Sparsity Promoted Non-Negative Matrix Factorization for Source Separation and Detection

Yanlin Wang, Yun Li, K. C. Ho, A. Zare, M. Skubic
Department of Electrical and Computer Engineering
University of Missouri
Columbia, MO 65211, USA
e-mail: yw3z8@mail.missouri.edu, yl874@mail.missouri.edu, hod@missouri.edu, zarea@missouri.edu, skubicm@missouri.edu

Abstract— **The effectiveness of non-negative matrix factorization (NMF) depends on a suitable choice of the number of bases, which is often difficult to decide in practice. This paper imposes sparseness on the factorization coefficients in order to determine the number of bases automatically during the decomposition process. The benefit of sparse promotion for NMF is demonstrated through application to sound source separation as well as acoustic-based human fall detection under strong interference.**

Keywords—elder care, non-negative matrix factorization, source separation, sparse promotion

I. INTRODUCTION

Non-negative Matrix Factorization (NMF) decomposes a matrix of non-negative elements into a product of two matrices each consisting only non-negative elements. Since the introduction of NMF and the multiplicative algorithm by Lee and Seung [1] in 1999, NMF has become a popular research topic in signal processing. Research on NMF ranges from developing more accurate and efficient algorithms for factorization to its usage in various applications including acoustic/audio signal processing [2], source separation [3], hyperspectral analysis [4] and cognitive radio [5].

NMF assumes an over-represented matrix in which each column can be decomposed into a linear combination of a small number of basis vectors. In the factorization, the first is a collection of these basis vectors and the second matrix provides the coefficients of decomposition. In general, given a low reconstruction error, fewer the number of basis vectors provides more effective NMF result. One challenge in the use of NMF is that the number of basis vectors has to be chosen before the factorization, which is often difficult to decide in practice. The best value for the number of basis vectors is often application dependent. Selection of an incorrect number of bases significantly reduces the effectiveness and performance of NMF.

This paper imposes a sparsity-promoting prior on the decomposition coefficients so that the factorization process will be able to determine the appropriate number of basis vectors automatically. Sparsity is promoted through the addition of a penalty term in the cost function. Starting with more bases than necessary, the sparsity promoting prior drives insignificant elements in the coefficient matrix to zero such that the associated basis vectors can be removed. Only the necessary basis vectors will remain in the final result. Sparse NMF has found to provide more reliable and accurate results than NMF [6]-[7].

Previous works on sparse NMF [6]-[7] focus on image representation and processing. In these methods, sparseness is imposed on both the basis matrix and the coefficient matrix. In this work, we only enforce sparsity in the coefficient matrix. In addition, we examine the use of sparse NMF for signal detection. Although sparse NMF has been applied to various applications, there are relatively few studies on its ability to improve signal detection.

We shall use the source separation problem [8]-[9] to demonstrate the advantage of incorporating sparsity promotion in NMF. Given the magnitude of the Short-Time Fourier Transform (STFT) of a mixture of two signals, we apply sparsity promoted NMF and separate the bases into two classes that correspond to the two signal components. The two signals are then reconstructed through their separated STFT magnitude spectra, and the error in signal recovery will be contrasted with a normal NMF that uses a fixed number of bases to do factorization.

The second application is on the use of sparsity promoted NMF for fall detection in elder care through the acoustic signal generated from a human fall. A fall signal often is corrupted by strong interference (e.g. audio signal from TV) and the role of NMF is to separate the fall signal from the interference for detection. The detection will be accomplished by using the Mel-Frequency Cepstrum Coefficients (MFCC) features [10] for fall versus nonfall classification. Results from the data containing 120 falls and 120 nonfalls will be demonstrated.

The next section gives a brief overview of NMF. Section III presents the sparsity promoted NMF. Section IV describes the source separation results and Section V the fall detection performance. Section VI concludes the paper.

This work was supported in part by the National Science Foundation under Grant CNS-0931607.

