EURASIP Journal on
Audio, Speech, and Music Processing
a SpringerOpen Journal

RESEARCH Open Access

Bayesian group sparse learning for music source separation

Jen-Tzung Chien* and Hsin-Lung Hsieh

Abstract

Nonnegative matrix factorization (NMF) is developed for parts-based representation of nonnegative signals with the sparseness constraint. The signals are adequately represented by a set of basis vectors and the corresponding weight parameters. NMF has been successfully applied for blind source separation and many other signal processing systems. Typically, controlling the degree of sparseness and characterizing the uncertainty of model parameters are two critical issues for model regularization using NMF. This paper presents the *Bayesian group sparse learning* for NMF and applies it for single-channel music source separation. This method reconstructs the rhythmic or repetitive signal from a *common subspace* spanned by the shared bases for the whole signal and simultaneously decodes the harmonic or residual signal from an *individual subspace* consisting of separate bases for different signal segments. A *Laplacian scale mixture* distribution is introduced for sparse coding given a sparseness control parameter. The relevance of basis vectors for reconstructing two groups of music signals is automatically determined. A Markov chain Monte Carlo procedure is presented to infer two sets of model parameters and hyperparameters through a sampling procedure based on the conditional posterior distributions. Experiments on separating single-channel audio signals into rhythmic and harmonic source signals show that the proposed method outperforms baseline NMF, Bayesian NMF, and other group-based NMF in terms of signal-to-interference ratio.

Keywords: Bayesian sparse learning; Signal reconstruction; Subspace approach; Group sparsity; Nonnegative matrix factorization; Single-channel source separation

1 Introduction

Many problems in audio, speech and music processing can be tackled through matrix factorization. Different cost functions and constraints may lead to different factorized matrices. This procedure can identify underlying sources from the mixed signals through blind source separation [1]. Nonnegative matrix factorization (NMF) is designed to find an approximate factorization $X \approx AS$ for a data matrix X into a basis matrix A and a weight matrix S which are all nonnegative [2]. Some divergence measures have been proposed to derive solutions to NMF [3,4]. NMF provides a useful learning tool for clustering as well as for classification. When a portion of labeled data are available, the semi-supervised NMF was developed for an improved classification system [5]. Different from standard principal component analysis (PCA) and independent component analysis (ICA), NMF only allows additive combination due to the nonnegative constraints on matrices A and S. Nevertheless, nonnegative PCA and nonnegative ICA were proposed for blind source separation in the presence of nonnegative image and music sources [6].

On the other hand, NMF conducts a parts-based sparse representation where only a few components or bases are relevant for representation of input nonnegative matrix X. The sparseness constraint is imposed in objective function [2]. An automatic relevance determination (ARD) scheme [7-9] is developed to determine relevant bases for sparse representation. Such sparse coding is efficient and robust. However, controlling the sparseness or smoothness is influential for system performance. Bayesian learning is beneficial to deal with sparse representation [9] and model regularization [7]. In [10], Bayesian learning was performed for sparse representation of image data where Laplacian distribution was used as prior density. The ℓ_1-regularized optimization was comparably performed. In addition, the group-based NMF [11] was proposed to

*Correspondence: jtchien@nctu.edu.tw
Department of Electrical and Computer Engineering, National Chiao Tung University, Taiwan 30010, Republic of China

capture the intra-subject variations and the inter-subject variations in EEG signals. In [12], the group sparse NMF was proposed by minimizing the Itakura-Saito divergence between X and AS. In [13], NMF was applied for drum source separation where the factorized components were partitioned into rhythmic sources and harmonic sources. No Bayesian learning was performed in [11-13].

More recently, a Bayesian NMF approach [14] was proposed for model selection and image reconstruction. This approach inferred NMF model by a variational Bayes method and a Markov chain Monte Carlo (MCMC) algorithm. In [15], a Bayesian NMF with gamma priors for source signals and mixture weights was implemented through a MCMC algorithm. In [16], the Bayesian NMF with Gaussian likelihood and exponential prior was constructed for image feature extraction where the posterior distribution was approximated by Gibbs sampling procedure. In [17], a Bayesian approach for blind separation of linear mixtures of sources was developed. The Student t distribution for mixture weights was introduced to achieve sparse basis representation. The underdetermined noisy mixtures were separated. However, the case of nonnegative source was not applied. Besides, single-channel source separation is known as an underdetermined problem. In [18], the harmonic structure information was adopted to estimate the demixed instrumental sources. In [19], the NMF was applied for single-channel speech separation where the speech of target speaker over that of masking speaker was enhanced by using sparse dictionaries learned on a phoneme level for individual speakers.

This paper addresses the problem of underdetermined source separation based on NMF for an application to music source separation [20]. The uses of NMF and Bayesian theory to source separation are not new since they have been many papers [11-13,15]. But, to our best knowledge, the novelty of this paper is to propose *Bayesian group sparse* (*BGS*) *learning* using *Laplacian distribution* and *Laplacian scale mixture* (*LSM*) *distribution* and apply it for single-channel music signal separation. We present a group-based NMF where the groups of common bases and individual bases are estimated for blind separation of rhythmic sources and harmonic sources, respectively. Bayesian sparse learning is developed by introducing LSM distributions as the priors for two groups of reconstruction weights. Gamma priors are used to represent two groups of nonnegative basis components. The BGS-NMF algorithm is accordingly established. A MCMC algorithm is derived to infer BGS-NMF parameters and hyperparameters according to *full Bayesian* theory. The rhythmic sources and harmonic sources are reconstructed through the relevant bases in common subspace and individual subspace, respectively. In the experiments, the proposed BGS-NMF is evaluated and compared with the other NMF methods for single-channel separation of audio signals into rhythmic signals and harmonic signals. From comparative study, we find that the improvement of separation performance benefits from Bayesian modeling, group basis representation, and sparse signal reconstruction. Sparser priors identify fewer but more relevant bases and correspondingly lead to a better performance in terms of signal-to-interference ratio.

The remaining of this paper is organized as follows. In the next section, the related studies on NMF and group basis representation are surveyed. Some Bayesian learning approaches are addressed. Section 3 highlights on the construction of BGS-NMF model as well as the inference procedure based on MCMC algorithm. The conditional posterior distributions of different parameters and hyperparameters are derived in the sampling procedure. Section 4 reports a series of experiments on underdetermined music source separation with different music sources. The convergence condition in MCMC sampling is investigated. The evaluation of demixed signals in terms of signal-to-interference ratio is reported. Finally, the conclusions drawn by this study are provided in Section 5.

2 Background survey

In what follows, nonnegative matrix factorization (NMF) and its extensions to different regularization functions are introduced. Several approaches to group basis representation are addressed. Group sparse coding is surveyed. Then Bayesian learning methods for matrix factorization and other related tasks are introduced.

2.1 Nonnegative matrix factorization

NMF is a linear model where the observed signals, factorized signals, and source signals are all assumed to be nonnegative. Given a data matrix $X = \{X_{ik}\}$, NMF estimates two factorized matrices $A = \{A_{ij}\}$ and $S = \{S_{jk}\}$ by minimizing the reconstruction error between X and AS. In [2], the sparseness constraint was imposed on minimization of an objective function $\mathcal{F}$ which is based on a regularized error function

$$\|X - AS\|^2 + \eta_a \sum_i \sum_j f(A_{ij}) + \eta_s \sum_j \sum_k f(S_{jk}) \quad (1)$$

where $\eta_{\mathrm{a}} \geq 0$ and $\eta_{\mathrm{s}} \geq 0$ are regularization parameters and different sparseness measures could be used, e.g., $f(S_{jk}) = |S_{jk}|$, $f(S_{jk}) = S_{jk}$, $f(S_{jk}) = S_{jk}\ln(S_{jk})$, etc. Several extensions of NMF have been proposed. In [21], the nonnegative matrix partial co-factorization (NMPCF) was proposed for rhythmic source separation. Given the magnitude spectrogram as input data matrix X, NMPCF decomposes the music signal into a drum or rhythmic part and a residual or harmonic part $X \approx A_{\mathrm{r}}S_{\mathrm{r}} + A_{\mathrm{h}}S_{\mathrm{h}}$ with the factorized matrices including basis matrix and

weight matrix for rhythmic source $\{A_\mathrm{r}, S_\mathrm{r}\}$ and for harmonic source $\{A_\mathrm{h}, S_\mathrm{h}\}$. The prior knowledge from drum-only signal $Y \approx A_\mathrm{r} S_\mathrm{r}$ given the same rhythmic bases A_r is incorporated in joint minimization of two Euclidean error functions

$$\|X - A_\mathrm{r}S_\mathrm{r} - A_\mathrm{h}S_\mathrm{h}\|^2 + \eta\,\|Y - A_\mathrm{r}S_\mathrm{r}\|^2 \tag{2}$$

where η is a trade-off between the first and the second reconstruction errors due to X and Y, respectively. In [22], the mixed signals were divided into L segments. Each segment $X^{(l)}$ is decomposed into common and individual parts which reflect the rhythmic and harmonic sources, respectively. The common bases A_r are shared for different segments due to high temporal repeatability in rhythmic sources. The individual bases $A_\mathrm{h}^{(l)}$ are separate for individual segment l due to the changing frequency and low temporal repeatability. The resulting objective function consists of a weighted Euclidean error function and the regularization terms due to bases A_r and $A_\mathrm{h}^{(l)}$ which are expressed by

$$\sum_{l=1}^{L} \omega^{(l)} \|X^{(l)} - A_\mathrm{r}S_\mathrm{r}^{(l)} - A_\mathrm{h}^{(l)}S_\mathrm{h}^{(l)}\|^2 + \eta L \|A_\mathrm{r}\|^2 + \eta \sum_{l=1}^{L} \|A_\mathrm{h}^{(l)}\|^2 \tag{3}$$

where $\{\omega^{(l)}, S_\mathrm{r}^{(l)}, S_\mathrm{h}^{(l)}\}$ denotes the segment-dependent weights and weight matrices for common basis and individual basis, respectively. This is a NMPCF for L segments. The solutions to these NMFs are derived and implemented by the multiplicative update rules so that nonnegative constraints are met for individual model parameters. For example, the terms in gradient of objective function $\mathcal{F}$ with respect to nonnegative parameter A are divided into positive terms and negative terms $\frac{\partial \mathcal{F}}{\partial A} = \left[\frac{\partial \mathcal{F}}{\partial A}\right]^+ - \left[\frac{\partial \mathcal{F}}{\partial A}\right]^-$ where $\left[\frac{\partial \mathcal{F}}{\partial A}\right]^+ > 0$ and $\left[\frac{\partial \mathcal{F}}{\partial A}\right]^- > 0$. The multiplicative update rule is yielded by

$$A \leftarrow A \otimes \left[\frac{\partial \mathcal{F}}{\partial A}\right]^- \oslash \left[\frac{\partial \mathcal{F}}{\partial A}\right]^+ \tag{4}$$

where $\otimes$ and $\oslash$ denote element-wise multiplication and division, respectively.

