Audio Source Separation Based on Nonnegative Matrix Factorization with Graph Harmonic Structure

Tomohiro Ichita*, Seisuke Kyochi*, and Keisuke Imoto†
* The University of Kitakyushu, Fukuoka, Japan
E-mail: s-kyochi@kitakyu-u.ac.jp
† Ritsumeikan University, Shiga, Japan
E-mail: k-imoto@fc.ritsumei.ac.jp

***Abstract*—This paper proposes a novel single-channel audio source separation based on graph-regularized nonnegative matrix factorization (NMF) taking harmonic frequency structure of each instrument into account. Since the original NMF, which is regarded as unsupervised learning, cannot readily identify the corresponding basis matrix for each target source, supervised NMFs (SNMFs) using given basis matrices learned from training sources have been extensively studied. Although SNMFs usually separate a mixed source better than NMF, the performance is degraded when training sources different from the observed source. The proposed SNMF does not use learned basis matrices but uses learned graph Laplacian matrices characterizing a harmonic frequency structure of training sources for regularization. Even if training sources are different from target sources, the graph structures from observed and training sources are more correlated, thus, as experimental results show, it can separate more robustly.**

I. Introduction

Audio source separation is essential for automatic music transcription, speech recognition, and so on [1], [2]. Nonnegative matrix factorization (NMF) plays an important role, which factorizes an observed spectrogram into two nonnegative matrices as

$$\mathbf{Y} \approx \mathbf{D}^{\star}\mathbf{X}^{\star}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^{\Omega \times T}$ denotes the spectrogram of an input signal, $\mathbf{D}^{\star} \in \mathbb{R}^{T \times L}$ is usually called as the basis matrix, and $\mathbf{X}^{\star} \in \mathbb{R}^{L \times T}$ is as the activation one (for mathematical expressions, see *Notations* in the end of this section). Each basis vector in the basis matrix typically characterizes a spectral pattern of each source, and we can extract a particular source from a product of a set of vectors in the basis matrix and a set of rows in the activation matrix.

Since the original NMF is an unsupervised NMF (UNMF), i.e., it does not use any prior information on each source in a mixed input source, it suffers from the problem. Specifically, it is difficult to identify a set of basis vectors corresponding to its source. This degrades the accuracy of BSS significantly. To overcome the UNMF's problem, supervised NMFs (SNMFs) have been extensively studied [3], [4]. In SNMFs, a set of basis vectors corresponding to a target source is given by learned vectors obtained from training sources. By introducing the pre-learned matrix, SNMFs can efficiently separate the mixed source better than NMF.

However, SNMFs suffer from a problem. Even though the harmonic frequency structures among the same type of musical instruments (e.g., piano, violin) are close, every individual instrument does not necessarily have the exact same structure. Thus, if a training source of a instrument are different from a test source, the learned vectors are not matched to the input source. Consequently, the source separation accuracy would be degraded.

To overcome this problem, this paper focuses on a graph structure of the harmonic frequency structure of each type of the instrument. The harmonic frequency structure of each note exhibits an impulsive spectral pattern. Its structure, i.e., positions and heights of peaks (see the top-left and the top-right figures in Fig. 1), tends to be close among individuals in the same kind of a musical instrument. Thus, if this tendency can be naturally integrated into the cost function of SNMFs, it is expected that more robust separation can be realized.

Our contributions in this paper are summarized as follows.

1) The harmonic frequency structures are characterized by using the graph Laplacian matrix. It represents how strongly two nodes (i.e., two frequencies) are connected.
2) A multiplicative update algorithm of the PSNMF [4] with graph regularization is presented.

The rest of this paper is organized as follows. Section II reviews the conventional UNMF and SNMFs and fundamentals of graph signal processing. Then, the proposed graph-regularized PSNMF (GPSNMF) with graph harmonic structure in Section III. The proposed method is evaluated in the experiments of compressed image sensing in Section IV. Finally, this paper is concluded in Section V.