II. NMF

NMF operates on a matrix **V** that is formed by a collection of *M* column vectors each has a length *N*. Each element of **V** is known to be non-negative (positive or zero). The columns of **V** are over-represented and are expected to be a linear combination of a number of basis vectors that is much less than *N*. The goal of NMF is to decompose **V** into

$$\mathbf{V} \approx \mathbf{W}\,\mathbf{H} \tag{1}$$

where the basis matrix **W** is $N \times K$ and the coefficient matrix **H** is $K \times M$. The parameter *K* is the number of bases for the columns of **V**. The elements of **W** and **H** are restricted to be non-negative as well. In order for NMF to be effective, *K* must be much smaller than *M* and *N* such that

$$M\,K + K\,N << M\,N\,. \tag{2}$$

Each column of **W** is a basis and it is constrained to sum to unity to avoid scaling ambiguity in the product representation (1). The basis vectors are not restricted to be orthogonal. The only requirement is that they have non-negative elements. NMF represents each column of **V** as a linear combination of the basis vectors. Denoting the *m*-th column of **V** as $\boldsymbol{v}_m$, the *k*-th column of **W** as $\boldsymbol{w}_k$, the *m*-th column of **H** as $\mathbf{h}_m$ whose elements are $h_{k,m}$, then (1) is equivalent to

$$\mathbf{v}_m \approx \mathbf{W}\,\mathbf{h}_m = \sum_{k=1}^{K} h_{k,m}\,\mathbf{w}_k\,. \tag{3}$$

The decomposition (1) is accomplished by minimizing the error function [11]-[12]

$$\zeta(\mathbf{V},\mathbf{WH}) = \sum_{m=1}^{M}\sum_{n=1}^{N} d(v_{n,m},\widetilde{v}_{n,m}) \tag{4}$$

where $v_{n,m}$ and $\widetilde{v}_{n,m}$ are the (n,m)-th element of **V** and $\widetilde{\mathbf{V}}$, and $\widetilde{\mathbf{V}} = \mathbf{W}\,\mathbf{H}$. The term $d(a,b)$ denotes either a distance or divergence measure between *a* and *b*. The most popular distance measure is the Euclidean distance:

$$d(a,b) = (a-b)^2\,. \tag{5}$$

The Kullback-Leibler (KL) divergence is often found to be more suitable for music/audio applications [13]-[14]:

$$d(a,b) = a\ln\frac{a}{b} - a + b \tag{6}$$

where $\ln(\bullet)$ is the natural logarithm operation.

An analytic solution that minimizes (4) does not exist. Perhaps the simplest and most widely used iterative solution is the multiplicative updates proposed by Lee and Seung [1]. The multiplicative updates can be derived using the steepest descent optimization. The update equations for the KL divergence measure are [14]

$$h_{k,m} \leftarrow h_{k,m}\frac{\sum_{n=1}^{N} w_{n,k}\,v_{n,m}/\widetilde{v}_{n,m}}{\sum_{n=1}^{N} w_{n,k}}, \tag{7a}$$

$$w_{n,k} \leftarrow w_{n,k}\frac{\sum_{m=1}^{M} h_{k,m}\,v_{n,m}/\widetilde{v}_{n,m}}{\sum_{m=1}^{M} h_{k,m}}. \tag{7b}$$

Updates using the Euclidean distance measure can be found in [14]. The multiplicative algorithms for both the KL and Euclidean measures have been shown to achieve monotonic convergence and ensure a locally optimal solution. Due to the complexity of the error function (4) and the use of an iterative approach, the decomposition performance is dependent on initialization.

III. Sparsity Promoted NMF

The effectiveness of NMF for various applications is sensitive to the choice of parameter *K*, the number of basis vectors. A *K* value that is too small results in large decomposition errors such that the pair **W** and **H** do not provide a good representation of **V**. A *K* value that is too large restricts the decomposition efficiency and performance.

In order to avoid manual selection of the *K* parameter *a priori*, the sparsity promoted NMF automatically identifies a suitable number of basis vectors. It is achieved by promoting sparseness in the coefficient vector $\mathbf{h}_m$.

