



# Regularized Non-negative Matrix Factorization Using Alternating Direction Method of Multipliers and Its Application to Source Separation

Shaofei Zhang<sup>1</sup>, Dongyan Huang<sup>2</sup>, Lei Xie<sup>1</sup>, Eng Siong Chng<sup>3</sup>, Haizhou Li<sup>2,3</sup>, Minghui Dong<sup>2</sup>

<sup>1</sup>School of Computer Science, Northwestern Polytechnical University, Xi'an, China

<sup>2</sup>Institute for Infocomm Research, A\*STAR, Singapore

<sup>3</sup>School of Computer Engineering, Nanyang Technological University, Singapore

sfzhang@nwpu-aslp.org, huang@i2r.a-star.edu.sg, lxie@nwpu.edu.cn

## Abstract

Non-negative matrix factorization (NMF) aims at finding non-negative representations of nonnegative data. Among different NMF algorithms, alternating direction method of multipliers (ADMM) is a popular one with superior performance. However, we find that ADMM shows instability and inferior performance on real-world data like speech signals. In this paper, to solve this problem, we develop a class of advanced regularized ADMM algorithms for NMF. Efficient and robust learning rules are achieved by incorporating  $l_1$ -norm and the Frobenius norm regularization. The prior information of Laplacian distribution of data is used to solve the problem with a unique solution. We evaluate this class of ADMM algorithms using both synthetic and real speech signals for a source separation task at different cost functions, i.e., Euclidean distance (EUD), Kullback-Leibler (KL) divergence and Itakura-Saito (IS) divergence. Results demonstrate that the proposed algorithms converge faster and yield more stable and accurate results than the original ADMM algorithm.

**Index Terms:** Regularized non-negative matrix factorization, Alternating direction method of multipliers, Beta-divergence, Source separation

## 1. Introduction

In the real-world, it is necessary and desirable to perform both non-negative and sparse decompositions of data such that the underlying components have a physical interpretation. NMF has been found to be a useful decomposition for multivariate data. In particular, NMF has been widely used in audio signal processing such as polyphonic music transcription [1] and source separation [2].

Given a data matrix  $V \in \mathcal{R}^{M \times N}$  with non-negative entries, NMF aims at finding two low-rank matrices  $W \in \mathcal{R}^{M \times K}$  and  $H \in \mathcal{R}^{N \times K}$  ( $K < M, N$ ) such that

$$(W^*, H^*) = \underset{W \geq 0, H \geq 0}{\text{minimize}} D(V|WH^T), \quad (1)$$

where  $D(\cdot)$  represents some measure of divergence. The commonly used measures in NMF are the special cases of beta-divergence [3]: EUD [4, 5, 6, 7, 8], KL-divergence [4, 9] and IS-divergence [10, 11]. Different cost functions can produce significantly different results [12]. In general, the problem is bi-convex in  $W$  and  $H$  separately, so many algorithms adopt an alternating minimization approach to find the locally optimal solution.

The most popular approach is a simple multiplicative update method proposed by Lee and Seung [4], but the conver-

gence of the algorithm to a stationary point has not yet been proved [13]. Under the Euclidean NMF framework, numerous algorithms have been proposed, such as the block principal pivoting (BPP) method [7], the hierarchical alternating least squares (HALS) method [8] and the basic gradient projection (GPSR) method [6]. These algorithms have a common property that every produced limit point is a stationary point [7]. Based on KL-divergence and IS-divergence, several algorithms have also been proposed for specific applications. For example, Virtanen *et al.* considered gamma chains for regularization of KL-NMF [14], while Févotte *et al.* considered the inverse-gamma and gamma Markov chain priors for IS-NMF [11]. All of these mentioned algorithms rely on special properties of EUD, KL-divergence and IS-divergence and are not universal for different applications.

To extend the NMF algorithm to diverse applications using different cost functions, recently, Sun *et al* [15] developed an ADMM based universal update framework which features faster convergence and better accuracy than the state-of-the-art algorithms on synthetic data sets. However, we find that for real non-negative data, e.g., speech spectrum, ADMM results in lower performance than our expectation in terms of stability and accuracy.

