathbf{A}}\|/\|\mathbf{A}\|\}$, where $\hat{\mathbf{A}}$ denotes the optimally ordered and scaled estimate of $\mathbf{A}$.

First, we consider blind identification of a 3×4 underdetermined mixtures with synthesized sources. Each of the sources is generated according to the following GMM PDF $f_s = 0.5\mathcal{N}(s; -5, 1) + 0.5\mathcal{N}(s; 5, 1)$. The sources are mixed by a mixing matrix whose elements are randomly drawn from [0, 2]. Additive Gaussian noise is added in the mixing process. The FOOBI algorithm presented in [15] is implemented as the baseline algorithm. 100 Monte Carlo experiments are run.

Figure 2 shows the PI as a function of the SNR when $N = 5000$ samples are transmitted. Figure 3 shows the PI as a function of the length of the data set, when the SNR is 20 dB. We can see that on average the proposed GMM-PARAFAC algorithm has better performance than the FOOBI algorithm in terms of the relative error. The main reason is that the GMM-PARAFAC method provides the ML estimate of the

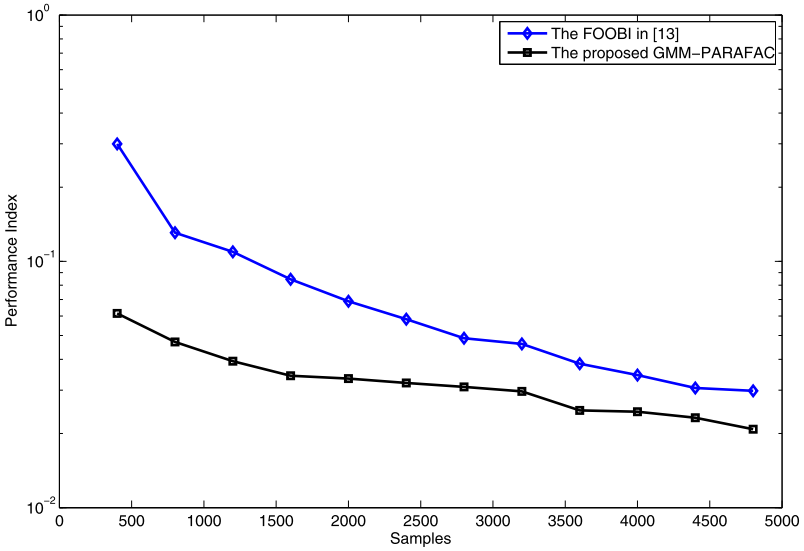

Fig. 3 The average relative error of the tested algorithms versus the number of the samples in the 3×4 underdetermined mixture with synthetic sources

mixing matrix. Moreover, it should be pointed out that the superiority of the proposed GMM-PARAFAC method is clearer when the data block is short, because estimating higher-order (HO) statistics from sample data generally requires longer data sets to obtain sufficient estimation accuracy.

Second, we also consider blind identification of a 3×4 underdetermined mixtures in which the sources are speech signals (available at http://www.kecl.ntt.co.jp/icl/signal/sawada/webdemo/bssdemo.html), which are sampled at 8000 Hz, and recorded with 8 bits per sample. The time duration of the sources is 5 s.

The sources are mixed by a mixing matrix whose elements are randomly drawn from [0, 2]. Additive Gaussian noise is added in the mixing process, and the SNR of the observations ranges from 0 dB to 25 dB. The SOBIUM algorithm in [13] and FOOBI in [15] are implemented as the baseline algorithms. The tested algorithms are operated under the following settings: (1) the GMM order in the GMM-PARAFAC algorithm is determined according to the Bayesian information criterion (BIC) [21]. Here, the GMM order is set to 30; (2) the mean vectors estimated by the GMM-PARAFAC algorithm is enforced to zero; (3) the number of delays in the SOBIUM algorithm is set to 10. 100 Monte Carlo experiments are run.