Following the approach developed in the algorithm presented in [15], we introduce an *L*-1 norm penalty term on the coefficients to the cost function as shown in (8). Since the coefficients are constrained to be non-negative, the absolute value of the *L*-1 norm is not needed and the complexity is reduced.

$$\zeta(\mathbf{V},\mathbf{WH}) = \sum_{m=1}^{M}\sum_{n=1}^{N} d(v_{n,m},\widetilde{v}_{n,m}) + \sum_{k=1}^{K}\gamma_k\sum_{m=1}^{M} h_{k,m} \tag{8a}$$

$$\gamma_k = \frac{\gamma}{\sum_{m=1}^{M} h_{k,m}^{(old)}}\,. \tag{8b}$$

In (8b), $h_{k,m}^{(old)}$ are the decomposition coefficients from the previous iteration. The parameter γ_k controls the degree of sparsity. A larger γ value yields a smaller number of basis vectors. Note that the sparsity should be enforced over each row of **H** during the minimization process**,** which is achieved by pruning the entire row of **H** in each iteration. The non-negativity constraint should also be imposed to all elements in **W** after each iteration.

We shall start with a large *K* value, and the optimization will automatically yield a matrix **H** with many rows equal to zero. The actual number of basis vectors will be the number of non-zero rows in **H**. Eliminating those zero rows and the corresponding columns in **W** yields a compact NMF representation. In this work, we perform the optimization based on code from [15] to obtain the sparse-promoted NMF results. We refer the readers to [15] for the details of the optimization.

IV. SOURCE SEPARATION

Source separation is the process of obtaining the signals of different sources from a mixture. It is a fundamental problem in many signal processing applications [8, 9, 18].

Let us consider a scenario in which a microphone captures a composite of the acoustic signals from two different sources:

$$x(n) = s_1(n) + s_2(n) . \quad (9)$$

We have ignored the additive noise in the measurement to simplify the development and n is the time index. $s_1(n)$ and $s_2(n)$ are the signal components to be found.

We generate the non-negative matrix $\mathbf{V}$ for decomposition by applying the Short-Time Fourier Transform (STFT) to the mixed signal $x(n)$. Each column of $\mathbf{V}$ is the magnitude spectrum of the STFT in each frame and the number of columns of $\mathbf{V}$ is the number of frames. The magnitude spectrum is symmetric and only the first half is necessary to construct $\mathbf{V}$, resulting in a computation reduction for decomposition.

The idea of using NMF for source separation is that the basis matrix $\mathbf{W}$ can be grouped into two submatrices $\mathbf{W}_1$ and $\mathbf{W}_2$ whose columns belong to the basis components of the two different sources. Accordingly, the rows of $\mathbf{H}$ are partitioned into two blocks $\mathbf{H}_1$ and $\mathbf{H}_2$ that correspond to the bases in $\mathbf{W}_1$ and $\mathbf{W}_2$. This NMF representation yields

$$\begin{aligned} \mathbf{V} &= \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_2 \end{bmatrix} \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix} = \mathbf{W}_1\mathbf{H}_1 + \mathbf{W}_2\mathbf{H}_2 \\ &= \mathbf{V}_1 + \mathbf{V}_2 . \end{aligned} \quad (10)$$

In (10), $\mathbf{V}_1$ and $\mathbf{V}_2$ approximate the magnitude spectra of $s_1(n)$ and $s_2(n)$. Using the phase of the composite signal in different frames, we can reconstruct the two signal components using the inverse Fourier transform.

The approach to separate the basis vectors into two sets is dependent on the applications and the characteristics of the two signals. Considering the human fall detection application in the next Section, we shall use a mixture of a human fall signal and a TV audio signal here for the source separation problem.

The fall signal is burst like and has higher frequency content than a TV audio signal. These properties enable a relatively straightforward grouping of the basis vectors for the two sources. The grouping is based on the Euclidean norms of the rows of $\mathbf{H}$ using the method described in [17].