Notations: Bold-faced lower-case and upper-case letters denote vectors and matrices, respectively. Sets $\mathbb{R}$ and $\mathbb{R}_{+}$ respectively denote real and non-negative real numbers, respectively. Real-valued and non-negative real-valued matrices of size N_r [row] and N_c [column] are described as $\mathbb{R}^{N_r \times N_c}$ and $\mathbb{R}_{+}^{N_r \times N_c}$. $A_{i,j}$ or $[\mathbf{A}]_{i,j}$ denotes the i-th row j-th column

This work was supported in part by JSPS Grants-in-Aid (17K14683).

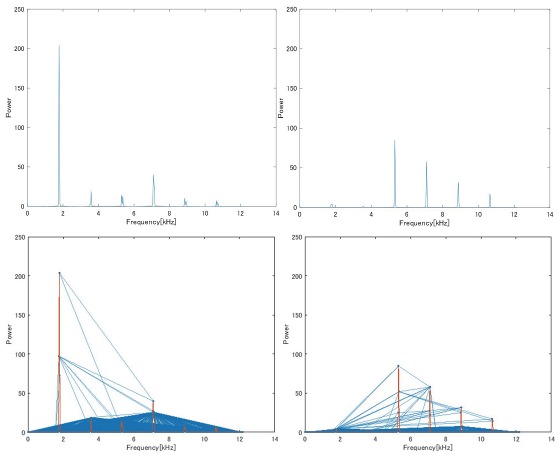


Fig. 1. Harmonic structures: clarinet (top-left), violin (top-right), graph harmonic structures: clarinet (bottom-left), violin (bottom-right)

element of a matrix $\mathbf{A} \in \mathbb{R}^{N_r \times N_c}$. $\|\cdot\|_F$ is the Frobenius norm. $\mathbf{1}$ is reserved for the all-ones vector.

II. Preliminaries

A. Nonnegative Matrix Factorization

The factorization problem of (1) is mathematically formulated as:

$$(\mathbf{D}^\star, \mathbf{X}^\star) = \underset{\mathbf{D},\mathbf{X}}{\operatorname{argmin}} \mathcal{J}_{\mathrm{NMF}}(\mathbf{D}, \mathbf{X}),$$
$$\mathcal{J}_{\mathrm{NMF}}(\mathbf{D}, \mathbf{X}) := \mathcal{D}(\mathbf{Y} \| \mathbf{D}\mathbf{X}), \tag{2}$$

where $\mathcal{D}$ is a distance function which is typically set to Euclid distance, general Kullback-Leibler (KL) divergence, or Itakura-Saito divergence [4]. For example, for given x and y be the random variables, KL divergence can be given as follows:

$$\mathcal{D}_{\mathrm{KL}}(y\|x) = y(\log y - \log x) + x - y. \tag{3}$$

According to (3), the cost function (2) is represented as the following equation:

$$\begin{aligned}&\mathcal{J}_{\mathrm{NMF}}(\mathbf{D}, \mathbf{X}) \\ &= \mathcal{D}_{\mathrm{KL}}(\mathbf{Y}\|\mathbf{D}\mathbf{X}) \\ &= \sum_{\omega=0}^{\Omega-1}\sum_{t=0}^{T-1}\left(Y_{\omega,t}\log\frac{Y_{\omega,t}}{[\mathbf{D}\mathbf{X}]_{\omega,t}} - (Y_{\omega,t} - [\mathbf{D}\mathbf{X}]_{\omega,t})\right).\end{aligned} \tag{4}$$

Hereafter, we only consider the KL divergence as a distance function.

B. SNMF for BSS

The problem of the original unsupervised NMF (UNMF) is that it is difficult to identify vectors in the basis matrix for the corresponding target source because the UNMF does not use any prior information. To tackle this problem, for a given K-source mixed input signal, the SNMF [3], [4] tries to find a basis matrix $\mathbf{D} = [\mathbf{D}_0,\ \mathbf{D}_1]$ that includes a K_0 pre-learned sub-basis matrices $\mathbf{D}_0 := \left[\mathbf{D}^{(1)} \quad \cdots \quad \mathbf{D}^{(K_0)}\right]$ for K_0-sources fixed and finds the minimizer $\mathbf{D}_1^\star := \mathbf{D}^{(K_0+1)}$ and $\mathbf{X}^\star := \left[\mathbf{X}_0^{\star\top} \quad \mathbf{X}_1^{\star\top}\right]^\top$ of the following equation.