Figure 4 depicts the behavior of the average relative error of the tested algorithms versus the SNR of the observations in the 3×4 underdetermined mixture. According to Fig. 4, the identification performance of the tested algorithms improves with the increase of SNR of the observations from 0 dB to 25 dB. The proposed GMM-PARAFAC algorithm has the best performance, followed by the SOBIUM in [13], and the FOOBI in [15]. It is due to the fact that the flexible source distribution is applied in the GMM-PARAFAC algorithm. The simulation results with both synthe-

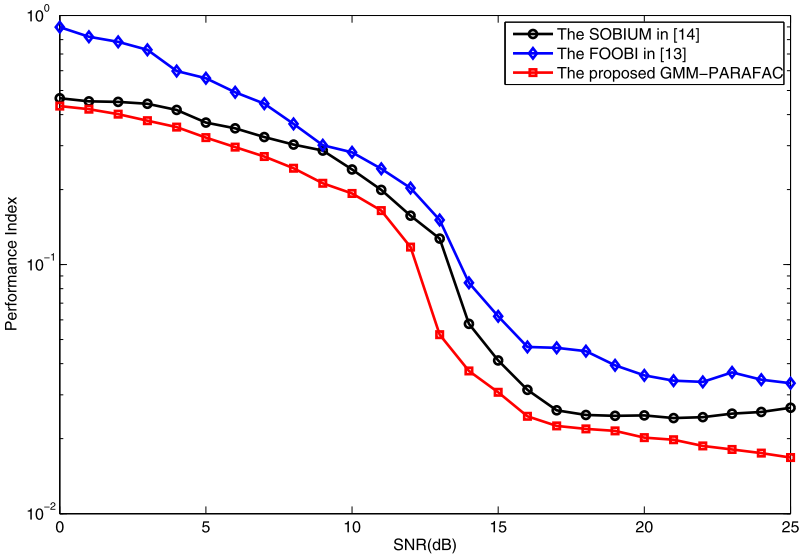

Fig. 4 The average relative error of the tested algorithms versus the SNR of the observations in the 3×4 underdetermined mixture with speech signals

sized data and real speech signals validate the superiority of the proposed GMM-PARAFAC method as compared with the baseline methods.

6 Conclusions

We have presented a novel algorithm named GMM-PARAFAC for blind identification of noisy instantaneous linear mixture. The GMM-PARAFAC algorithm uses GMM to model non-Gaussianity of the independent sources. In order to derive an ML estimation of the mixing matrix, we formulate an ML function of the mixing matrix by estimating the GMM parameters of the observation via the EM algorithm. Aiming for decreasing the computational complexity of the EM algorithm, a tight upper bound of the log-likelihood, instead of the log-likelihood itself, is used to estimate the mixing matrix. The mixing matrix is estimated by the PARAFAC decomposition of a three-way tensor which is composed of the covariance matrices of the GMM of the observations. Owing to the powerful uniqueness properties of the PARAFAC decomposition, the proposed algorithm works well in the underdetermined mixture. The simulation results show that the proposed GMM-PARAFAC algorithm has superior identification performance to the SOBIUM and FOOBI algorithms.

Acknowledgements The authors would like to thank Lieven De Lathauwer for sharing with us the Matlab codes of the FOOBI algorithm. They also thank the anonymous reviewers for their careful reading and helpful remarks, which have contributed to improving the clarity of the paper. This work is supported in part by the Natural Science Foundation of China under Grant 61001106 and the National Program on Key Basic Research Project of China under Grant 2009CB320400.

Appendix

In this appendix, it is shown that (17) can be formulated as (20).

According to (19)

$$\mathcal{F}\big(\Theta_{\mathbf{x}}, \hat{\Theta}^*_{\mathbf{x}}\big) = \frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma^*_{t,m}\log\big(\omega_m N(\mathbf{x}_t|\boldsymbol{\eta}_m, \mathbf{R}_m)\big) \tag{31}$$

where $\gamma^*_{t,m} = \hat{\omega}^*_m N(\mathbf{x}_t|\hat{\boldsymbol{\eta}}^*_m, \hat{\mathbf{R}}^*_m)/\sum_{m'=1}^{M}\hat{\omega}^*_{m'}N(\mathbf{x}_t|\hat{\boldsymbol{\eta}}^*_{m'}, \hat{\mathbf{R}}^*_{m'})$. Therefore

$$\begin{aligned}\mathcal{F}\big(\Theta_{\mathbf{x}}, \hat{\Theta}^*_{\mathbf{x}}\big) = -\frac{1}{T}\sum_{m=1}^{M}\Bigg\{&\frac{1}{2}\sum_{t=1}^{T}\gamma^*_{t,m}\log|2\pi\mathbf{R}_m| \\ &+\frac{1}{2}\sum_{t=1}^{T}\gamma^*_{t,m}\operatorname{tr}\big[\mathbf{R}_m^{-1}(\mathbf{x}_t-\boldsymbol{\eta}_m)(\mathbf{x}_t-\boldsymbol{\eta}_m)^{\mathrm{T}}\big] - \sum_{t=1}^{T}\gamma^*_{t,m}\log\omega_m\Bigg\}\end{aligned} \tag{32}$$