To illustrate, we create a composite signal mixture by adding a fall signal, a 10-second TV audio signal and white Gaussian noise together. The signal to interference ratio (SIR) and the signal to noise ratio (SNR) are both 10 dB. The fall signal is normalized to unity energy. Fig. 1(a) displays the original clean fall signal. The mixture is shown in Fig. 1(b). In the STFT to generate $\mathbf{V}$, the frame size is 1024 samples at 16kHz sampling, the window function is Hamming and the amount of overlap is 50%. The separated fall signals using NMF and sparse NMF are shown in Figs. 1(c) and (d), where K in both algorithms were set to 60 as recommended in [17] for the same source separation problem. Both algorithms had the same initialization and the number of iterations was 200. Careful inspection shows sparse NMF gives better performance and it estimates the actual number of bases to be 19. The apparently nosier output from sparse NMF around 1 sec. is due to the random initialization in this simulation. Fig. 2 confirms statistically the better performance of sparse NMF.

Fig. 2 shows the Mean Square Error (MSE) between the original and the separated fall signal, where SNR=10dB and SIR=10dB, as the parameter K increases. The MSE is computed over 50 trials each with different random initialization. The MSE for sparse NMF reduces only slightly when K is larger than 20 while the MSE for NMF requires 60 basis vectors to settle at a stable value. Sparse NMF performs consistently better than NMF and the improvement is around 4 dB for K larger than 60.

Fig. 3 shows the actual number of bases estimated by sparse NMF for the same simulation in Fig. 2. The values are the average from the 50 trials. Sparse NMF is able to determine the number of bases automatically and even when K keeps increasing, it remains to settle at the value of 17.

Fig. 4 gives the performance comparison as the SNR increases, when SIR=10dB and K= 60. Unless the SNR is below -14dB, sparse NMF always yields better result.

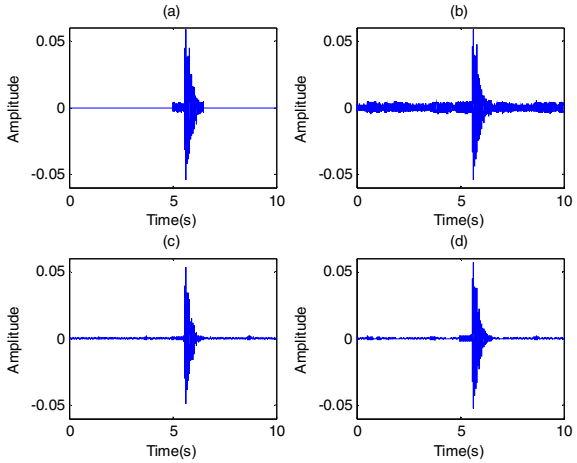

Fig. 1. (a) 'Clean' fall signal. (b) Mixture at SNR=10dB and SIR = 10dB. (c) Extracted fall signal by NMF. (d) Extracted fall signal by sparse NMF.

V. FALL DETECTION WITH SPARSE NMF

Human fall is a major factor that causes injuries for elderly. The injuries could become serious and lead to death if a caregiver does not come to rescue in time. Fall detection is an active research subject for elder care in the United States.

Apart from the video based system, one promising and low cost approach for in-home fall detection is by using the sound generated when a human falls, which can be easily acquired by a microphone. An acoustic system can complement the video based system when occlusion occurs. A typical home environment has many sources of audio interferences,

including radio and TV. The main challenge for acoustics-based fall detection is the need to remove audio interferences.

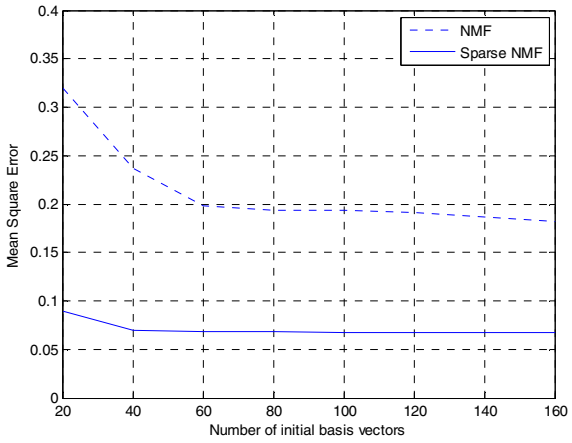

Fig. 2. Mean square error for separated fall signal when SIR=10dB and SNR=10dB.