$$(\mathbf{D}_1^\star, \mathbf{X}^\star) = \underset{\mathbf{D}_1,\mathbf{X}}{\operatorname{argmin}} \mathcal{J}_{\mathrm{SNMF}}(\mathbf{D}_1, \mathbf{X}),$$
$$\mathcal{J}_{\mathrm{SNMF}}(\mathbf{D}_1, \mathbf{X}) = \mathcal{D}_{\mathrm{KL}}\left(\mathbf{Y}\,\|\mathbf{D}_0\mathbf{X}_0 + \mathbf{D}_1\mathbf{X}_1\right). \tag{5}$$

To robustify SNMF, the penalized SNMF (PSNMF) [4] introduces penalty terms for orthogonal condition into its cost function to make $\mathbf{D}_0$ and $\mathbf{D}_1$ uncorrelated.The cost function is given as:

$$\mathcal{J}_{\mathrm{PSNMF}}(\mathbf{D}_1, \mathbf{X}) = \mathcal{J}_{\mathrm{SNMF}}(\mathbf{D}_1, \mathbf{X}) + \mu\|\mathbf{D}_0^\top\mathbf{D}_1\|_F^2. \tag{6}$$

For simple discussion, we set $K_0 = 1$ (i.e., $\mathbf{D}_0 := \mathbf{D}^{(1)}$ and $\mathbf{D}_1 := \mathbf{D}^{(2)}$), then the above cost function is reduced to:

$$\mathcal{J}_{\mathrm{PSNMF}}(\mathbf{D}^{(2)}, \mathbf{X}) = \mathcal{J}_{\mathrm{SNMF}}(\mathbf{D}^{(2)}, \mathbf{X}) + \mu\|\mathbf{D}^{(1)\top}\mathbf{D}^{(2)}\|_F^2. \tag{7}$$

Applying Jensen's inequality, introducing the auxiliary variables, and taking derivative to the auxiliary function, the multiplicative rule for the above function can be given as:

$$\begin{aligned}X_{i,t}^{(1)} &\leftarrow X_{i,t}^{(1)}\frac{\sum_\omega Y_{\omega,t}D_{\omega,i}^{(1)}\left(\sum_{k=1}^{2}[\mathbf{D}^{(k)}\mathbf{X}^{(k)}]_{\omega,t}\right)^{-1}}{\sum_\omega D_{\omega,i}^{(1)}},\\ D_{\omega,j}^{(2)} &\leftarrow D_{\omega,j}^{(2)}\frac{\sum_t Y_{\omega,t}X_{j,t}^{(2)}\left(\sum_{k=1}^{2}[\mathbf{D}^{(k)}\mathbf{X}^{(k)}]_{\omega,t}\right)^{-1}}{\sum_t X_{j,t}^{(2)} + 2\mu\sum_{i'}D_{\omega,I'}^{(1)}\sum_{\omega'}D_{\omega',i'}^{(1)}D_{\omega',j}^{(2)}},\\ X_{j,t}^{(2)} &\leftarrow X_{j,t}^{(2)}\frac{\sum_\omega Y_{\omega,t}D_{\omega,j}^{(2)}\left(\sum_{k=1}^{2}[\mathbf{D}^{(k)}\mathbf{X}^{(k)}]_{\omega,t}\right)^{-1}}{\sum_\omega D_{\omega,j}^{(2)}}.\end{aligned} \tag{8}$$

Note that the update rule (8) can be reduced to the one for the UNMF by setting $\mu = 0$.

C. Graph-based Reguralization

This section reviews the fundamental elements of graph signal processing (GSP) [5]. Graph signals have intensity at each node, and the geometrical information of graphs (connection among nodes) determined by an adjacent matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$. Fig. 2 shows an example of undirected graph signals. Its adjacent matrix can be given as:

$$\mathbf{W} = \begin{bmatrix} 0 & 0 & 1 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0.3 \\ 1 & 0 & 0 & 0 & 0 \\ 0.3 & 0.5 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 & 0 \end{bmatrix}. \tag{9}$$