2.2 Group basis representation

The signal reconstruction methods in (2) and (3) correspond to the group basis representation where two groups of bases A_r and $A_\mathrm{h}^{(l)}$ are applied. The separation of single-channel mixed signal into two source signals is achieved. The issue of underdetermined source separation is resolved. In [11], the group-based NMF (GNMF) was developed by conducting group analysis and constructing two groups of bases. The intra-subject variations for a subject in different trials and the inter-subject variations for different subjects could be compensated. Given the L subjects or segments, the lth segment is generated by $X^{(l)} \approx A_\mathrm{r}^{(l)}S_\mathrm{r}^{(l)} + A_\mathrm{h}^{(l)}S_\mathrm{h}^{(l)}$ where $A_\mathrm{r}^{(l)}$ denotes the common bases which capture the intra and inter-subject variations and $A_\mathrm{h}^{(l)}$ denotes the individual bases which reflect the residual information. In general, different common bases $A_\mathrm{r}^{(l)}$ should be close together since these bases represent the shared information in mixed signal. Contrarily, individual bases $A_\mathrm{h}^{(l)}$ characterize individual features which should be discriminated and mutually far apart [11]. The object function of GNMF is formed by

$$\begin{aligned}
&\sum_{l=1}^{L} \|X^{(l)} - A_\mathrm{r}^{(l)}S_\mathrm{r}^{(l)} - A_\mathrm{h}^{(l)}S_\mathrm{h}^{(l)}\|^2 + \eta_\mathrm{a} \sum_{l=1}^{L} \|A_\mathrm{r}^{(l)}\|^2 \\
&\quad + \eta_\mathrm{a} \sum_{l=1}^{L} \|A_\mathrm{h}^{(l)}\|^2 \\
&\quad + \eta_{\mathrm{a_r}} \sum_{l=1}^{L}\sum_{m=1}^{L} \|A_\mathrm{r}^{(l)} - A_\mathrm{r}^{(m)}\|^2 \\
&\quad - \eta_{\mathrm{a_h}} \sum_{l=1}^{L}\sum_{m=1}^{L} \|A_\mathrm{h}^{(l)} - A_\mathrm{h}^{(m)}\|^2.
\end{aligned} \tag{5}$$

In (5), the second and third terms are seen as the ℓ_2 regularization functions, the fourth term enforces the distance between different common bases to be small, and the fifth term enforces the distance between different individual bases to be large. Regularization parameters $\{\eta_\mathrm{a}, \eta_{\mathrm{a_r}}, \eta_{\mathrm{a_h}}\}$ are used. The NMPCFs in [21,22] and GNMF in [11] did not consider sparsity in group basis representation.

More generally, a group sparse coding algorithm [23] was proposed for basis representation of group instances $\{X_k, k \in \mathcal{G}\}$ where objective function is defined by

$$\sum_{k\in\mathcal{G}} \left\| X_k - \sum_{j=1}^{|D|} S_j^k A_j \right\|^2 + \eta \sum_{j=1}^{|D|} \|S_j\|. \tag{6}$$

All the instances within a group $\mathcal{G}$ share the same dictionary D with basis vectors $\{A_j\}_{j=1}^{|D|}$. The weight matrix $\{S_j\}_{j=1}^{|D|}$ consists of nonnegative vectors $S_j = [\,S_j^1, \ldots, S_j^{|\mathcal{G}|}]^T$. The weight parameters $\{S_j^k\}$ are estimated for different group instances $k \in \mathcal{G}$ using different bases $j \in \mathcal{D}$. In (6), ℓ_1 regularization term is incorporated to carry out group sparse coding. The group sparsity was further extended to structural sparsity for dictionary learning and basis representation. Nevertheless, nonnegative constraints were not imposed on bases $\{A_j\}$ and observed signals $\{X_k\}$. Basically, all the above-mentioned methods [2,11,21-24] did not apply probabilistic framework. No Bayesian learning was considered.

2.3 Bayesian learning approaches

Model regularization is critical for improving the generalization of a learning machine to new data [7]. Conducting Bayesian learning shall compensate the variations of the estimated parameters and accordingly improve model regularization. Typically, NMF and group basis representation are viewed as learning machine which is based on a set of bases. Following the perspective of relevance vector machines [8,9], Bayesian sparse learning is beneficial to identify relevant bases for regularized basis representation. To do so, sparse priors based on Student t distribution [17] and Laplacian distribution [10,25] could act as regularization functions and merged with likelihood function to come up with *a posteriori* probability. Maximizing the logarithm of *a posteriori* probability is equivalent to minimizing the ℓ_1-regularized error function if Laplacian prior is applied. Hyperparameters of sparse priors are then used as the regularization parameter which controls the trade-off between a reconstruction error function and a sparsity-favorable penalty function.

In the literature, a probabilistic matrix factorization (PMF) [26] for $X = A^{\mathrm{T}}S$ was proposed by assuming Gaussian noise for each independent entry of data matrix $X = \{X_{ik}\}$ by $p(X|A,S,\alpha) = \prod_{i=1}^{N}\prod_{k=1}^{M}\mathcal{N}(X_{ik}|A_i^T S_k, \alpha^{-1})$ and assuming Gaussian priors $p(A|\alpha_{\mathrm{a}}) = \prod_{i=1}^{N}\mathcal{N}(A_i|0,\alpha_{\mathrm{a}}^{-1}I)$ and $p(S|\alpha_{\mathrm{s}}) = \prod_{k=1}^{M}\mathcal{N}(S_k|0,\alpha_{\mathrm{s}}^{-1}I)$ where $\{\alpha,\alpha_{\mathrm{a}},\alpha_{\mathrm{s}}\}$ is a set of precision parameters of Gaussians. Here, A_i denotes the ith column of A and S_k denotes the kth column of S. Learning for PMF is equivalent to maximizing the log posterior likelihood

$$\ln p(A,S|X,\alpha,\alpha_{\mathrm{a}},\alpha_{\mathrm{s}}) = \ln p(X|A,S,\alpha) + \ln p(A|\alpha_{\mathrm{a}}) + \ln p(S|\alpha_{\mathrm{s}}) + C \tag{7}$$

with respect to A and S. In (7), C is a constant. This optimization turns out to minimizing the sum-of-squares error function with quadratic regularization terms

$$\sum_{i=1}^{N}\sum_{k=1}^{M}(X_{ik}-A_i^T S_k)^2+\eta_{\mathrm{a}}\sum_{i=1}^{N}\|A_i\|^2+\eta_{\mathrm{s}}\sum_{k=1}^{M}\|S_k\|^2. \tag{8}$$

The regularization terms are determined from hyperparameters by $\eta_{\mathrm{a}} = \alpha_{\mathrm{a}}/\alpha$ and $\eta_{\mathrm{s}} = \alpha_{\mathrm{s}}/\alpha$. Bayesian learning of PMF was performed through MCMC algorithm where Gaussian-Wishart priors for Gaussian mean vectors and precision matrices were assumed. There was no constraint on nonnegative matrices by using PMF. No sparse learning was considered.

In [27], a full Bayesian NMF was implemented to determine the number of bases according to the marginal likelihood. Furthermore, Bayesian nonparametric approach to NMF was proposed in [28] where model structure was determined through Gamma process NMF. This method was applied to find both latent sources in spectrograms and their number. In [25], the group sparse coding [23] was upgraded with Bayesian interpretation. Bayesian sparse learning was only developed for single-sample basis representation. In [29], the group sparse priors were presented for maximum *a posteriori* estimation of covariance matrix which was used in Gaussian graphical model. More recently, the group sparse hidden Markov models (HMMs) [30] were proposed to represent a sequence of observations and have been successfully applied for speech recognition. A set of common bases were shared for representation of speech samples across HMM states, while a set of individual bases were employed to represent speech samples within individual HMM states. Bayesian group sparse learning was performed for speech recognition [30] and signal separation [20] by using Laplacian scale mixture distribution.

3 Bayesian group sparse matrix factorization

Previous NMF methods [11,13,21] were developed to extract task-specific nonnegative factors, but they did not simultaneously consider the *uncertainty* of model parameters and control the *sparsity* of weight parameters. In [23,25], the group sparse coding and its Bayesian extension did not impose *nonnegative* constraints in data matrix X and factorized matrices A and S. This paper presents a new Bayesian group sparse learning for NMF (denoted by BGS-NMF) and applied it for single-channel music source separation.

3.1 Model construction

In this study, magnitude spectrogram $X = \{X^{(l)}\}$ of a mixed audio signal is calculated and chopped into L segments for implementation of BGS-NMF algorithm. The audio signal is assumed to be mixed from two kinds of source signals. One is rhythmic or repetitive source signal and the other is harmonic or residual source signal. As illustrated in Figure 1, BGS-NMF aims to decompose a nonnegative matrix $X^{(l)} \in \mathcal{R}_+^{N\times M}$ of the lth segment into a product of two nonnegative matrices $A^{(l)}S^{(l)}$. A linear decomposition model is constructed in a form of

$$X^{(l)} = A_{\mathrm{r}}S_{\mathrm{r}}^{(l)} + A_{\mathrm{h}}^{(l)}S_{\mathrm{h}}^{(l)} + E^{(l)} \tag{9}$$

where $A_{\mathrm{r}} \in \mathcal{R}_+^{N\times D_{\mathrm{r}}}$ denotes the shared basis matrix for all segments $\{X^{(l)}, l = 1,\ldots,L\}$; $A_h^{(l)} \in \mathcal{R}_+^{N\times D_h}$ and $E^{(l)}$ denotes the individual matrix and the noise matrix for a given segment l, respectively. Typically, common bases capture the repetitive patterns which continuously happen in different segments of a whole signal. Individual bases are used to compensate the residual information that common bases could not handle. Without loss of generality, common bases and individual bases are applied

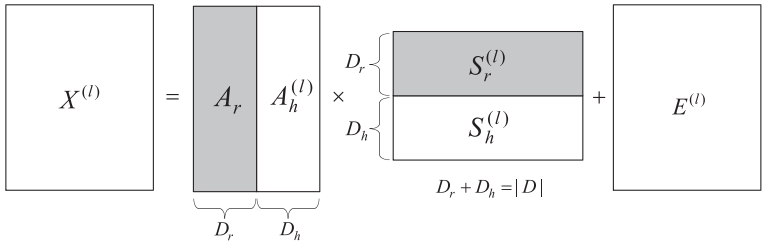

Figure 1 Illustration for group basis representation. There are $|D|$ bases in the dictionary.

to recover the rhythmic signal and the harmonic signal, respectively, from a mixed audio signal. Such a signal recovery problem could be interpreted from a perspective of subspace approach. Namely, an observed signal is demixed into one signal from *principal subspace* spanned by common bases and the other signal from *minor subspace* spanned by individual bases [31]. Moreover, the sparseness constraint is imposed on two groups of reconstruction weights $S_r^{(l)} \in \mathcal{R}_+^{D_r \times M}$ and $S_h^{(l)} \in \mathcal{R}_+^{D_h \times M}$. It is assumed that the reconstruction weights of rhythmic sources $S_r^{(l)}$ and harmonic sources $S_h^{(l)}$ are independent, but the dependencies between reconstruction weights within each group are allowed. Assuming that the kth noise vector $E_k^{(l)}$ is Gaussian distributed with zero mean and $N \times N$ diagonal covariance matrix $\Sigma^{(l)} = \text{diag}\{[\,\Sigma^{(l)}]_{ii}\}$ which is shared for all samples within a segment l, the likelihood function of an audio signal segment $X^{(l)}$ is expressed by

$$p(X^{(l)}|\Theta^{(l)}) = \prod_{i=1}^{N}\prod_{k=1}^{M} \mathcal{N}(X_{ik}^{(l)} \,|\, [A_{\mathrm{r}} S_{\mathrm{r}}^{(l)}]_{ik} + [A_{\mathrm{h}}^{(l)} S_{\mathrm{h}}^{(l)}]_{ik}, [\,\Sigma^{(l)}]_{ii}\,). \tag{10}$$

BGS-NMF model is therefore constructed with parameters $\Theta^{(l)} = \{A_{\mathrm{r}}, A_{\mathrm{h}}^{(l)}, S_{\mathrm{r}}^{(l)}, S_{\mathrm{h}}^{(l)}, \Sigma^{(l)}\}$.