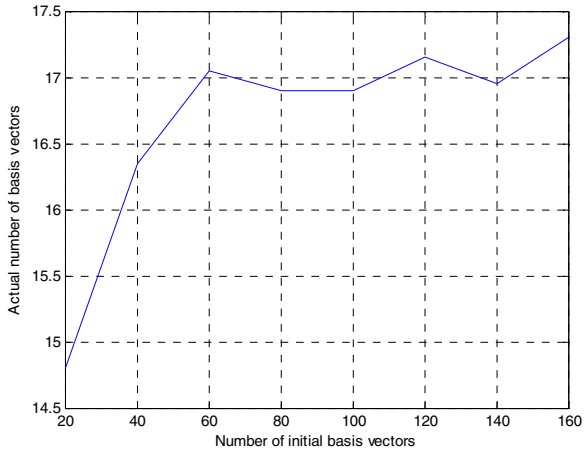

Fig. 3. Actual number of basis vectors found by sparse NMF when SIR=10dB and SNR=10dB.

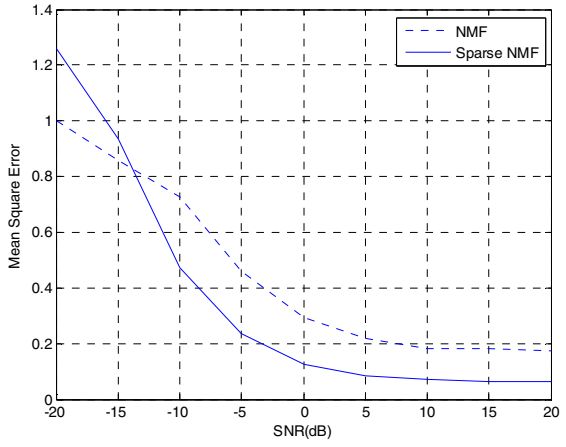

Fig. 4. Mean square error for separated fall signal when SIR=10dB and K=60.

The study in [17] illustrates NMF can separate a fall signal from interference and improve performance of acoustic-based fall detection. We shall apply sparse NMF to the same problem.

TABLE I. DESCRIPTION OF DATASET [18]

Falls (format: 'type'-'trend')	#files	Nonfalls	#file
Balance-Forwards	6	Closing window	6
Balance-Backwards	6	Typing keyboard	6
Balance-Left	6	Key shaking	6
Balance-Right	6	Machine noise	6
Loose consciousness-Forwards	6	Phone ringing	6
Loose consciousness-Backwards	6	Knocking door	6
Loose consciousness-Left	6	Talking	6
Loose consciousness-Right	6	Lying on a bed	6
Trip and fall-Forwards	6	Lying on a couch	6
Trip and fall-Sideways	6	Sitting on a chair	6
Slip and fall-Forwards	6	Normal walking	6
Slip and fall-Sideways	6	Slow walking	6
Slip and fall-Backwards	6	Rubbing	6
Reach fall (chair)-Forwards	6	Dropping books	6
Reach fall (chair)-Left	6	Dropping balls	6
Reach fall (chair)-Right	6	Dropping cans	6
Slide fall-Forwards	6	Dropping wood cases	6
Slide fall-Backwards	6	Dropping plastic bottles	6
Couch fall-Upper body first	6	Rolling a can	6
Couch fall-Hips first	6	Rolling a plastic bottle	6
Total # fall files	120	Total # non fall files	120

The dataset in this experiment was obtained in a simulated home environment. The data collection was approved by the Institutional Review Board at the University of Missouri. The dataset contains 20 different kinds of falls performed by trained stunt actors and 20 typical activities of nonfalls, as illustrated in Table I. Each was performed six times, yielding 120 falls and 120 nonfalls. The fall and nonfall data are 1 second long sampled at a rate of 16 kHz. The detailed descriptions of the fall and nonfall types can be found in [17]-[18].