Since trace is a linear operator, the summation with regard to t in the mid-term of (32) can be inserted into the trace operator. Hence, (32) can be rewritten in the following form:

$$\begin{aligned}\mathcal{F}\big(\Theta_{\mathbf{x}}, \hat{\Theta}^*_{\mathbf{x}}\big) = -\frac{1}{T}\sum_{m=1}^{M}\Bigg\{&\frac{1}{2}\sum_{t=1}^{T}\gamma^*_{t,m}\log|2\pi\mathbf{R}_m| \\ &+\frac{1}{2}\operatorname{tr}\Bigg[\mathbf{R}_m^{-1}\sum_{t=1}^{T}\gamma^*_{t,m}(\mathbf{x}_t-\boldsymbol{\eta}_m)(\mathbf{x}_t-\boldsymbol{\eta}_m)^{\mathrm{T}}\Bigg] - \sum_{t=1}^{T}\gamma^*_{t,m}\log\omega_m\Bigg\}\end{aligned} \tag{33}$$

The factor $\sum_{t=1}^{T}\gamma^*_{t,m}$ can be extracted out of the main brackets and (33) can be formulated as follows:

$$\begin{aligned}\mathcal{F}\big(\Theta_{\mathbf{x}}, \hat{\Theta}^*_{\mathbf{x}}\big) = -\frac{1}{T}\sum_{m=1}^{M}\sum_{t=1}^{T}\gamma^*_{t,m}\Bigg\{&\frac{1}{2}\operatorname{tr}\Bigg[\mathbf{R}_m^{-1}\times\underbrace{\Bigg(\frac{\sum_{t=1}^{T}\gamma^*_{t,m}(\mathbf{x}_t-\boldsymbol{\eta}_m)(\mathbf{x}_t-\boldsymbol{\eta}_m)^{\mathrm{T}}}{\sum_{t=1}^{T}\gamma^*_{t,m}}\Bigg)}_{\mathbf{G}_m}\Bigg] \\ &+\frac{1}{2}\log|2\pi\mathbf{R}_m| - \log\omega_m\Bigg\}\end{aligned} \tag{34}$$

The updating equations of the EM algorithm for GMM parameter estimation can be formulated as follows:

$$\begin{cases}\hat{\omega}^*_m = \frac{1}{T}\sum_{t=1}^{T}\gamma^*_{t,m} \\ \hat{\boldsymbol{\eta}}^*_m = \frac{\sum_{t=1}^{T}\gamma^*_{t,m}\mathbf{x}_t}{\sum_{t=1}^{T}\gamma^*_{t,m}} \\ \hat{\mathbf{R}}^*_m + \hat{\boldsymbol{\eta}}^*_m(\hat{\boldsymbol{\eta}}^*_m)^{\mathrm{T}} = \frac{\sum_{t=1}^{T}\gamma^*_{t,m}\mathbf{x}_t\mathbf{x}_t^{\mathrm{T}}}{\sum_{t=1}^{T}\gamma^*_{t,m}}\end{cases} \tag{35}$$

Therefore

$$\begin{aligned}\mathbf{G}_m &= \hat{\mathbf{R}}_m^* + \hat{\boldsymbol{\eta}}_m^* \big(\hat{\boldsymbol{\eta}}_m^*\big)^{\mathrm{T}} - \boldsymbol{\eta}_m \big(\hat{\boldsymbol{\eta}}_m^*\big)^{\mathrm{T}} - \hat{\boldsymbol{\eta}}_m^* \boldsymbol{\eta}_m^{\mathrm{T}} + \boldsymbol{\eta}_m \boldsymbol{\eta}_m^{\mathrm{T}} \\ &= \hat{\mathbf{R}}_m^* + \big(\hat{\boldsymbol{\eta}}_m^* - \boldsymbol{\eta}_m\big)\big(\hat{\boldsymbol{\eta}}_m^* - \boldsymbol{\eta}_m\big)^{\mathrm{T}} \end{aligned} \tag{36}$$