In this paper, in order to guarantee the stability of ADMM and the accuracy of NMF, we propose a novel ADMM update framework for NMF. Specifically, we develop a class of advanced regularized ADMM algorithms for NMF to achieve efficient and robust learning (update) rules by incorporating  $l_1$ -norm and the Frobenius-norm regularization. The  $l_1$ -norm promotes sparse solutions which can guarantee a uniqueness of solution for NMF [16]. On the other hand, from the Bayesian perspective, the  $l_1$ -norm can improve the accuracy of the algorithm for NMF if the priori information of a variable follows a Laplacian distribution [16], which coincides with the distribution of some real data, e.g., speech signal [17]. The Frobenius-norm can be used to stabilize the algorithm. As a special case of Tikhonov regularization [18], it can be viewed as a regularization technique to overcome the ill-condition problem. The proposed framework turns the updates of variables  $W$  and  $H$  into a nonnegativity-constrained least squares (NNLS) problem, then numerous Euclidean NMF algorithms can be adopted to update  $W$  and  $H$ . Instead of solving the close-form solution, the new update rules can make the solution much more stable and accelerate the convergence speed.

In this study, we adopt the state-of-the-art Euclidean NMF algorithms including the BPP method [7], the HALS method [8] and the GPSR method [6]. We evaluate this class of advanced ADMM algorithms using synthetic data and real non-

negative data. We compare the performance of ADMM algorithms on source separation using EUD, KL-divergence and IS-divergence as cost functions. Experimental results show that the proposed algorithms have fast convergence and yield more stable and accurate results than the original ADMM algorithm.

## 2. Conventional ADMM for NMF

We note that NMF with different divergences are used in different applications. In order to apply ADMM algorithm to various applications, Sun and Févotte [15] developed an ADMM-based update framework to extend the NMF problem to  $\beta$ -divergence. They proposed to split divergence with dictionary and activation which makes optimization simpler and much more universal for different divergences than the existing algorithms.

### 2.1. Non-negative matrix factorization

In the approach proposed by Sun [15], the NMF can be formulated as follows

$$\begin{aligned} & \text{minimize} && D_\beta(V|X) \\ & \text{subject to} && X = WH^T, \\ & && W = W_+, H = H_+ \\ & && W_+ \geq 0, H_+ \geq 0, \end{aligned} \quad (2)$$

where  $D_\beta(V|X)$  is considered as some measure of  $\beta$ -divergence between  $V$  and  $X$ .  $W$  and  $H$  are the dictionary matrix and activation matrix, respectively. This approach introduces new variables  $W_+$  and  $H_+$  as the nonnegative constraints. A new variable  $X$  splits the divergence with  $WH^T$  which makes optimization problem simpler and also more universal for different divergences. The augmented Lagrangian corresponding to Eq.(2) is as follows.

$$\begin{aligned} L_\rho(X, W, H, W_+, H_+, \alpha_X, \alpha_W, \alpha_H) = & \\ D_\beta(V|X) + \langle \alpha_X, X - WH^T \rangle + \frac{\rho}{2} \|X - WH^T\|_F^2 & \\ + \langle \alpha_W, W - W_+ \rangle + \frac{\rho}{2} \|W - W_+\|_F^2 & \\ + \langle \alpha_H, H - H_+ \rangle + \frac{\rho}{2} \|H - H_+\|_F^2. & \end{aligned} \quad (3)$$

The  $\alpha_X, \alpha_W, \alpha_H$  are three dual variables.  $(W, H, X)$  and  $(W_+, H_+)$  are five primal variables.  $\rho$  is dual step size.

### 2.2. Alternating optimization

In the approach proposed by Sun [15], the updates alternately optimize  $L_\rho$  with respect to five primal variables, followed by gradient ascent in each of the three dual variables. In particular, the updates of  $W$  and  $H$  rely on the differentiability of  $L_\rho$  on  $W$  and  $H$ . We can adopt conventional method of solving extreme value to set the first derivative equal to zero. The root formula and Cardan formula are used to update  $X$ , which corresponds to  $\beta=1$  and  $\beta=0$ , respectively. The updates of  $W_+$  and  $H_+$  adopt methods based on proximity operator [15, 19]. The detailed update rules can be found in [15].