3.2 Priors for Bayesian group sparse learning

From Bayesian perspective, the uncertainties of BGS-NMF parameters, expressed by prior densities, are considered to assure *model regularization*. Using BGS-NMF model, the common bases A_{r} are constructed to represent the characteristics of repetitive patterns for different data segments, while the individual bases $A_{\mathrm{h}}^{(l)}$ are estimated to reflect unique information in each segment l. Sparsity control is enforced in the corresponding reconstruction weights $S_{\mathrm{r}}^{(l)}$ and $S_{\mathrm{h}}^{(l)}$ so that relevant bases are retrieved for group basis representation. In accordance with [15], the nonnegative basis parameters are assumed to be gamma distributed by

$$p(A_{\mathrm{r}}) = \prod_{i=1}^{N}\prod_{j=1}^{D_{\mathrm{r}}} \mathcal{G}([A_{\mathrm{r}}]_{ij} \,|\alpha_{rj}, \beta_{rj}) \tag{11}$$

$$p(A_{\mathrm{h}}^{(l)}) = \prod_{i=1}^{N}\prod_{j=1}^{D_{\mathrm{h}}} \mathcal{G}([A_{\mathrm{h}}^{(l)}]_{ij} \,|\alpha_{hj}^{(l)}, \beta_{hj}^{(l)}) \tag{12}$$

where $\Phi_{\mathrm{a}}^{(l)} = \{\{\alpha_{rj}, \beta_{rj}\}, \{\alpha_{hj}^{(l)}, \beta_{hj}^{(l)}\}\}$ denotes the hyperparameters of gamma distributions and $\{D_{\mathrm{r}}, D_{\mathrm{h}}\}$ denote the numbers of common bases and individual bases, respectively. Gamma distribution is an exponential family distribution for *nonnegative data*. Its two parameters $\{\alpha, \beta\}$ can be adjusted to fit different shapes of distributions. In (11) and (12), all entries in matrices A_{r} and $A_{\mathrm{h}}^{(l)}$ are assumed to be independent.

Importantly, we control the sparsity of reconstruction weights by using prior density based on the *Laplacian scale mixture* (LSM) distribution [25]. The LSM of a reconstruction weight of common basis is constructed by $[\,S_{\mathrm{r}}^{(l)}]_{jk} = (\lambda_{rj}^{(l)})^{-1} u_{rj}^{(l)}$ where $u_{rj}^{(l)}$ is a Laplacian distribution $p(u_{rj}^{(l)}) = \frac{1}{2}\exp\{-|u_{rj}^{(l)}|\}$ with scale 1 and $\lambda_{rj}^{(l)}$ is an inverse scale parameter. Accordingly, the parameter $[\,S_{\mathrm{r}}^{(l)}]_{jk}$ has a Laplacian distribution

$$p([\,S_{\mathrm{r}}^{(l)}]_{jk} \,|\lambda_{rj}^{(l)}) = \frac{\lambda_{rj}^{(l)}}{2} \exp\{-\lambda_{rj}^{(l)} [\,S_{\mathrm{r}}^{(l)}]_{jk}\} \tag{13}$$

which is controlled by a positive continuous mixture parameter $\lambda_{rj}^{(l)} \geq 0$. Considering a gamma distribution for inverse scale parameter, i.e., $p(\lambda_{rj}^{(l)}) = \mathcal{G}(\lambda_{rj}^{(l)} | \gamma_{rj}^{(l)}, \delta_{rj}^{(l)})$, the marginal distribution of a reconstruction weight can be calculated by [25]

$$\begin{aligned} p([\,S_r^{(l)}]_{jk}) &= \int_0^{\infty} p([\,S_r^{(l)}]_{jk} \,|\lambda_{rj}^{(l)}) p(\lambda_{rj}^{(l)}) d\lambda_{rj}^{(l)} \\ &= \frac{\gamma_{rj}^{(l)} (\delta_{rj}^{(l)})^{\gamma_{rj}^{(l)}}}{2(\delta_{rj}^{(l)} + [\,S_r^{(l)}]_{jk})^{\gamma_{rj}^{(l)}+1}}. \end{aligned} \tag{14}$$

In (13) and (14), the constraint $[S_r^{(l)}]_{jk} \geq 0$ has been considered. This LSM distribution is obtained by adopting the property that gamma distribution is the *conjugate prior* for Laplacian distribution. In application of image coding, LSM distribution was estimated and measured to be sparser than Laplacian distribution by approximately a factor of 2 [25]. Figure 2 compares Gaussian, Laplacian, and LSM distributions with specific parameters. In this example, LSM is the sharpest distribution among these distributions. In addition, a truncated LSM prior for nonnegative parameter $[S_r^{(l)}]_{jk} \in \mathcal{R}_+$ is adopted, namely, the distribution of negative parameter is forced to be zero. The sparse prior for reconstruction weight for individual basis $[S_h^{(l)}]_{jk}$ is also expressed by LSM distribution with hyperparameter $\{\gamma_{hj}^{(l)}, \delta_{hj}^{(l)}\}$. The hyperparameters of BGS-NMF is formed by $\Phi^{(l)} = \{\Phi_a^{(l)}, \Phi_s^{(l)} = \{\gamma_{rj}^{(l)}, \delta_{rj}^{(l)}, \gamma_{hj}^{(l)}, \delta_{hj}^{(l)}\}\}$. Figure 3 displays a graphical representation for construction of BGS-NMF with different parameters $\Theta^{(l)}$ and hyperparameters $\Phi^{(l)}$.

By combining the likelihood function in (10) and the prior densities in (11) to (13), the negative logarithm of posterior distribution $-\ln p(A_{\mathrm{r}}, A_{\mathrm{h}}^{(l)}, S_{\mathrm{r}}^{(l)}, S_{\mathrm{h}}^{(l)}|X)$ can be calculated and arranged as a new objective function expressed by

$$\begin{aligned}
&\sum_{l=1}^{L}\sum_{i=1}^{N}\sum_{k=1}^{M}(X_{ik}^{(l)} - [A_{\mathrm{r}}S_{\mathrm{r}}^{(l)}]_{ik} - [A_{\mathrm{h}}^{(l)}S_{\mathrm{h}}^{(l)}]_{ik})^2 \\
&\quad + \eta_a L \sum_{i=1}^{N}\sum_{j=1}^{D_{\mathrm{r}}}((1-\alpha_{rj})\ln[A_{\mathrm{r}}]_{ij} \\
&\quad + \beta_{rj}[A_{\mathrm{r}}]_{ij}) \\
&\quad + \eta_{\mathrm{a}} \sum_{l=1}^{L}\sum_{i=1}^{N}\sum_{j=1}^{D_{\mathrm{h}}}((1-\alpha_{hj}^{(l)})\ln[A_{\mathrm{h}}^{(l)}]_{ij} \\
&\quad + \beta_{hj}^{(l)}[A_{\mathrm{h}}^{(l)}]_{ij}) + \eta_{\mathrm{s_r}} \sum_{l=1}^{L}\sum_{j=1}^{D_{\mathrm{r}}}\sum_{k=1}^{M}[S_{\mathrm{r}}^{(l)}]_{jk} \\
&\quad + \eta_{\mathrm{s_h}} \sum_{l=1}^{L}\sum_{j=1}^{D_{\mathrm{h}}}\sum_{k=1}^{M}[S_{\mathrm{h}}^{(l)}]_{jk}
\end{aligned} \tag{15}$$

where $\{\eta_{\mathrm{a}}, \eta_{\mathrm{s_r}}, \eta_{\mathrm{s_h}}\}$ denote the regularization parameters for two groups of bases and reconstruction weights. Some BGS-NMF parameters or hyperparameters have been absorbed in these regularization parameters. Comparing with the objective functions (3) for NMPCF, (5) for GNMF, and (8) for PMF, the optimization of (15) for BGS-NMF shall lead to two groups of signals which are reconstructed from the sparse common bases A_{r} and sparse individual bases $A_{\mathrm{h}}^{(l)}$. The regularization terms due to two gamma bases are additionally considered. Different

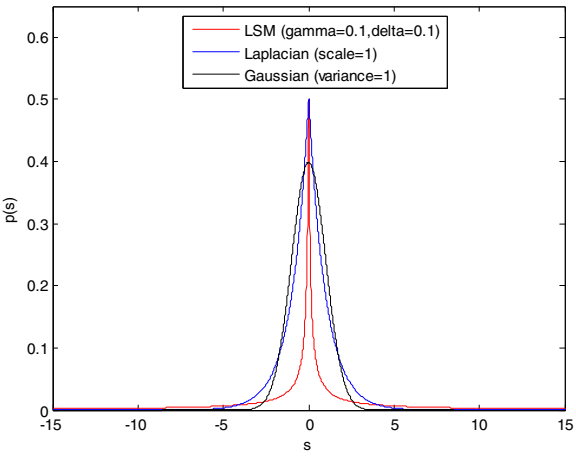

Figure 2 Comparison of Gaussian, Laplacian, and LSM distributions.

from the Bayesian NMF (BNMF) [15], BGS-NMF conducts group sparse learning which does not only characterize the within-segment harmonic information but also represent the across-segment rhythmic regularity. Sparse sets of basis vectors are further determined for sparse representation. Basically, BGS-NMF follows a general objective function. By applying different hyperparameter values $\{\alpha_{rj}, \beta_{rj}, \alpha_{hj}^{(l)}, \beta_{hj}^{(l)}\}$, probability structures, and prior distributions for $\{A_{\mathrm{r}}, A_{\mathrm{h}}^{(l)}, S_{\mathrm{r}}^{(l)}, S_{\mathrm{h}}^{(l)}\}$, BGS-NMF can be realized to find solutions to NMF [2], NMPCF [21], GNMF [11], PMF [26], and BNMF [15]. Notably, the objective function in (15) is written for comparative study among different methods. This function only considers BGS-NMF

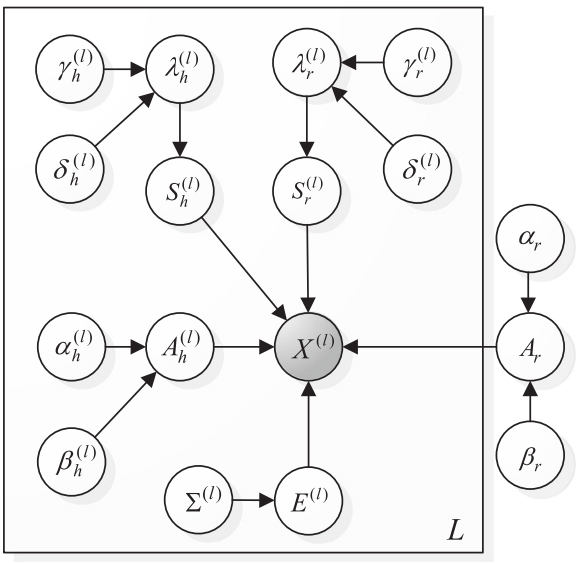

Figure 3 A graphical representation for BGS-NMF.

based on Laplacian prior. BGS-NMF algorithms with Laplacian prior and LSM prior shall be both implemented in the experiments. Nevertheless, in what follows, we address the model inference procedure for *BGS-NMF with LSM prior*.