Substitution of (36) into (35), yields the expression (37) as follows:

$$\begin{aligned}&\mathcal{F}\big(\Theta_{\mathbf{x}}, \hat{\Theta}_{\mathbf{x}}^*\big) \\ &\quad = -\sum_{m=1}^{M} \hat{\omega}_m^* \\ &\qquad \times \underbrace{\left\{\frac{1}{2}\operatorname{tr}\big[\mathbf{R}_m^{-1}\hat{\mathbf{R}}_m^*\big] + \frac{1}{2}\big(\hat{\boldsymbol{\eta}}_m^* - \boldsymbol{\eta}_m\big)^{\mathrm{T}}\mathbf{R}_m^{-1}\big(\hat{\boldsymbol{\eta}}_m^* - \boldsymbol{\eta}_m\big) + \frac{1}{2}\log|2\pi\mathbf{R}_m| - \log\omega_m\right\}}_{\mathbf{H}_m}\end{aligned} \tag{37}$$

Applying that $\boldsymbol{\eta}_m = \mathbf{A}\boldsymbol{\mu}_m$ and $\mathbf{R}_m = \mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}$

$$\begin{aligned}\mathbf{H}_m &= \frac{1}{2}\operatorname{tr}\big[\big(\mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}\big)^{-1}\hat{\mathbf{R}}_m^*\big] \\ &\quad + \frac{1}{2}\big(\hat{\boldsymbol{\eta}}_m^* - \mathbf{A}\boldsymbol{\mu}_m\big)^{\mathrm{T}}\big(\mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}\big)^{-1}\big(\hat{\boldsymbol{\eta}}_m^* - \mathbf{A}\boldsymbol{\mu}_m\big) \\ &\quad + \frac{1}{2}Q\log 2\pi - \frac{1}{2}\log\big|\big(\mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}\big)^{-1}\big| - \frac{1}{2}\log\big|\hat{\mathbf{R}}_m^*\big| \\ &\quad + \frac{1}{2}\log\big|\hat{\mathbf{R}}_m^*\big| - \frac{1}{2}Q + \frac{1}{2}Q - \log\omega_m\end{aligned} \tag{38}$$

Express the KL divergence of a zero-mean Q-variate normal density with covariance matrix Σ_2 from a zero-mean Q-variate normal density with covariance matrix Σ_1 as

$$\mathrm{KL}_{\mathrm{norm}}[\Sigma_1|\Sigma_2] = \frac{1}{2}\operatorname{tr}\big[\Sigma_2^{-1}\Sigma_1\big] - \frac{1}{2}\log\big|\Sigma_2^{-1}\Sigma_1\big| - \frac{1}{2}Q \tag{39}$$

Then (38) can be formulated in the following form:

$$\begin{aligned}\mathbf{H}_m &= \underbrace{\frac{1}{2}\operatorname{tr}\big[\big(\mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}\big)^{-1}\hat{\mathbf{R}}_m^*\big] - \frac{1}{2}\log\big|\big(\mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}\big)^{-1}\hat{\mathbf{R}}_m^*\big| - \frac{1}{2}Q}_{\mathrm{KL}_{\mathrm{norm}}[\hat{\mathbf{R}}_m^*|\mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}}+\mathbf{R}_{\mathbf{w}}]} \\ &\quad + \frac{1}{2}\big(\hat{\boldsymbol{\eta}}_m^* - \mathbf{A}\boldsymbol{\mu}_m\big)^{\mathrm{T}}\big(\mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}\big)^{-1}\big(\hat{\boldsymbol{\eta}}_m^* - \mathbf{A}\boldsymbol{\mu}_m\big) \\ &\quad + \underbrace{\frac{1}{2}Q\log 2\pi e + \frac{1}{2}\log\big|\hat{\mathbf{R}}_m^*\big| - \log\omega_m}_{\mathrm{const}}\end{aligned} \tag{40}$$

where $\mathrm{KL}_{\mathrm{norm}}[\hat{\mathbf{R}}_m^*|\mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}]$ is the KL divergence between two zero-mean Q-variate normal densities with a covariance matrix $\hat{\mathbf{R}}_m^*$ and $\mathbf{A}\mathbf{C}_m\mathbf{A}^{\mathrm{T}} + \mathbf{R}_{\mathbf{w}}$, respectively. Therefore, (20) can be derived by inserting (40) into (37).