## 3. A class of advanced regularized ADMM for NMF

### 3.1. Regularized NMF model

We incorporate regularization including  $l_1$ -norm and the Frobenius-norm in the universal ADMM framework.  $l_1$ -norm

can promote a uniqueness of solution with sparse constrains. Moreover, the incorporation of the  $l_1$ -norm is equal to bringing in a priori information that the distribution of data is Laplacian [16]. This priori coincides with the distribution of some real data, e.g., speech signal [17]. This can improve the accuracy of the algorithm. The Frobenius-norm can be used to stabilize the algorithm. It can be viewed as a special case of Tikhonov regularization.

The incorporation of  $l_1$ -norm can improve the accuracy and sparsity. Meanwhile, the Frobenius-norm can improve the stability. However, the instability is still a problem while solving the close-form solution. Therefore, we propose a novel ADMM algorithm framework for NMF which turns the updates of variable  $W$  and  $H$  into a non-negative least squares (NNLS) problem. Under this framework, the updating value of  $W$  or  $H$  is non-negative. Therefore,  $W_+$  and  $H_+$  can be discarded. If we consider the  $l_1$ -norm of  $H$  and the Frobenius-norm of  $W$ <sup>1</sup>, the NMF formulation is proposed as follows

$$\begin{aligned} & \text{minimize}_{W \geq 0, H \geq 0} && D_\beta(V|X) + \lambda_{H,l_1} \|H\|_1 + \lambda_{W,F} \|W\|_F^2 \\ & \text{subject to} && X = WH^T, \end{aligned} \quad (4)$$

where the scale parameters  $\lambda_{H,l_1}$  or  $\lambda_{W,F}$  are used to control the strength of regularization.  $X$  is for the same purpose as in Eq.(2). The augmented Lagrangian for Eq.(4) is as follows

$$\begin{aligned} L_\rho(X, W, H, \alpha_X) = & D_\beta(V|X) + \langle \alpha_X, X - WH^T \rangle \\ & + \frac{\rho}{2} \|X - WH^T\|_F^2 \\ & + \lambda_{H,l_1} \|H\|_1 + \lambda_{W,F} \|W\|_F^2. \end{aligned} \quad (5)$$

The proposed ADMM algorithm consists of three primal variables  $(X, W, H)$  and one dual variable  $\alpha_X$ . It simplifies the form of NMF more than the form in Eq.(2) proposed by Sun *et al* [15].

### 3.2. Alternating optimization

According to the ADMM algorithm, Eq.(5) can be solved by the following iterations

$$W^{k+1} := \underset{W}{\operatorname{argmin}} L_\rho(X^k, W, H^k, \alpha_X^k) = \min L_{\rho,W} \quad (6a)$$

$$H^{k+1} := \underset{H}{\operatorname{argmin}} L_\rho(X^k, W^{k+1}, H, \alpha_X^k) = \min L_{\rho,H} \quad (6b)$$

$$X^{k+1} := \underset{X}{\operatorname{argmin}} L_\rho(X, W^{k+1}, H^{k+1}, \alpha_X^k) = \min L_{\rho,X} \quad (6c)$$

$$\alpha_X^{k+1} := \alpha_X^k + \rho(X^{k+1} - W^{k+1}(H^T)^{k+1}). \quad (6d)$$

Applying (5) to (6c), the minimization form on  $X$  is the same with that proposed by Sun [15]. Similarly, we adopt the same update rules as the approach proposed by Sun. For the update rules on  $W$  and  $H$ , we firstly propose a new optimization framework which combines the scale form of ADMM with the properties of regularization item. Under this framework, the minimization problem on  $W$  or  $H$  becomes an NNLS problem. We adopt the state-of-the-art Euclidean NMF algorithms including BPP [7], HALS [8] and GPSR [6] to update  $W$  and  $H$ .