3.3 Model inference

The full Bayesian framework for BGS-NMF model based on the posterior distribution of parameters and hyperparameters $p(\Theta, \Phi|X)$ is not analytically tractable. A stochastic optimization scheme is adopted. We develop a MCMC sampling algorithm for approximate inference through iteratively generating samples of parameters Θ and hyperparameters Φ according to the posterior distribution. This algorithm converges by those samples. The key idea of MCMC sampling is to simulate a stationary ergodic Markov chain whose samples asymptotically follow the posterior distribution $p(\Theta, \Phi|X)$. The estimates of parameters Θ and hyperparameters Φ are then computed via Monte Carlo integrations on the simulated Markov chains. For simplicity, the segment index l is neglected in derivation of MCMC algorithm for BGS-NMF. At each new iteration $t+1$, the BGS-NMF parameters $\Theta^{(t+1)}$ and hyperparameters $\Phi^{(t+1)}$ are sequentially sampled in an order of $\{A_\mathrm{r}, S_\mathrm{r}, A_\mathrm{h}, S_\mathrm{h}, \Sigma, \alpha_\mathrm{r}, \beta_\mathrm{r}, \alpha_\mathrm{h}, \beta_\mathrm{h}, \lambda_\mathrm{r}, \lambda_\mathrm{h}, \gamma_\mathrm{r}, \delta_\mathrm{r}, \gamma_\mathrm{h}, \delta_\mathrm{h}\}$ according to their corresponding conditional posterior distributions. In this subsection, we describe the calculation of conditional posterior distributions under BGS-NMF parameters $\{A_\mathrm{r}, S_\mathrm{r}, A_\mathrm{h}, S_\mathrm{h}, \Sigma\}$. The conditional posterior distributions for hyperparameters $\{\alpha_\mathrm{r}, \beta_\mathrm{r}, \alpha_\mathrm{h}, \beta_\mathrm{h}, \lambda_\mathrm{r}, \lambda_\mathrm{h}, \gamma_\mathrm{r}, \delta_\mathrm{r}, \gamma_\mathrm{h}, \delta_\mathrm{h}\}$ are derived in the Appendix.

1. Sampling of $[A_\mathrm{r}]_{ij}$. First of all, the common basis parameter $[A_\mathrm{r}^{(t+1)}]_{ij}$ is sampled by the conditional posterior distribution

$$p([A_\mathrm{r}]_{ij}|X_i^T, \Theta_{A_{rij}}^{(t)}, \Phi_{A_{rij}}^{(t)}) \propto p(X_i^T|\Theta_{A_{rij}}^{(t)}) p([A_\mathrm{r}]_{ij}|\Phi_{A_{rij}}^{(t)}) \tag{16}$$

where $\Theta_{A_{rij}}^{(t)} = \{[A_\mathrm{r}^{(t+1)}]_{i(1:j-1)}, [A_\mathrm{r}^{(t)}]_{i(j+1:D_\mathrm{r})}, S_\mathrm{r}^{(t)}, A_\mathrm{h}^{(t)}, S_\mathrm{h}^{(t)}, \Sigma^{(t)}\}$ and $\Phi_{A_{rij}}^{(t)} = \{\alpha_{rj}^{(t)}, \beta_{rj}^{(t)}\}$. Here, X_i denotes the ith row vector of X. Notably, for each sampling, we use the preceding bases $[A_\mathrm{r}^{(t+1)}]_{i(1:j-1)}$ at new iteration $t+1$ and subsequent bases $[A_\mathrm{r}]_{i(j+1:D_\mathrm{r})}^{(t)}$ at current iteration t. The likelihood function can be arranged as a Gaussian distribution of $[A_\mathrm{r}]_{ij}$

$$p(X_i^T|\Theta_{A_{rij}}^{(t)}) \propto \exp\left\{-\frac{([A_\mathrm{r}]_{ij} - \mu_{A_{rij}}^{\mathrm{likel}})^2}{2[\sigma_{A_{rij}}^{\mathrm{likel}}]^2}\right\} \tag{17}$$

where $\mu_{A_{rij}}^{\mathrm{likel}} = [\sigma_{A_{rij}}^{\mathrm{likel}}]^{-2} \sum_{k=1}^{M} ([S_\mathrm{r}^{(t)}]_{jk}\, \varepsilon_{ik}^{(-j)})$, $\varepsilon_{ik}^{(-j)} = X_{ik} - (\sum_{m=1}^{j-1} [A_\mathrm{r}^{(t+1)}]_{im} [S_\mathrm{r}^{(t)}]_{mk} + \sum_{m=j+1}^{D_\mathrm{r}} [A_\mathrm{r}^{(t)}]_{im} [S_\mathrm{r}^{(t)}]_{mk}) - \sum_{m=1}^{D_\mathrm{h}} [A_\mathrm{h}^{(t)}]_{im} [S_\mathrm{h}^{(t)}]_{mk}$ and $[\sigma_{A_{rij}}^{\mathrm{likel}}]^2 = [\Sigma^{(t)}]_{ii} (\sum_{k=1}^{M} [S_\mathrm{r}^{(t)}]_{jk})^{-1}$. By combining likelihood function of (17) and gamma prior $p([A_\mathrm{r}]_{ij}|\Phi_{A_{rij}}^{(t)})$ of (11), the conditional posterior distribution in (16) is derived in a form of

$$[A_r]_{ij}^{\alpha_{rj}^{(t)}-1} \exp\left\{-\frac{([A_r]_{ij} - \mu_{A_{rij}}^{\mathrm{post}})^2}{2[\sigma_{A_{rij}}^{\mathrm{post}}]^2}\right\} \mathbb{I}_{[0,+\infty[}([A_r]_{ij}) \tag{18}$$

where $\mu_{A_{rij}}^{\mathrm{post}} = \mu_{A_{rij}}^{\mathrm{likel}} - \beta_{rj}^{(t)} [\sigma_{A_{rij}}^{\mathrm{likel}}]^2$, $[\sigma_{A_{rij}}^{\mathrm{post}}]^2 = [\sigma_{A_{rij}}^{\mathrm{likel}}]^2$, and $\mathbb{I}_{[0,+\infty[}(z)$ denotes an indicator function which has value either 1 if $z \in [0, +\infty[$ or 0 for the other case. In (18), the posterior distribution for negative $[A_\mathrm{r}]_{ij}$ is forced to be zero. Derivations of (17) and (18) are detailed in the Appendix. However, (18) is not an usual distribution, therefore its sampling requires the use of a *rejection sampling* method, such as the Metropolis-Hastings algorithm [32]. Using this algorithm, an *instrumental distribution* $q([A_\mathrm{r}]_{ij})$ is chosen to fit at best the target distribution (18) so that high rejection condition is avoided or equivalently rapid convergence toward true parameter could be achieved. In case of rejection, the previous parameter sample is used, namely, $[A_\mathrm{r}^{(t+1)}]_{ij} \leftarrow [A_\mathrm{r}^{(t)}]_{ij}$. Generally, the shape of target distribution is characterized by its mode and width. The instrumental distribution is constructed as a truncated Gaussian distribution which is calculated by

$$q([A_\mathrm{r}]_{ij}) = \mathcal{N}_+([A_\mathrm{r}]_{ij}|\mu_{A_{rij}}^{\mathrm{inst}}, [\sigma_{A_{rij}}^{\mathrm{inst}}]^2). \tag{19}$$

In (19), the mode $\mu_{A_{rij}}^{\mathrm{inst}}$ is obtained by finding the roots of a quadratic equation of $[A_\mathrm{r}]_{ij}$ which appears in the exponent of the posterior distribution in (18). Derivation for the mode $\mu_{A_{rij}}^{\mathrm{inst}}$ is detailed in the Appendix. In case of complex-valued root or negative-valued root, the mode is forced by $\mu_{A_{rij}}^{\mathrm{inst}} = 0$. The width of instrumental distribution is controlled by $[\sigma_{A_{rij}}^{\mathrm{inst}}]^2 = [\sigma_{A_{rij}}^{\mathrm{post}}]^2$.

2. Sampling of $[S_\mathrm{r}]_{jk}$. The sampling of reconstruction weight of common basis $[S_\mathrm{r}^{(t+1)}]_{jk}$ depends on the conditional posterior distribution

$$p([S_\mathrm{r}]_{jk}|X_k, \Theta_{S_{rjk}}^{(t)}, \Phi_{S_{rjk}}^{(t)}) \propto p(X_k|[S_\mathrm{r}]_{jk}, \Theta_{S_{rjk}}^{(t)}) \, p([S_\mathrm{r}]_{jk}|\Phi_{S_{rjk}}^{(t)}) \tag{20}$$

where $\Theta_{S_{rjk}}^{(t)} = \{A_\mathrm{r}^{(t+1)}, [S_\mathrm{r}^{(t+1)}]_{(1:j-1)k}, [S_\mathrm{r}^{(t)}]_{(j+1:D_\mathrm{r})k}, A_\mathrm{h}^{(t)}, S_\mathrm{h}^{(t)}, \Sigma^{(t)}\}$ and $\Phi_{S_{rjk}}^{(t)} = \lambda_{rj}^{(t)}$. X_k is the kth column of X. Again, the preceding weights $[S_\mathrm{r}^{(t+1)}]_{(1:j-1)k}$ at new iteration $t+1$ and subsequent weights $[S_\mathrm{r}^{(t)}]_{(j+1:D_\mathrm{r})k}$ at current

iteration tx are used. The likelihood function is rewritten as a Gaussian distribution of $[S_{\mathrm{r}}]_{jk}$ given by

$$p(X_k \,|[S_{\mathrm{r}}]_{jk}, \Theta^{(t)}_{S_{rjk}}) \propto \exp\left\{-\frac{([S_{\mathrm{r}}]_{jk}-\mu^{\mathrm{likel}}_{S_{rjk}})^2}{2[\sigma^{\mathrm{likel}}_{S_{rjk}}]^2}\right\}. \quad (21)$$