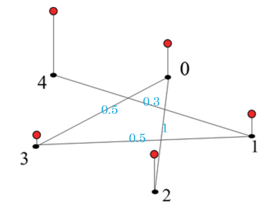
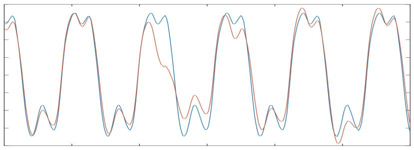
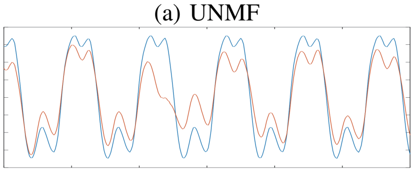


Fig. 2. Undirected graph signal



(a) UNMF



(b) SNMF

Fig. 3. Reconstructed signals from the UNMF and the SNMF (blue line: original signal, orange line: reconstructed signal from the UNMF or the SNMF)

Graph Laplacian matrices are often used in GSP, which is defined by using incident matrix $\mathbf{\Delta} = \mathrm{diag}(\sum_j W_{1,j}, \ldots, \sum_j W_{N,j})$ and adjacent matrix as $\mathbf{L} = \mathbf{\Delta} - \mathbf{W}$. The graph Laplacian matrix is incorporated into many signal recovery problem. For example, consider the problem to recover a signal $\mathbf{d} \in \mathbb{R}^N$ smoothly distributing over a graph with an adjacent matrix $\mathbf{W}$. In this case, two values in the optimal signal should have a small difference where two nodes are connected with a large weight. Thus, the following term is usually integrated into the cost function:

$$\sum_{i,j=0}^{N-1} w_{i,j}(d_i - d_j)^2 = \mathbf{d}^\top \mathbf{L} \mathbf{d}. \tag{10}$$

III. Graph-regularized NMF with Harmonic Structural Information

Although the conventional SNMFs can improve the UNMFs by using pre-trained vectors in the basis matrix, it might degrade its performance when training signals have a different sound structure from the corresponding source mixed in the input audio signal. Specifically,

1) the accuracy of a target source extraction decrease due to the mismatch of pre-trained vectors,
2) the overall reconstruction quality (i.e., $\mathcal{D}(\mathbf{Y}\|\mathbf{DX})$ in (2)) also decreases due to fixing some free parameters for the basis matrix in optimization.

We demonstrate the second problem. Fig. 3 indicates an original mixed signal (blue line) and the estimated overall one (orange line) after NMF. Obviously, the approximation of the SNMF is worse than the UNMF. On the other hand, although the UNMF approximates the input source, the UNMF cannot identify the basis vectors for the corresponding source. In order to overcome the problems of the UNMF and the SNMF simultaneously, we integrate graph harmonic structure into NMF framework.

A. Graph Characterization of Harmonic Structure

As shown in the top figures of Fig. 1, a power spectrum of each musical instrument has sparse peaks. Depending on a instrument, heights of peaks form a different shape. From the viewpoint of GSP, we regard a power spectrum signal as ring undirected graph signal, and term it as *graph harmonic structure*.

In this paper, we take a conventional learning approach [6] for graph Laplacian matrices of each note as :

$$\begin{aligned} \mathbf{L}^\star = \ & \operatorname*{argmin}_{\mathbf{L}} \alpha \mathrm{Tr}(\mathbf{D}^\top \mathbf{L} \mathbf{D}) + \mu \|\mathbf{L}\|_F^2 \\ & \text{s.t} \ \ L_{i,j} = L_{j,i} \le 0, \ \mathbf{L} \cdot \mathbf{1} = \mathbf{0}, \ \mathrm{Tr}(\mathbf{L}) = N, \end{aligned} \tag{11}$$

where the first and the second constraints imposes $\mathbf{L}$ being graph Laplacian matrix and the third one is to avoid $\mathbf{L}$ being zero matrix. $\mathbf{D}$ consists of power spectra obtained by STFT of a given single note sound.

An example of the graph harmonic structure is shown in the bottom figures of Fig. 1, where blue lines show the connectivity between frequencies. We can observe that tall peaks have few connections compared with low peaks. The graph harmonic structure can be useful as prior information on each instrument.