<sup>1</sup>The Frobenius-norm of  $H$  and the  $l_1$ -norm of  $W$  can also be adopted, but here we leave out them for simplify of form. If they are considered, the scale parameter are  $\lambda_{H,F}$  and  $\lambda_{W,l_1}$ , respectively.

### 3.2.1. Proposed optimization framework

**Update on  $W$ :** Applying (5) to (6a) while considering the incorporation of the Frobenius-norm, the minimization problem is stated as

$$L_{\rho,W} = \langle \alpha_X, X - WH^T \rangle + \frac{\rho}{2} \|X - WH^T\|_F^2 + \lambda_{W,F} \|W\|_F^2.$$

The scaled form of  $L_{\rho,W}$  is

$$L_{\rho,W} = \frac{\rho}{2} \|X - WH^T + \mu\|_F^2 + \lambda_{W,F} \|W\|_F^2 + \text{const}, \quad (7)$$

where  $\mu = \frac{1}{\rho} \alpha_X$ . Since const is a constant value which has no effect on the minimization problem, the above form can be rewritten as

$$L_{\rho,W} = \frac{\rho}{2} \|HW^T - (X + \mu)^T\|_F^2 + \lambda_{W,F} \|W^T\|_F^2. \quad (8)$$

According to properties of Frobenius-norm, Eq.(8) can be rewritten as follows

$$L_{\rho,W} = \left\| \begin{pmatrix} \sqrt{\rho/2}H \\ \sqrt{\lambda_{W,F}}E_K \end{pmatrix} W^T - \begin{pmatrix} (\sqrt{\rho/2})(X+\mu)^T \\ 0_{K \times M} \end{pmatrix} \right\|_F^2, \quad (9)$$

where  $E_K$  is a  $K \times K$  identity matrix and  $0_{K \times M}$  is a  $K \times M$  zero matrix.

**Update on  $H$ :** Applying (5) to (6b) while considering the  $l_1$ -norm, the minimization problem becomes

$$L_{\rho,H} = \langle \alpha_X, X - WH^T \rangle + \frac{\rho}{2} \|X - WH^T\|_F^2 + \lambda_{H,l_1} \|H\|_1^2,$$

where we adopt squares of the  $l_1$ -norm of the columns of  $H$  as  $\|H\|_1^2$ , i.e.,

$$\|H\|_1^2 = \sum_{n=1}^N \|h_n\|_1^2,$$

where  $h_n \in R^{1 \times K}$  is a row of  $H$ . According to the same manner of Eq.7 and Eq.8, the  $L_{\rho,H}$  is as follows

$$L_{\rho,H} = \frac{\rho}{2} \|WH^T - (X + \mu)\|_F^2 + \lambda_{H,l_1} \sum_{n=1}^N \|h_n\|_1^2. \quad (10)$$

Then according to the properties of  $l_1$ -norm,  $L_{\rho,H}$  can be rewritten as

$$L_{\rho,H} = \left\| \begin{pmatrix} \sqrt{\rho/2}W \\ \sqrt{\lambda_{H,l_1}}\mathbf{1}_{1 \times K} \end{pmatrix} H^T - \begin{pmatrix} (\sqrt{\rho/2})(X+\mu) \\ 0_{1 \times N} \end{pmatrix} \right\|_F^2, \quad (11)$$

where  $\mathbf{1}_{1 \times K}$  is a row vector containing only ones.

### 3.2.2. Euclidean NMF Optimization

If we can guarantee the non-negativity of  $X + \mu$  in Eq.(9) and Eq.(11), the minimization of Eq.(9) and Eq.(11) can be viewed as a Euclidean NMF problem.  $X + \mu$  can be rewritten as

$$X + \mu = X + \frac{1}{\rho} \alpha_X, \quad (12)$$

where  $X$  is the estimation of  $V$  and is non-negative, and the sign of  $\alpha_X$  is uncertain. Therefore, our algorithm is to adapt the step size  $\rho$  to guarantee the non-negativity of  $X + \mu$ . In particular, we firstly find the index of negative value in  $\alpha_X$ , and extract the value in  $X$  according to the index. Then, we add the corresponding value according to Eq.(12) and set each equation

equal to zero. Finally, we select the maximal value of  $\rho$  as the current value. When the value of  $\rho$  is settled, we can apply the Euclidean NMF algorithms in the proposed framework.