The Gaussian parameters are obtained by $\mu^{\mathrm{likel}}_{S_{rjk}} = [\sigma^{\mathrm{likel}}_{S_{rjk}}]^{-2}\sum_{i=1}^{N}([\Sigma^{(t)}]^{-1}_{ii}[A^{(t+1)}_{\mathrm{r}}]_{ij}\,\varepsilon^{(-j)}_{ik})$, $\varepsilon^{(-j)}_{ik} = X_{ik} - (\sum_{m=1}^{j-1}[A^{(t+1)}_{\mathrm{r}}]_{im}[S^{(t+1)}_{\mathrm{r}}]_{mk} + \sum_{m=j+1}^{D_{\mathrm{r}}}[A^{(t+1)}_{\mathrm{r}}]_{im}[S^{(t)}_{\mathrm{r}}]_{mk}) - \sum_{m=1}^{D_{\mathrm{h}}}[A^{(t)}_{h}]_{im}[S^{(t)}_{\mathrm{h}}]_{mk}$ and $[\sigma^{\mathrm{likel}}_{S_{rjk}}]^2 = (\sum_{i=1}^{N}[\Sigma^{(t)}]^{-1}_{ii}([A^{(t+1)}_{\mathrm{r}}]_{ij})^2)^{-1}$. Given the Gaussian likelihood and Laplacian prior, the conditional posterior distribution is calculated by

$$\lambda^{(t)}_{rj}\exp\left\{-\frac{([S_{\mathrm{r}}]_{jk}-\mu^{\mathrm{post}}_{S_{rjk}})^2}{2[\sigma^{\mathrm{post}}_{S_{rjk}}]^2}\right\}\mathbb{I}_{[0,+\infty[}([S_{\mathrm{r}}]_{jk}) \quad (22)$$

where $\mu^{\mathrm{post}}_{S_{rjk}} = \mu^{\mathrm{likel}}_{S_{rjk}} - \lambda^{(t)}_{rj}[\sigma^{\mathrm{likel}}_{S_{rjk}}]^2$ and $[\sigma^{\mathrm{post}}_{S_{rjk}}]^2 = [\sigma^{\mathrm{likel}}_{S_{rjk}}]^2$. Notably, the hyperparameters $\{\gamma^{(t+1)}_{rj}, \delta^{(t+1)}_{rj}\}$ in LSM prior are also sampled and used to sample LSM parameter $\lambda^{(t+1)}_{rj}$ based on a gamma distribution. Here, Metropolis-Hastings algorithm is applied again. The best instrumental distribution $q([S_r]_{jk})$ is selected to fit (22). This distribution is derived as a truncated Gaussian distribution $\mathcal{N}_+([S_r]_{jk}\,|\mu^{\mathrm{inst}}_{S_{rjk}}, [\sigma^{\mathrm{inst}}_{S_{rjk}}]^2)$ where the mode $\mu^{\mathrm{inst}}_{S_{rjk}}$ is derived by finding the root of a quadratic equation of $[S_{\mathrm{r}}]_{jk}$ and the width is obtained by $[\sigma^{\mathrm{inst}}_{S_{rjk}}]^2 = [\sigma^{\mathrm{post}}_{S_{rjk}}]^2$. In addition, the conditional posterior distributions for sampling the individual basis parameter $[A^{(t+1)}_{\mathrm{h}}]_{ij}$ and its reconstruction weight $[S^{(t+1)}_{\mathrm{h}}]_{jk}$ are similar to those for sampling $[A^{(t+1)}_{\mathrm{r}}]_{ij}$ and $[S^{(t+1)}_{\mathrm{r}}]_{jk}$, respectively. We do not address these two distributions.

3. Sampling of $[\Sigma]^{-1}_{ii}$. The sampling of the inverse of noise variance $([\Sigma]^{(t+1)}_{ii})^{-1}$ is performed according to the conditional posterior distribution

$$p([\Sigma]^{-1}_{ii}|X^T_i,\Theta^{(t)}_{\Sigma_{ii}},\Phi^{(t)}_{\Sigma_{ii}}) \propto p(X^T_i|[\Sigma]^{-1}_{ii},\Theta^{(t)}_{\Sigma_{ii}})p([\Sigma]^{-1}_{ii}|\Phi^{(t)}_{\Sigma_{ii}}) \quad (23)$$

where $\Theta^{(t)}_{\Sigma_{ii}} = \{A^{(t+1)}_{\mathrm{r}}, S^{(t+1)}_{\mathrm{r}}, A^{(t+1)}_{\mathrm{h}}, S^{(t+1)}_{\mathrm{h}}\}$ and $p([\Sigma]^{-1}_{ii}\,|\Phi^{(t)}_{\Sigma_{ii}}) = \mathcal{G}([\Sigma]^{-1}_{ii}\,|\alpha_{\Sigma_{ii}}, \beta_{\Sigma_{ii}})$. The resulting posterior distribution can be derived as a new gamma distribution with updated hyperparameters $\alpha^{\mathrm{post}}_{\Sigma_{ii}} = \frac{M}{2} + \alpha_{\Sigma_{ii}}$ and $\beta^{\mathrm{post}}_{\Sigma_{ii}} = \frac{1}{2}\sum_{k=1}^{M}(X_{ik} - \sum_{m=1}^{D_r}[A^{(t+1)}_{r}]_{im}[S^{(t+1)}_{r}]_{mk} - \sum_{m=1}^{D_h}[A^{(t+1)}_{h}]_{im}[S^{(t+1)}_{h}]_{mk})^2 + \beta_{\Sigma_{ii}}$. In the experiments, we conduct MCMC sampling procedure for $t_{\max}$ iterations. However, the first $t_{\min}$ iterations are not stable. These burn-in samples are abandoned. The marginal posterior estimates of common basis $[\hat{A}_{\mathrm{r}}]_{ij}$, individual basis $[\hat{A}_{\mathrm{h}}]_{ij}$ and their reconstruction weights $[\hat{S}_{\mathrm{r}}]_{jk}$ and $[\hat{S}_{\mathrm{h}}]_{jk}$ are calculated by finding the following sample means, e.g.,

$$[\hat{A}_{\mathrm{r}}]_{ij} = \frac{1}{t_{\max}-t_{\min}}\sum_{t=t_{\min}+1}^{t_{\max}}[A_{\mathrm{r}}]^{(t)}_{ij}. \quad (24)$$

With these posterior estimates, the rhythmic source and the harmonic source are calculated by $\hat{A}_{\mathrm{r}}\hat{S}_{\mathrm{r}}$ and $\hat{A}_{\mathrm{h}}\hat{S}_{\mathrm{h}}$, respectively. The BGS-NMF algorithm is completed. Different from BNMF [15], the proposed BGS-NMF conducts a *group sparse learning* based on *LSM distribution*. Common bases A_{r} are shared for different data segments l. The group sparse learning performs well in our experiments.

4 Experiments

In this study, BGS-NMF is implemented to estimate two audio source signals from a single-channel mixed signal. One source signal contains rhythmic pattern which is constructed by the bases shared for all audio segments while the other source contains harmonic information which is represented via bases from individual segments. Bayesian sparse learning is performed to conduct probabilistic reconstruction based on the relevant group bases. Some experiments are reported to evaluate the performance of model inference and signal reconstruction.

4.1 Experimental setup

In the experiments, we sampled six rhythmic signals and six harmonic signals from http://www.free-scores.com/index_uk.php3 and http://www.freesound.org/. Six mixed music signals were collected as follows: 'music 1,' bass+piano; 'music 2,' drum+guitar; 'music 3,' drum+violin; 'music 4,' cymbal+organ; 'music 5,' drum+saxophone; and 'music 6,' cymbal+singing, which contained combinations of different rhythmic and harmonic source signals. Three different drum signals and two different cymbal signals were included. For each set of experimental data, we applied a different mixing matrix music 1 (1.2667 − 1.9136), music 2 (1.1667 − 1.9136), music 3 (−1.2667 1.6136), music 4 (1.8667 1.1136), music 5 (−1.1667 2.8136), and music 6 (1.9617 1.1510) to simulate the corresponding single-channel mixed signal. Each audio signal was 21 s long. Readers may access http://chien.cm.nctu.edu.tw/bgs-nmf to listen to the twelve source signals and the corresponding six mixed signals.

The specification of 44,100-Hz sampling rate and 16-bit resolution was used in the collected audio signals. In our implementation, the magnitude of fast Fourier transform of audio signal was extracted every 1,024 samples with 512 samples in frame overlapping. Each mixed signal was equally chopped into L segments for music source separation. Each segment had a length of 3 s. Sufficient rhythmic signal existed within a segment. The numbers of common bases and individual bases were empirically set to be 15 and 10, respectively, i.e., $D_{\mathrm{r}} = 15$ and $D_{\mathrm{h}} = 10$. The common bases were sufficiently allocated so as to capture the shared base information from different segments. The initial common bases $A_{\mathrm{r}}^{(0)}$ and individual bases $A_{\mathrm{h}}^{(0)}$ were estimated by applying k-means clustering using the automatically detected rhythmic and harmonic segments, respectively. The detection was based on a classifier using Gaussian mixture model. We performed 1,000 Gibbs sampling iterations ($t_{\max} = 1,000$). The separation performance was evaluated according to the signal-to-interference ratio (SIR) in decibels

$$\mathrm{SIR\ (dB)} = 10\log_{10}\left[\frac{\sum_{l=1}^{L}\sum_{k=1}^{M}\|X_k^{(l)}\|^2}{\sum_{l=1}^{L}\sum_{k=1}^{M}\|\hat{X}_k^{(l)} - X_k^{(l)}\|^2}\right]. \tag{25}$$

The interference was measured by the Euclidean distance between original signal $\{X_k^{(l)}\}$ and reconstructed signal $\{\hat{X}_k^{(l)}\}$ for different samples k in different segments l. These signals include rhythmic signals $\{[\hat{A}_{\mathrm{r}}\hat{S}_{\mathrm{r}}^{(l)}]_k\}$ and harmonic signals $\{[\hat{A}_{\mathrm{h}}^{(l)}\hat{S}_{\mathrm{h}}^{(l)}]_k\}$.

For system initialization at $t = 0$, we detected two short segments with only rhythmic signal and harmonic signal and applied them for finding rhythmic parameters $\{A_{\mathrm{r}}^{(0)}, S_{\mathrm{r}}^{(0)}\}$ and harmonic parameters $\{A_{\mathrm{h}}^{(0)}, S_{\mathrm{h}}^{(0)}\}$, respectively. This prior information was used to implement five NMF methods for single-channel source separation. We carried out baseline NMF [2], Bayesian NMF (BNMF) [15], group-based NMF (GNMF) [11] (or NMPCF [22]), and the proposed BGS-NMF under consistent experimental conditions. To evaluate the effect of sparse priors in BGS-NMF for music source separation, we additionally realized BGS-NMF by applying Laplacian distribution. For this realization, the sampling steps of LSM parameters $\{\gamma_{r_j}, \delta_{r_j}, \gamma_{h_j}, \delta_{h_j}\}$ were ignored. The BGS-NMFs with Laplacian distribution (denoted by BGS-NMF-LP) and LSM distribution (BGS-NMF-LSM) were compared. All these NMFs were implemented for different segments l. Basically, the NMF model [2] was realized by using multiplicative updating algorithm in (4). The BNMF [15] conducted Bayesian learning of NMF model where MCMC sampling was performed, and gamma distributions were assumed for bases and reconstruction weights. No group sparse learning was considered in NMF and BNMF. Using NMPCF [22] or GNMF [11], the common bases and individual bases were constructed by applying multiplicative updating algorithm. No probabilistic framework was involved. The ℓ_2-norm regularization for basis parameters A_{r} and $A_{\mathrm{h}}^{(l)}$ was considered. There was no sparseness constraint imposed on reconstruction weight parameters $S_{\mathrm{r}}^{(l)}$ and $S_{\mathrm{h}}^{(l)}$. Only the result of GNMF method was reported. Using GNMF, the regularization parameters in (5) were empirically determined as $\{\eta_{\mathrm{a}} = 0.35, \eta_{\mathrm{a_r}} = 0.2, \eta_{\mathrm{a_h}} = 0.2\}$. Nevertheless, the Bayesian group sparse learning is presented in BGS-NMF-LP and BGS-NMF-LSM algorithms. Using this algorithm, the uncertainties of bases and reconstruction weights are represented by gamma distributions and LSM distributions, respectively. MCMC algorithm is developed to sample BGS-NMF parameters $\Theta^{(t+1)}$ and hyperparameters $\Phi^{(t+1)}$. The groups of common bases A_{r} and individual bases A_{h} are estimated to capture between-segment repetitive patterns and within-segment residual information, respectively. The relevant bases are detected via sparse priors in accordance with Laplacian or LSM distributions. Using BGS-NMF-LP, we sampled the parameters and hyperparameters by using different frames from six music signals and automatically calculated the averaged values of regularization parameters in (15) as $\{\eta_{\mathrm{a}} = 0.41, \eta_{\mathrm{s_r}} = 0.31, \eta_{\mathrm{s_h}} = 0.26\}$. The regularization parameters in (5) and (15) reflect different physical meanings in objective function. The computational cost and the model size are also examined. The computation times of running MATLAB codes were measured by a personal computer with Intel Core 2 Duo 2.4-GHz CPU and 4-GB RAM. In our investigation, the computation times of demixing an audio signal with 21 s long were measured as 3.1, 12.1, 16.2, 20.9, and 21.2 min by using NMF, BNMF, GNMF, and the proposed BGS-NMF-LP and BGS-NMF-LSM respectively. In addition, BNMF, GNMF, BGS-NMF-LP, and BGS-NMF-LSM were measured to be 2.5, 4.5, 5.2, and 5.3 times the model size of the baseline NMF respectively.