B. Graph-regularized PSNMF

Conventionally, graph-regularized NMF approaches have been proposed [7], [8]. Inspired by the conventional approaches, we integrate learned graph Laplacian matrices of notes from a musical instrument into the cost function as follows:

$$\begin{aligned} \mathcal{J}_{\mathrm{GPSNMF}}(\mathbf{D}, \mathbf{X}) = \mathcal{J}_{\mathrm{NMF}}(\mathbf{D}, \mathbf{X}) + \sum_{i,j,(i \ne j)} \mu_{i,j} \|\mathbf{D}^{(i)\top} \mathbf{D}^{(j)}\|_F^2 \\ + \sum_{k=1}^{K} \alpha_k \mathrm{Tr}\left(\mathbf{D}^{(k)\top} \mathbf{L}^{(k)} \mathbf{D}^{(k)}\right) \end{aligned} \tag{12}$$

This problem is termed as graph-regularized PSNMF (GPSNMF). By using graph regularization, the cost function can naturally take the prior information on the basis matrix for each source into account without fixing some parameters of the basis matrices for optimization. Thus, we can find appropriate matrices $\mathbf{D}_k$ for the corresponding source, even though the training sources are different from an input source. Furthermore, because all the parameters of the basis matrix can be fully optimized (unlike the conventional SNMFs), the reconstruction error can be decreased.

C. Multiplicative Update Rule

In this section, we show the update rule for GPSNMF. For simple discussion, we restrict the number of sources mixed in an input signal $K = 2$, and customize the cost function (12) into the following eqation:

$$\begin{aligned} \mathcal{J}_{\mathrm{GPSNMF}}(\mathbf{D}, \mathbf{X}) = \mathcal{J}_{\mathrm{NMF}}(\mathbf{D}, \mathbf{X}) + \mu \|\mathbf{D}^{(1)\top} \mathbf{D}^{(2)}\|_F^2 \\ + \alpha \mathrm{Tr}\left(\mathbf{D}^{(1)\top} \mathbf{L}^{(1)} \mathbf{D}^{(1)}\right). \end{aligned} \tag{13}$$

In similar discussion in the UNMF, the PSNMF [4], and the graph-regularized NMF methods [7], [8], by applying Jensen's inequality, introducing auxiliary variables, and taking derivative to the auxiliary function, the multiplicative update rule for (13) can be derived as follows:

$$D^{(1)}_{\omega,i} \leftarrow D^{(1)}_{\omega,i}\frac{\sum_t Y_{\omega,t}X^{(1)}_{i,t}P^{-1}_{\omega,t} + \alpha[\mathbf{W}\mathbf{D}^{(1)}]_{\omega,i}}{\sum_t X^{(1)}_{i,t} + 2\mu Q^{(2,1)}_{\omega,i} + \alpha[\mathbf{\Delta}\mathbf{D}^{(1)}]_{\omega,i}},$$

$$X^{(1)}_{i,t} \leftarrow X^{(1)}_{i,t}\frac{\sum_\omega Y_{\omega,t}D^{(1)}_{\omega,i}\left(\sum_{k=1}^{2}[\mathbf{D}^{(k)}\mathbf{X}^{(k)}]_{\omega,t}\right)^{-1}}{\sum_\omega D^{(1)}_{\omega,i}},$$

$$D^{(2)}_{\omega,j} \leftarrow D^{(2)}_{\omega,j}\frac{\sum_t Y_{\omega,t}X^{(2)}_{i,t}P^{-1}_{\omega,t}}{\sum_t X^{(2)}_{j,t} + 2\mu Q^{(1,2)}_{\omega,i}},$$

$$X^{(2)}_{j,t} \leftarrow X^{(2)}_{j,t}\frac{\sum_\omega Y_{\omega,t}D^{(2)}_{\omega,j}\left(\sum_{k=1}^{2}[\mathbf{D}^{(k)}\mathbf{X}^{(k)}]_{\omega,t}\right)^{-1}}{\sum_\omega D^{(2)}_{\omega,j}},$$

$$P_{\omega,t} = \sum_{i'} D^{(1)}_{\omega,i'}X^{(1)}_{i',t} + \sum_{j'} D^{(2)}_{\omega,j'}X^{(2)}_{j',t},$$

$$Q^{(\ell_1,\ell_2)}_{\omega,i} = \sum_{j'} D^{(\ell_1)}_{\omega,j'}\sum_{\omega'} D^{(\ell_1)}_{\omega',j'}D^{(\ell_2)}_{\omega',i}. \tag{14}$$

IV. Experimental results

We evaluated the performance of the proposed method, compared with the UNMF, the SNMF, and the PSNMF. The test mixed musical sources are artificially synthesized from two of oboe, clarinet, piano, and trombone. The data set of [9] was used in this experiment. First, we apply the proposed or the conventional NMFs to the input mixed sources, then reconstruct only the target sources from the corresponding basis and activation matrices. The training signal consisted of 24 notes over two octaves as in Fig. 4(a). The other experimental conditions are summarized in Table I.