To simulate an in-home interference environment, we add TV interference and white Gaussian noise to each fall or nonfall signal. The signal mixture for fall detection is 10 seconds long.

We apply the same processing structure as used in [17] where the data is first processed by sparse NMF, which is followed by feature extraction and classification for fall detection. The features are the MFCC and the classifier is the nearest neighbor (NN) [17]. Leave one out cross-validation [18] is used to generate the detection performance in terms of the Receiver Operating Characteristic (ROC) curve.

Fig. 5 shows the detection rate as the false alarm rate increases, where the SIR and SNR are both 10 dB. The performance of NMF and sparse NMF differ when the detection rate is above 32%. This observation indicates that sparse NMF can not only provide an estimate of the correct

number of the basis vectors but also better detection performance.

At lower SIR of -10 dB while keeping SNR at 10 dB, Fig. 6 illustrates the improvement offered by sparse NMF increases. In particular, at 100% detection rate, sparse NMF has a false alarm rate of only 10% while NMF has 79%.

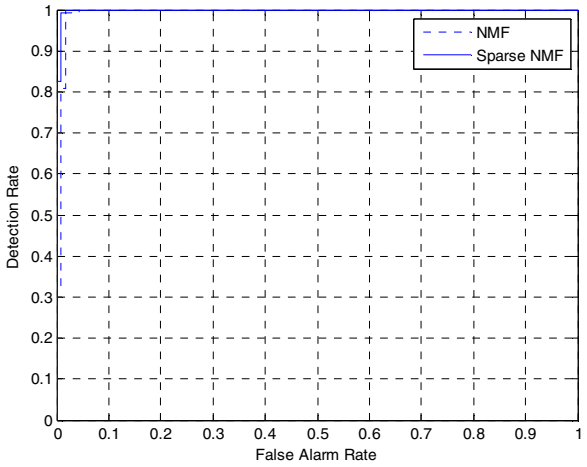

Fig. 5. ROC curve for fall detection using NMF and sparse NMF when SIR=10dB and SNR=10dB.

Table II tabulates the area under the curve (AUC) from the ROC at different SIR values, where the SNR is 10 dB. Table III gives the AUC values when holding SIR to 10 dB and decreasing SNR. The AUC is between 0 and 1 and larger AUC means better performance. In both cases, sparse NMF always yield better results. The improvement is particularly significant when the SNR or SIR becomes small.

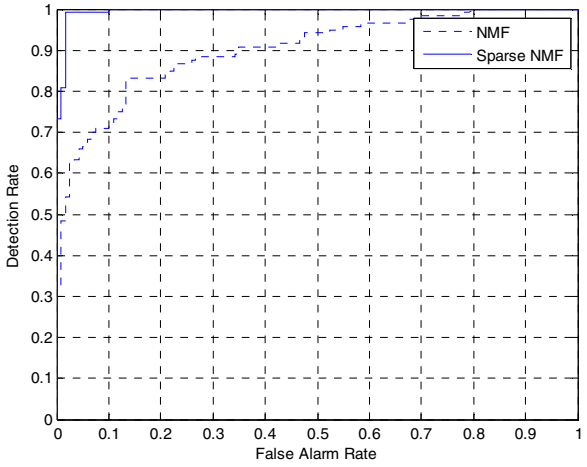

Fig. 6. ROC curve for fall detection using NMF and sparse NMF when SIR=-10dB and SNR=10dB.

TABLE II. AUC of NMF and sparse NMF at different SIR levels and SNR=10dB

		'clean'	10dB	0dB	-10dB	-20dB
NMF	AUC	0.9948	0.9925	0.9878	0.9016	0.4960
Sparse NMF	AUC	0.9989	0.9985	0.9978	0.9955	0.9044

TABLE III. AUC of NMF and sparse NMF at different SNR levels and SIR=10dB

		'clean'	10dB	0dB	-10dB	-20dB
NMF	AUC	0.9978	0.9915	0.9791	0.9042	0.5245
Sparse NMF	AUC	0.9992	0.9988	0.9978	0.9897	0.7026

VI. CONCLUSION

This paper introduces a sparsity-promoting term for NMF. It is shown that the inclusion of this term provides an improvement in performance. The method is particularly useful when the number of basis components K for NMF decomposition is not known *a priori*. The sparse NMF method is applied to source separation and fall detection and demonstrates better performance over NMF.