4.2 Evaluation for MCMC iterative procedure

In this set of experiments, the sampling process of BGS-NMF algorithm is evaluated. The control parameter of sparsity λ_{rj} and its hyperparameters γ_{rj} and δ_{rj} for common basis are investigated. Figure 4 displays an example of MCMC iterative sampling process for LSM parameter $\lambda_{rj}^{(t+1)}$. The value of samples converges after 200 iterations. Also, Figure 5 shows an example of iterative sampling process for LSM hyperparameters $\gamma_{rj}^{(t+1)}$ and $\delta_{rj}^{(t+1)}$. Convergence condition is good in these examples. MCMC samples converge after 200 iterations. Empirically, the parameter t_{min} is specified as 200 when

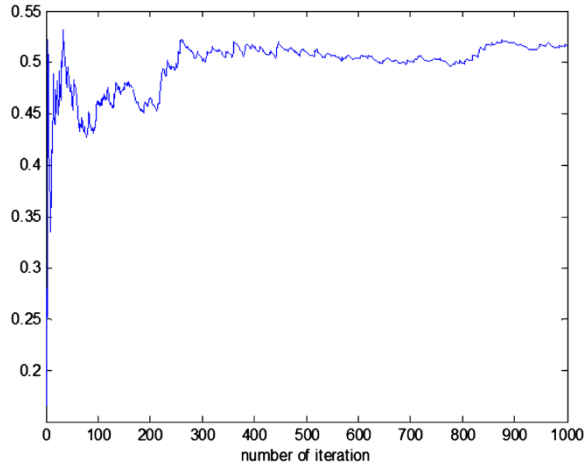

Figure 4 An example of iterative sampling process for LSM parameter $\lambda_{rj}^{(t+1)}$.

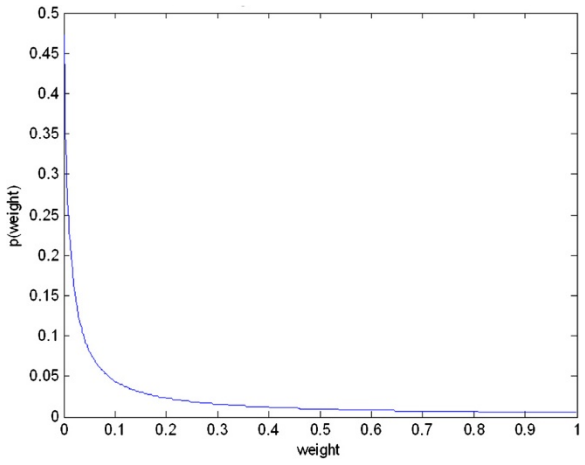

Figure 6 An estimated distribution of reconstruction weight of common basis $p([S_r]_{jk}|\gamma_{rj},\delta_{rj})$.

calculating posterior estimates of BGS-NMF parameters as given in (24). In addition, Figure 6 shows an estimated distribution of reconstruction weight of common basis $p([S_r]_{jk}|\gamma_{rj},\delta_{rj})$ where only nonnegative $[S_r]_{jk}$ is valid in the distribution. This distribution is shaped as a LSM distribution which is estimated from the 2^{nd} segment of "music 2".

4.3 Evaluation for single-channel music source separation

A quantitative comparison over different NMFs is conducted by measuring SIRs of reconstructed rhythmic signal and reconstructed harmonic signal. Table 1 shows the experimental results on six mixed music signals. These six signals come from twelve different source signals.

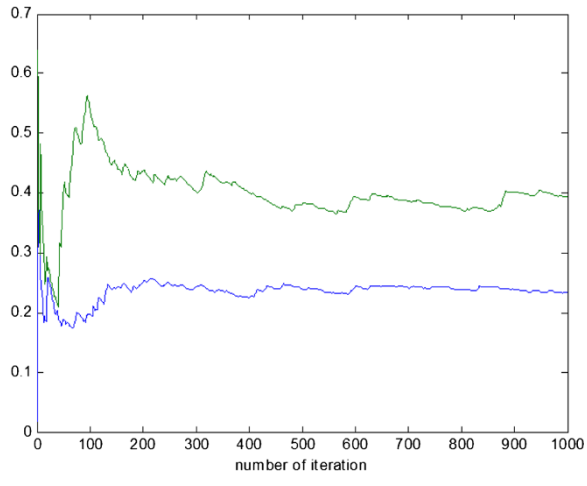

Figure 5 An example of iterative sampling process for LSM hyperparameters $\gamma_{rj}^{(t+1)}$ (green curve) and $\delta_{rj}^{(t+1)}$ (blue curve).

The averaged SIRs are reported in the last row. Comparing NMF and BNMF, we find that BNMF obtains higher SIRs on the reconstructed signals. Further, BNMF is more robust to different combination of rhythmic signals and harmonic signals. The variation of SIRs using NMF is relatively high. Bayesian learning provides model regularization for NMF. On the other hand, GNMF (or NMPCF) performs better than BNMF in terms of averaged SIR of the reconstructed signals. The key difference between BNMF and GNMF is the reconstruction of rhythmic signal. BNMF estimates the rhythmic bases for individual segments while GNMF (or NMPCF) calculates the shared rhythmic bases for different segments. Prior information $\{A_r^{(0)}, S_r^{(0)}, A_h^{(0)}, S_h^{(0)}\}$ is applied for these methods. From these results, we confirm the importance of basis grouping in signal reconstruction based on NMF. In particular, BGS-NMF-LP and BGS-NMF-LSM perform better than other NMF methods. BGS-NMF-LSM even outperforms BGS-NMF-LP in terms of SIRs. Reconstruction weights modeled by LSM distributions are better than those by Laplacian distributions. Sparser reconstruction weights identify fewer but more relevant basis vectors for signal separation. Nevertheless, among these five related NMFs, the highest SIRs of reconstructed signals are achieved by using BGS-NMF-LSM. The SIRs of reconstructed rhythmic and harmonic signals are measured as 8.13 dB and 8.40 dB which are higher than 3.71 dB and 3.38 dB by using NMF, 4.87 dB and 4.61 dB by using BNMF, 5.63 dB and 5.71 dB by using GNMF and 7.91 dB and 8.11 dB by using BGS-NMF-LP, respectively. Basically, the superiority of BGS-NMF-LSM to other NMFs is three-fold, i.e. *Bayesian probabilistic modeling*, *group basis representation* and *sparse reconstruction weight*. Again, compared to GNMF, the proposed BGS-NMF-LP and BGS-NMF-LSM

Table 1 Comparison of SIR (in dB) of the reconstructed rhythmic signal and harmonic signal based on NMF, BNMF, GNMF, BGS-NMF-LP and BGS-NMF-LSM

	NMF		BNMF		GNMF		BGS-NMF-LP		BGS-NMF-LSM	
	Rhythmic	Harmonic	Rhythmic	Harmonic	Rhythmic	Harmonic	Rhythmic	Harmonic	Rhythmic	Harmonic
Music 1	6.47	4.17	6.33	4.29	9.19	6.10	9.61	8.32	9.86	8.63
Music 2	6.30	1.10	8.08	5.18	8.22	3.03	8.33	7.13	8.55	7.45
Music 3	3.89	-1.11	5.16	3.80	6.01	3.22	8.44	8.52	8.63	8.79
Music 4	2.66	6.03	3.28	6.28	3.59	8.36	7.97	9.52	8.20	9.78
Music 5	1.85	3.71	3.03	2.55	3.97	6.44	8.11	8.22	8.35	8.50
Music 6	1.06	6.37	3.34	5.56	2.78	7.10	5.00	6.93	5.19	7.23
Average	*3.71*	*3.38*	*4.87*	*4.61*	*5.63*	*5.71*	*7.91*	*8.11*	*8.13*	*8.40*

Six mixed music signals are investigated.

obtain a more robust performance in SIRs against different music source signals. Figure 7 shows the waveforms of a drum signal, a saxophone signal and the resulting mixed signal in "music 5". Figure 8 displays the spectrograms of these three signals. Figure 9 demonstrates the spectrograms of the reconstructed drum signal and saxophone signal using BGS-NMF-LSM. For the other five mixed signals, the performance of reconstructed signals in single-channel music source separation is shown at http://chien.cm.nctu.edu.tw/bgs-nmf.

5 Conclusions

This paper has presented the Bayesian group sparse learning and applied it for single-channel nonnegative source separation. The basis vectors in NMF were grouped into two partitions. The first group was the common bases which were used to explore the inter-segment repetitive characteristics, while the second was the individual bases which were applied to represent the intra-segment harmonic information. The LSM distribution was introduced to express sparse reconstruction weights for two groups of basis vectors. Bayesian learning was incorporated into group basis representation with model regularization. The MCMC algorithm or the Metropolis-Hastings algorithm was developed to conduct approximate inference of model parameters and hyperparameters. Model parameters were used to find the decomposed rhythmic signals and harmonic signals. Hyperparameters were used to control the sparsity of reconstructed weights and the generation of basis parameters. In the experiments, we implemented the

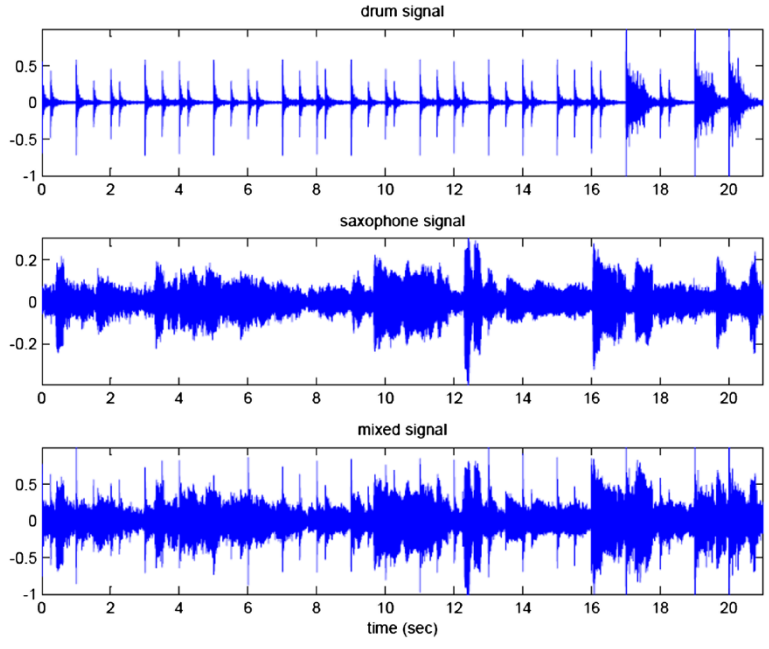

Figure 7 Waveforms of music 5 containing a drum signal, a saxophone signal, and their mixed signal.