References

1. A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, E. Moulines, A blind source separation technique using second-order statistics. IEEE Trans. Signal Process. **45**(2), 434–444 (1997)
2. J. Bilmes, A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, Tech. Rep. ICSI-TR-97-021, University of Berkely, 1997, http://www.Citeseer.nj.nec.com/blimes98gentle.html
3. C.F. Caiafa, A.N. Proto, Separation of statistically dependent sources using an l_2-distance non-Gaussianity measure. Signal Process. **86**(11), 3404–3420 (2006)
4. Y. Chen, D. Han, L. Qi, New ALS methods with extrapolating search directions and optimal step size for complex-valued tensor decompositions. IEEE Trans. Signal Process. **59**(12), 5888–5898 (2011)
5. A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc., Ser. B, Methodol. **39**(1), 1–38 (1977)
6. A. Ferréol, L. Albera, P. Chevalier, Fourth-order blind identification of underdetermined mixtures of sources (FOBIUM). IEEE Trans. Signal Process. **53**(5), 1640–1653 (2005)
7. F. Gu, H. Zhang, W. Wang, D. Zhu, Generalized generating function with Tucker decomposition and alternating least squares for underdetermined blind identification. EURASIP J. Adv. Signal Process. **2013**, 124 (2013). doi:10.1186/1687-6180-2013-124
8. F. Gu, H. Zhang, W. Wang, An expectation-maximization algorithm for blind separation of noisy mixtures using Gaussian mixture model, IEEE Trans. Signal Process. (revised)
9. A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. **10**(3), 1697–1710 (1999)
10. A. Karfoul, L. Albera, G. Birot, Blind underdetermined mixture identification by joint canonical decomposition of HO cumulants. IEEE Trans. Signal Process. **58**(2), 638–649 (2010)
11. T.G. Kolda, B.W. Bader, Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
12. S. Kullback, R.A. Leibler, On information and sufficiency. Ann. Math. Stat. **22**(1), 79–86 (1951)
13. L.D. Lathauwer, J. Castaing, Blind identification of underdetermined mixtures by simultaneous matrix diagonalization. IEEE Trans. Signal Process. **56**(3), 1096–1105 (2008)
14. L.D. Lathauwer, B.D. Moor, J. Vandewalle, On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors. SIAM J. Matrix Anal. Appl. **21**, 1324–1342 (2000)
15. L.D. Lathauwer, J. Castaing, J.F. Cardoso, Fourth-order cumulant-based blind identification of underdetermined mixtures. IEEE Trans. Signal Process. **55**(6), 2965–2973 (2007)
16. Q. Li, A. Barron, Mixture density estimation, in *Advances in Neural Information Processing Systems*, vol. 12 (MIT Press, Cambridge, 2000), pp. 279–285
17. É. Moulines, J.F. Cardoso, E. Gassiat, Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models, in *IEEE International Conference on Audio, Speech, and Signal Processing*, Munich, Germany (1997), pp. 3617–3620
18. D. Nion, L.D. Lathauwer, An enhanced line search scheme for complex-valued tensor decompositions: application in DS-CDMA. Signal Process. **88**(3), 749–755 (2008)
19. D.T. Pham, J.F. Cardoso, Blind separation of instantaneous mixtures of nonstationary sources. IEEE Trans. Signal Process. **49**(9), 1837–1848 (2001)
20. T. Routtenberg, J. Tabrikian, MIMO-AR system identification and blind source separation for GMM-distributed sources. IEEE Trans. Signal Process. **57**(5), 1717–1730 (2009)
21. G. Schwarz, Estimating the dimension of a model. Ann. Stat. **6**(2), 461–464 (1978)
22. H. Snoussi, A.M. Djafari, Bayesian source separation with mixture of Gaussians prior for sources and Gaussian prior for mixture coefficients, in *Proc. of MaxEnt, Amer. Inst. Physics. in Bayesian Inference and Maximum Entropy Methods*, France (2000), pp. 388–406
23. P. Tichavský, Z. Koldovský, Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources. IEEE Trans. Signal Process. **59**(3), 1037–1047 (2011)

24. K. Todros, J. Tabrikian, Blind separation of independent sources using Gaussian mixture model. IEEE Trans. Signal Process. **55**(7), 3645–3658 (2007)
25. H. Yang, S. Amari, Adaptive online learning algorithms for blind separation: maximum entropy and minimum mutual information. Neural Comput. **9**, 1457–1482 (1997)
26. Y. Zhang, X. Shi, C.H. Chen, A Gaussian mixture model for underdetermined independent component analysis. Signal Process. **86**, 1538–1549 (2006)