We conducted two experiments. In the first experiment, both training and test data are generated by the same synthesizer (YAMAHA MU-1000 PCM-based MIDI synthesizer). In the second experiment, different synthesizers are used. Training data is generated by the YAMAHA MU-1000 PCM-based MIDI synthesizer and test data is generated by the Microsoft GS Wavetable Synth.

We evaluate the separation performance by the signal-to-distortion ratio (SDR) (see [10] for detail information):

$$\mathrm{SDR} = 10\log\left(\frac{\sum_t(s_{target}(t))^2}{\sum_t(e_{interf}(t)+e_{noise}(t))^2}\right). \tag{15}$$

Numerical results are shown in Table II. In both experiments, the proposed GPSNMF can achieve higher SDR in most cases. In particular, in the second experiment, the GPSNMF significantly improve the BSS performance because the proposed method can robustly find the basis matrix by graph harmonic structure, even though the training and the test data are synthesized from different software.

V. Concluding Remarks

In this paper, we proposed the GPSNMF for audio source separation. Unlike the conventional SNMF, the GPSNMF uses graph harmonic structures as prior information for each instrument, instead of learned basis matrices. Even if the sound quality of a training source is different from the test source, the GPSNMF could extract the target source more robustly than the conventional UNMF and SNMFs.

(a) Training data

(b) Test data

Fig. 4. Test and training notes

TABLE I
Experimental Setup

Target signal (MIDI)	Oboe, Clarinet, Piano, Trombone
Observed signal (MIDI)	Mixture of 2 target signals
Sampling frequency	44.1 kHz
Frame length of STFT	4096 points
Shift length of STFT	2048 points
Target signal basis number	100
Interference signal basis number	50
Number of iterations NMF (conv./prop.)	500

TABLE II
Experimental Results (dB)

Experiment 1					
Target	Interference	UNMF	SNMF	PSNMF	GPRNMF
Oboe	Piano	4.93	7.00	8.67	**8.97**
	Trombone	4.76	6.72	7.73	**7.74**
Clarinet	Oboe	3.14	3.13	4.32	**5.45**
	Piano	5.03	5.49	9.23	**14.31**
	Trombone	4.89	5.36	9.10	**13.48**

Experiment 2					
Target	Interference	UNMF	SNMF	PSNMF	GPRNMF
Oboe	Piano	1.95	2.18	**3.68**	3.40
	Trombone	3.39	5.64	6.56	**6.80**
Clarinet	Oboe	2.70	2.78	3.44	**5.67**
	Piano	3.21	3.92	5.69	**6.69**
	Trombone	4.14	7.93	8.92	**9.69**

References

[1] C. Pierre and J. Christian, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic Press, Inc., Orlando, FL, USA, 1st edition, 2010.

[2] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proc. 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, Oct 2003, pp. 177–180.

[3] P. Smaragdis, B. Raj, and M. Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," in *Proc. the 7th International Conference on Independent Component Analysis and Signal Separation*, Berlin, Heidelberg, 2007, ICA'07, pp. 414–421, Springer-Verlag.

[4] D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi, and K. Kondo, "Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E97.A, no. 5, pp. 1113–1118, 2014.

[5] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[6] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Laplacian matrix learning for smooth graph signal representation," in *Proc. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 3736–3740.

[7] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug 2011.

[8] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2030–2048, July 2011.

[9] D. Kitamura, H. Saruwatari, H. Kameoka, Y. Takahashi, K. Kondo, and S. Nakamura, "Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 23, no. 4, pp. 654–669, April 2015. http://d-kitamura.net/.

[10] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. on Audio, Speech, and Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, July 2006.