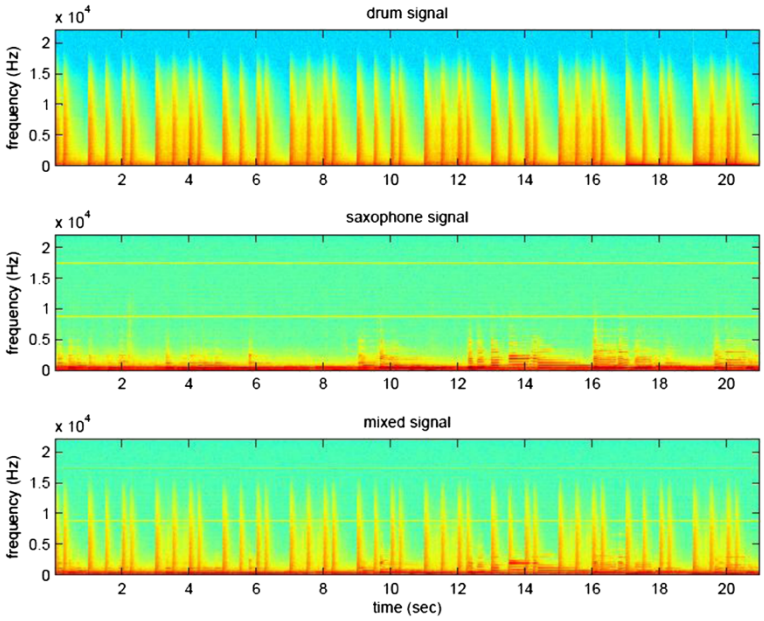

Figure 8 Spectrograms of music 5 containing a drum signal, a saxophone signal, and their mixed signal.

proposed BGS-NMFs for underdetermined source separation. The convergence condition of sampling procedure for approximate inference was investigated. The performance of BGS-NMF-LP and BGS-NMF-LSM was shown to be robust to the different kinds of rhythmic and harmonic sources and mixing conditions. BGS-NMF-LSM outperformed the other NMFs in terms of SIRs. The BGS-NMF controlled by LSM distribution performed better than that controlled by Laplacian distribution. In the future, the system performance of BGS-NMF may be further improved by some other considerations. For example, the numbers of common bases and individual bases could be automatically selected according to Bayesian framework by using marginal likelihood. The group sparse learning could be extended for constructing hierarchical NMF where hierarchical grouping of basis vectors is examined. The underdetermined separation under different number of sources and sensors could be tackled. Also, the online learning could be involved to update segment-based parameters and hyperparameters [33,34]. The evolutionary BGS-NMFs shall work for nonstationary single-channel blind source separation. In addition, more evaluations shall be conducted by using realistic data with larger amount of mixed speech signals from different application domains, such as meetings and call centers.

Appendix

Derivations for inference of BGS-NMF parameters and hyperparameters

We address some derivations for model inference of BGS-NMF parameters and hyperparameters. First, the

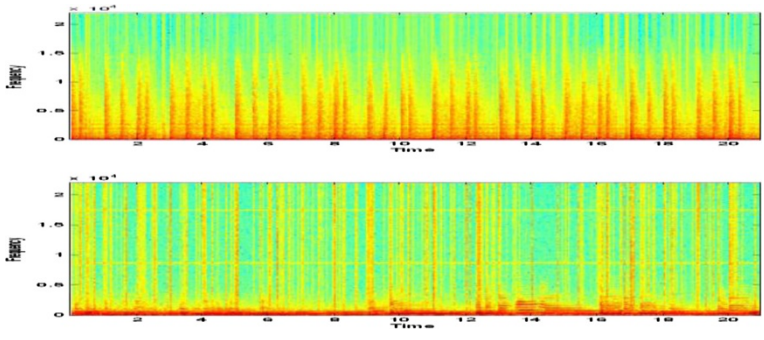

Figure 9 Spectrograms of the demixed drum signal (upper) and the demixed saxophone signal (lower).

exponent of the likelihood function $p(X_i^T|[A_{\mathrm{r}}^{(t+1)}]_{i(1:j-1)}, [A_{\mathrm{r}}^{(t)}]_{i(j+1:D_{\mathrm{r}})}, S_{\mathrm{r}}^{(t)}, A_{\mathrm{h}}^{(t)}, S_{\mathrm{h}}^{(t)}, \Sigma^{(t)})$ in (16) is expressed by

$$-\frac{1}{2[\,\Sigma^{(t)}]_{ii}}\sum_{k=1}^{M}\Bigg[X_{ik}-\sum_{m=1}^{j-1}[A_{\mathrm{r}}^{(t+1)}]_{im}\,[S_{\mathrm{r}}^{(t)}]_{mk} -[A_{\mathrm{r}}]_{ij}\,[S_{\mathrm{r}}^{(t)}]_{jk}-\sum_{m=j+1}^{D_{\mathrm{r}}}[A_{\mathrm{r}}^{(t)}]_{im}\,[S_{\mathrm{r}}^{(t)}]_{mk} -\sum_{m=1}^{D_{\mathrm{h}}}[A_{\mathrm{h}}^{(t)}]_{im}\,[S_{\mathrm{h}}^{(t)}]_{mk}\Bigg]^2 \tag{26}$$

which can be manipulated as a quadratic function of parameter $[A_{\mathrm{r}}]_{ij}$ and leads to (17). The conditional posterior distribution $p([A_{\mathrm{r}}]_{ij}|X_i^T, \Theta_{A_{rij}}^{(t)}, \Phi_{A_{rij}}^{(t)})$ is then derived by combining (17) and (11) and turns out to be

$$[A_{\mathrm{r}}]_{ij}^{\alpha_{rj}^{(t)}-1} \exp\left\{-\frac{[A_{\mathrm{r}}]_{ij}^2-2(\mu_{A_{rij}}^{\mathrm{likel}}-\beta_{rj}^{(t)}[\sigma_{A_{rij}}^{\mathrm{likel}}]^2)[A_{\mathrm{r}}]_{ij}+[\mu_{A_{rij}}^{\mathrm{likel}}]^2)}{2[\sigma_{A_{rij}}^{\mathrm{likel}}]^2}\right\} \mathbb{I}_{[0,+\infty[}([A_{\mathrm{r}}]_{ij}) \tag{27}$$

which is proportional to (18). In addition, when finding the mode of (18), we take logarithm of (18) and solve a corresponding quadratic equation of $[A_{\mathrm{r}}]_{ij}$ as

$$\frac{\partial}{\partial[A_{\mathrm{r}}]_{ij}}\left\{(\alpha_{rj}^{(t)}-1)\ln[A_{\mathrm{r}}]_{ij}-\frac{([A_{\mathrm{r}}]_{ij}-\mu_{A_{rij}}^{\mathrm{post}})^2}{2[\sigma_{A_{rij}}^{\mathrm{post}}]^2}\right\}=0 \Rightarrow [A_{\mathrm{r}}]_{ij}^2-\mu_{A_{rij}}^{\mathrm{post}}[A_{\mathrm{r}}]_{ij}-(\alpha_{rj}^{(t)}-1)[\sigma_{A_{rij}}^{\mathrm{post}}]^2=0. \tag{28}$$

By defining $\triangle=(\mu_{A_{rij}}^{\mathrm{post}})^2+4(\alpha_{rj}^{(t)}-1)[\sigma_{A_{rij}}^{\mathrm{post}}]^2$, the mode is determined by

$$\mu_{A_{rij}}^{\mathrm{inst}}=\begin{cases}0, & \text{if } \triangle<0\\ \max\{\frac{1}{2}(\mu_{A_{rij}}^{\mathrm{post}}+\sqrt{\triangle}),0\}, & \text{else.}\end{cases} \tag{29}$$

On the other hand, following the model inference in Section 3.3, we continue to describe the MCMC sampling algorithm and the calculation of conditional posterior distributions for the remaining BGS-NMF hyperparameters $\{\alpha_{\mathrm{r}}, \beta_{\mathrm{r}}, \alpha_{\mathrm{h}}, \beta_{\mathrm{h}}, \lambda_{\mathrm{r}}, \lambda_{\mathrm{h}}, \gamma_{\mathrm{r}}, \delta_{\mathrm{r}}, \gamma_{\mathrm{h}}, \delta_{\mathrm{h}}\}$.

4. Sampling of α_{rj}. The hyperparameter $\alpha_{rj}^{(t+1)}$ is sampled according to a conditional posterior distribution which is obtained by combining a likelihood function of $[A_{\mathrm{r}}]_{ij}$ and an exponential prior density of α_{rj} with parameter $\lambda_{\alpha_{rj}}$. The resulting distribution is written by

$$p(\alpha_{rj}|[A_{\mathrm{r}}^{(t+1)}]_{ij}, \beta_{rj}^{(t)}) \propto \left(\frac{1}{\Gamma(\alpha_{rj})}\exp\{\lambda_{\alpha_{rj}}^{\mathrm{post}}\alpha_{rj}\}\right)^{D_{\mathrm{r}}}\mathbb{I}_{[0,+\infty[}(\alpha_{rj}) \tag{30}$$

where $\lambda_{\alpha_{rj}}^{\mathrm{post}}=\ln\beta_{rj}^{(t)}+(1/D_{\mathrm{r}})\sum_{j=1}^{D_{\mathrm{r}}}\ln[A_{\mathrm{r}}^{(t+1)}]_{ij}-(1/D_{\mathrm{r}})\lambda_{\alpha_{rj}}$. This distribution does not belong to a known family, so the Metropolis-Hastings algorithm is applied. An instrumental distribution $q(\alpha_{rj})$ is obtained by fitting the term within the brackets of (30) through a gamma distribution as detailed in [15].

5. Sampling of β_{rj}. The hyperparameter $\beta_{rj}^{(t+1)}$ is sampled according to a conditional posterior distribution which is obtained by combining a likelihood function of $[A_{\mathrm{r}}]_{ij}$ and a gamma prior density of β_{rj} with parameters $\{\alpha_{\beta_{rj}}, \beta_{\beta_{rj}}\}$, i.e.,

$$p(\beta_{rj}|[A_{\mathrm{r}}^{(t+1)}]_{ij}, \alpha_{rj}^{(t+1)}) \propto (\beta_{rj})^{D_{\mathrm{r}}\alpha_{rj}^{(t+1)}} \times\exp\left\{-\beta_{rj}\sum_{j=1}^{D_{\mathrm{r}}}[A_{\mathrm{r}}^{(t+1)}]_{ij}\right\}\mathcal{G}(\beta_{rj}|\alpha_{\beta_{rj}}, \beta_{\beta_{rj}}). \tag{31}$$

The resulting distribution is arranged as a new gamma distribution $\mathcal{G}(\beta_{rj}|\alpha_{\beta_{rj}}^{\mathrm{post}}, \beta_{\beta_{rj}}^{\mathrm{post}})$ where $\alpha_{\beta_{rj}}^{\mathrm{post}}=1+D_{\mathrm{r}}\alpha_{rj}^{(t+1)}+\alpha_{\beta_{rj}}$ and $\beta_{\beta_{rj}}^{\mathrm{post}}=\sum_{j=1}^{D_{\mathrm{r}}}[A_r^{(t+1)}]_{ij}+\beta_{\beta_{rj}}$. Here, we do not describe the sampling of $\alpha_{hj}^{(t+1)}$ and $\beta_{hj}^{(t+1)}$ since the conditional posterior distributions for sampling these two hyperparameters are similar to those for sampling of $\alpha_{rj}^{(t+1)}$ and $\beta_{rj}^{(t+1)}$.