REFERENCES

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects with nonnegative matrix factorization," *Nature*, vol. 401, pp.788−791, Oct. 1999.

[2] M. Helén and T. Virtanen, "Separation of drums from polyphonic music using nonnegtive matrix factorization and support vector machine," in *Proc. Eur. Signal Process. Conf.*, Istanbul, Turkey, 2005.

[3] E. M. Grais and H. Erdogan, "Regularized nonnegative matrix factorization using Gaussian mixture priors for supervised single channel source separation," *Computer Speech & Language*, vol. 27, no. 3, pp. 746-762, May 2013.

[4] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens*., vol. 51, pp. 2815–2826, May 2013.

[5] M. Berry, M. Browne, A. Langville, P. Pauca, and R. Plemmons. "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics and Data Analysis*, vol. 52, 155–173, 2007.

[6] Patrik O. Hoyer. "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*," vol. 5, pp. 1457–1469, Nov. 2004.

[7] M. Heiler and C. SchnÖrr. "Learning sparse representations by non-negative matrix factorization andsequential cone programming," *J. Machine Learning Research*," vol. 7, pp. 1385–1407, Jul. 2006.

[8] X. Lu, B. Gao, L. C. Khor, W. L. Woo, S. Dlay, W. Ling, and C. S. Chin, "Single channel source separation using filterbank and 2D sparse matrix factorization," *J. Signal, Information Processing*, vol. 4, pp. 186-196, May 2013.

[9] B. Gao, W. L. Woo, and B. W.-K. Ling, "Machine learning source separation using maximum a posteriori nonnegative matrix factorization," in press, *IEEE Trans. Cybern.*, 2014.

[10] J. S. Jang, Audio Signal Processing and Recognition. [online]. Available: http://mirlab.org/jang/books/audioSignalProcessing/

[11] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio, Speech, Lang. Process*., vol. 18, pp. 550 − 563, Mar. 2010.

[12] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, “Efficient algorithms for multichannel extensions of Itakura-Saito nonnegative matrix factorization”, in *Proc. Int. Conf. Acoust., Speech, Signal Process*, Kyoto, Japan, Mar. 2012, pp. 261-264.

[13] C. Févotte, N. Bertin, and J. Durrieu, “Nonnegative matrix factorization with the itakura-saito divergence: with application to music analysis,” *Neural Comput.,* vol. 21, pp. 793-830, Mar. 2009.

[14] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” *Advances in Neural Information Processing Systems*, vol. 13, pp. 556-562, 2000.

[15] A. Zare and P. Gader, “Sparsity promoting iterated constrained endmember detection for hyperspectral imagery,” *IEEE Geosci. Remote Sens. Lett*., vol. 4, pp. 446–450, Jul. 2007.

[16] J.F. Cardoso. “Blind Signal Separation: Statistical Principles,"*Proc. IEEE*, Vol. 86, pp. 2009-2025, Oct. 1998.

[17] Y. Li, K. C. Ho, and M. Popescu, "Efficient source separation algorithms for acoustic fall detection using a Microsoft Kinect" *IEEE Trans. Biomed Eng.*, vol.61, pp. 745-755, Mar. 2014.

[18] Y. Li, K. C. Ho, and M. Popescu, “A microphone array system for automatic fall detection” *IEEE Trans. Biomed Eng.*, vol. 59, pp. 1291-1301, Feb. 2012.

[19] Y. Zigel, D. Litvak, and I. Gannot, “A method for automatic fall detection of elderly people using floor vibrations and sound--proof of concept on human mimicking doll falls,” *IEEE Trans. Biomed Eng.*, vol. 56, pp. 2858-2867, Dec. 2009.