6. Sampling of λ_{rj} or λ_{hj}. For sampling of scaling parameter $\lambda_{rj}^{(t+1)}$, the conditional posterior distribution is obtained by

$$p(\lambda_{rj}|[S_{\mathrm{r}}^{(t+1)}]_{j(k=1:M)}, \gamma_{rj}^{(t)}, \delta_{rj}^{(t)}) \propto \prod_{k=1}^{M}p([S_{\mathrm{r}}^{(t+1)}]_{jk}|\lambda_{rj})p(\lambda_{rj}|\gamma_{rj}^{(t)}, \delta_{rj}^{(t)}) \propto (\lambda_{rj})^{M\gamma_{rj}^{(t)}}\exp\left\{-M\lambda_{rj}\left(\delta_{rj}^{(t)}+\sum_{k=1}^{M}[S_{\mathrm{r}}^{(t+1)}]_{jk}\right)\right\}. \tag{32}$$

7. Sampling of γ_{rj}. The sampling of LSM parameter $\gamma_{rj}^{(t+1)}$ is performed by using the conditional posterior distribution which is derived by combining a

likelihood function of λ_{rj} and an exponential prior density of γ_{rj} with parameter $\lambda_{\gamma_{rj}}$. The resulting distribution is expressed as

$$p(\gamma_{rj}|\lambda_{rj}^{(t+1)},\delta_{rj}^{(t)}) \propto \frac{1}{\Gamma(\gamma_{rj})}\exp\{\lambda_{\gamma_{rj}}^{\text{post}}\gamma_{rj}\}\mathbb{I}_{[0,+\infty[}(\gamma_{rj}), \tag{33}$$

where $\lambda_{\gamma_{rj}}^{\text{post}} = \ln\delta_{rj}^{(t)} + \frac{\gamma_{rj}-1}{\gamma_{rj}}\ln\lambda_{rj}^{(t+1)} - \lambda_{\gamma_{rj}}$. Again, we need to find an instrumental distribution $q(\gamma_{rj})$ which optimally fits the conditional posterior distribution $p(\gamma_{rj}|\lambda_{rj}^{(t+1)},\delta_{rj}^{(t)})$. An approximate gamma distribution is found accordingly. The Metropolis-Hastings algorithm is then applied.

8. Sampling of δ_{rj}. The sampling of the other LSM parameter $\delta_{rj}^{(t+1)}$ is performed by using the conditional posterior distribution which is derived from a likelihood function of λ_{rj} and a gamma prior density of δ_{rj} with parameters $\{\alpha_{\delta_{rj}},\beta_{\delta_{rj}}\}$

$$\begin{aligned} p(\delta_{rj}|\lambda_{rj}^{(t+1)},\gamma_{rj}^{(t+1)}) &\propto (\delta_{rj})^{\gamma_{rj}^{(t+1)}} \\ &\exp\{-\delta_{rj}\lambda_{rj}^{(t+1)}\}\mathcal{G}(\delta_{rj}|\alpha_{\delta_{rj}},\beta_{\delta_{rj}}). \end{aligned} \tag{34}$$

This distribution can be arranged as a new gamma distribution $\mathcal{G}(\delta_{rj}|\alpha_{\delta_{rj}}^{\text{post}},\beta_{\delta_{rj}}^{\text{post}})$ where $\alpha_{\delta_{rj}}^{\text{post}} = D_{\text{r}}\gamma_{rj}^{(t+1)} + \alpha_{\delta_{rj}}$ and $\beta_{\delta_{rj}}^{\text{post}} = \lambda_{rj}^{(t+1)} + \beta_{\delta_{rj}}$. Similarly, the conditional posterior distributions for sampling $\gamma_{hj}^{(t+1)}$ and $\delta_{hj}^{(t+1)}$ could be formulated by referring those for sampling $\gamma_{rj}^{(t+1)}$ and $\delta_{rj}^{(t+1)}$, respectively.

Competing interests
Both authors declare that they have no competing interests.

Acknowledgments
The authors acknowledge anonymous reviewers for their constructive feedback and helpful suggestions. This work has been partially supported by the National Science Council, Taiwan, Republic of China, under contract NSC 100-2628-E-009-028-MY3.

Received: 28 October 2012 Accepted: 13 May 2013
Published: 5 July 2013

References

1. A Cichocki, R Zdunek, S Amari, in *Proceedings of International Conference on Acoustic, Speech and Signal Processing (ICASSP)*. New algorithms for non-negative matrix factorization in applications to blind source separation (IEEE, Piscataway, 2006), pp. 621–624
2. PO Hoyer, Non-negative matrix factorization with sparseness constraints. J. Mach. Lear. Res. **5**, 1457–1469 (2004)
3. J-T Chien, H-L Hsieh, Convex divergence ICA for blind source separation. IEEE Trans. Audio, Speech, Language Process. **20**(1), 290–301 (2012)
4. R Kompass, A generalized divergence measure for nonnegative matrix factorization. Neural Comput. **19**, 780–791 (2007)
5. H Lee, J Yoo, S Choi, Semi-supervised nonnegative matrix factorization. IEEE Signal Process. Lett. **17**(1), 4–7 (2010)
6. MD Plumbley, Algorithms for nonnegative independent component analysis. IEEE Trans. Neural Netw. **14**(3), 534–543 (2003)
7. CM Bishop, *Pattern Recognition and Machine Learning* (Springer Science, New York, 2006)
8. G Saon, J-T Chien, Bayesian sensing hidden Markov models. IEEE Trans. Audio, Speech Language, Process. **20**(1), 43–54 (2012)
9. ME Tipping, Sparse Bayesian learning and the relevance vector machine. J Mach. Learn. Res. **1**, 211–244 (2001)
10. SD Babacan, R Molina, AK Katsaggelos, Bayesian compressive sensing using Laplace priors. IEEE Trans. Image Process. **19**(1), 53–63 (2010)
11. H Lee, S Choi, in *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 15. Group nonnegative matrix factorization for EEG classification (JMLR, 2009), pp. 320–327
12. A Lefevre, F Bach, C Fevotte, Itakura-Saito, in *Proceedings of the International Conference on Acoustic, Speech and Signal Processing (ICASSP)*. Nonnegative matrix factorization with group sparsity (Prague Congress Center, 22–27 May 2011), pp. 21–24
13. M Kim, J Yoo, K Kang, S Choi, in *Proceedings of the International Conference on Acoustic, Speech and Signal Processing (ICASSP)*. Blind rhythmic source separation: nonnegativity and repeatability (IEEE, Piscataway, 2010), pp. 2006–2009
14. AT, Cemgil, Bayesian inference for nonnegative matrix factorization models. University of Cambridge, Technical Report CUED/F-INFENG/TR.609, 2008
15. S Moussaoui, D Brie, Mohammad-A Djafari, C Carteret, Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling. IEEE Trans. Signal Process. **54**(11), 4133–4145 (2006)
16. MN Schmidt, O Winther, LK Hansen, in *Proceedings of the International Conference on Independent Component Analysis and Signal Separation*, Paraty, March 2009. Lecture Notes in Computer Science, vol. 5441. Bayesian non-negative matrix factorization (Springer, Heidelberg, 2009), pp. 540–547
17. C Fevotte, SJ Godsill, A Bayesian approach for blind separation of sparse sources. IEEE Trans. Audio, Speech, Language Process. **14**(6), 2174–2188 (2006)
18. Z Duan, Y Zhang, C Zhang, Z Shi, Unsupervised single-channel music source separation by average harmonic structure modeling. IEEE Trans. on Audio, Speech, Language Process. **16**(4), 766–778 (2008)
19. MN Schmidt, RK Olsson, in *Proceedings of the Annual Conference of International Speech Communication Association (INTERSPEECH)* Single-channel speech separation using sparse non-negative matrix factorization (Pittsburgh, 17–21 September 2006), pp. 2614–2617
20. J-T Chien, H-L Hsieh, in *Proceedings of the Annual Conference of International Speech Communication Association (INTERSPEECH)*. Bayesian group sparse learning for nonnegative matrix factorization (Portland, 9–13 September 2012), pp. 1552–1555
21. J Yoo, M Kim, K Kang, Choi S, in *Proceedings of the International Conference on Acoustic, Speech and Signal Processing (ICASSP)* Nonnegative matrix partial co-factorization for drum source separation (IEEE, Piscataway, 2010), pp. 1942–1945
22. M Kim, J Yoo, K Kang, S Choi, Nonnegative matrix partial co-factorization for spectral and temporal drum source separation. IEEE J. Sel. Top. Signal Process. **5**(6), 1192-1204 (2011)
23. S Bengio, F Pereira, Y Singer, D Strelow, in *Advances in Neural Information Processing Systems (NIPS)*, vol. 22. Group sparse coding (NIPS La Jolla, 2009), pp. 82–89
24. R Jenatton, J Mairal, G Obozinski, F Bach, in *Proceedings of the International Conference on Machine Learning (ICML)*. Proximal methods for sparse hierarchical dictionary learning (Haifa, 21–25 June 2010)
25. PJ Garrigues, BA Olshausen, in *Advances in Neural Information Processing Systems (NIPS)*, vol. 23. Group sparse coding with a Laplacian scale mixture prior (NIPS La Jolla, 2010), pp. 676–684
26. R Salakhutdinov, A Mnih, in *Proceedings of the International Conference on Machine Learning (ICML)*. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo (Helsinki, 5–9 July 2008), pp. 880–887
27. M Zhong, M Girolami, in *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. Reversible jump MCMC for non-negative matrix factorization (Clearwater Beach, 16–18 April 2009), pp. 663–670
28. MD Hoffman, DM Blei, PR Cook, in *Proceedings of the International Conference on Machine Learning (ICML)*. Bayesian nonparametric matrix factorization for recorded music (Haifa, 21–24 June 2010)

29. M Marlin, BM, Schmidt, KP Murphy, in *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*. Group sparse priors for covariance estimation (Montreal, 18–21 June 2009), pp. 383–392
30. J-T Chien, C-C Chiang, in *Proceedings of the Annual Conference of International Speech Communication Association (INTERSPEECH)*. Group sparse hidden Markov models for speech recognition (Portland, 9–13 September 2012), pp. 2646–2649
31. J-T Chien, C-W Ting, Factor analyzed subspace modeling and selection. IEEE Trans. Audio, Speech Language Process. **16**(1), 239–248 (2008)
32. S Chib, E Greenberg, Understanding the Metropolis-Hastings algorithm. Am. Statistician. **49**(4), 327–335 (1995)
33. J-T Chien, H.-L Hsieh, Nonstationary source separation using sequential and variational Bayesian learning. IEEE Trans. Neural Netw. Learn. Syst. **24**(5), 681–694 (2013)
34. H-L Hsieh, J-T Chien, in *Proceedings of the International Conference on Acoustic, Speech and Signal Processing (ICASSP)*. Nonstationary and temporally-correlated source separation using Gaussian process (Prague Congress Center, 22–27 May 2011), pp. 2120–2123

doi:10.1186/1687-4722-2013-18
Cite this article as: Chien and Hsieh: **Bayesian group sparse learning for music source separation.** *EURASIP Journal on Audio, Speech, and Music Processing* 2013 **2013**:18.