In this study, we adopt the Euclidean NMF algorithms including BPP, HALS and GPSR. The BPP algorithm is an active-set-like method and allows exchanges of multiple variables between working sets. On the other hand, when  $W$  or  $H$  is not necessarily full column rank, the BPP algorithm may break down. However, the regularization item especially for the Frobenius-norm can be adopted to remedy this problem [7]. The HALS algorithm minimizes a set of local cost functions with the same global minima. It allows to formulate a very basic subproblem that can be solved in a closed form. In practice, a zero column could occur in  $W$  or  $H$  during the HALS algorithm. Typically a small number  $\varepsilon \approx 1e^{-16}$  is used to take place zero. Due to this modification, the result of HALS is not sparse. In the same manner as BPP, we incorporate Frobenius-norm in the formulation which can solve this problem [20]. The GPSR algorithm searches from each iterate  $W^k$  or  $H^k$  along the negative gradient of  $D(W^k)$  or  $D(H^k)$  and projects onto the non-negative orthant. Step length is chosen by a back-tracking line search to satisfy Armijo's rule. Compared with the projected method [5], GPSR uses an initial guess for step length. This yields the exact minimizer of  $D$  along this direction and also confines the step length to an interval which protects against the step length too small or too large. In this paper, the original ADMM algorithm is named as ADMM\_ORI and the proposed algorithms are named as ADMM\_BPP, ADMM\_HALS and ADMM\_GPSR, respectively.

## 4. Experimental Evaluations

We evaluate the proposed algorithms, i.e., ADMM\_BPP, ADMM\_HALS and ADMM\_GPSR using both synthetic and real data sets.

We firstly evaluate ADMM\_ORI and the proposed algorithms with synthetic data considering KL-divergence as a cost function and setting  $\rho = 1$ . The synthetic data  $V$  is constructed as  $V = W * H$  where the initial dictionary matrix  $W$  and the activation matrix  $H$  are generated as the absolute values of Gaussian noise<sup>2</sup>, and  $M = 513$ ,  $K = 25$  and  $N = 185$ . For the proposed algorithms, we set  $\lambda_{H,l_1} = 0.01$ ,  $\lambda_{H,F} = 0.01$ ,  $\lambda_{W,l_1} = 0$  and  $\lambda_{W,F} = 0.01$ . As shown in Fig 1, for synthetic data, the proposed algorithms almost have same accuracy and convergence speed with ADMM\_ORI, especially for ADMM\_BPP. On the other hand, compared with ADMM\_ORI, the proposed algorithms are much more stable. We also compare the performance of ADMM\_ORI and the proposed algorithms using real spectrum data, taking KL as a cost function and an setting  $\rho = 0.5$ . As shown in Fig.2, ADMM\_ORI results in instable and low-accuracy solution. In contrast, the proposed algorithms significantly outperform the ADMM\_ORI by a large margin in terms of stability and accuracy. In particular, ADMM\_BPP and ADMM\_GPSR possess a faster convergence speed and achieve better accuracy. Compared with ADMM\_GPSR, ADMM\_BPP and ADMM\_HALS are much more efficient.

We then compare the proposed algorithms with the ADMM\_ORI on source separation using EUD, KL-divergence and IS-divergence as cost functions. We adopt a general source separation pipeline using NMF [2]. We use the signal-to-distortion (SDR) ratio as the evaluation criterion [21]. The same initialization is adopted for all of the ADMM algorithms. We set

<sup>2</sup>In MATLAB notation:  $V = \text{abs}(\text{randn}(M,K)) * \text{abs}(\text{randn}(K,N))$

Table 1: Average SDR for different ADMM algorithms. The last column is the average SDR calculated on all 8 different noises.

ALGORITHMS	destroyerengine	f16	factory1	factory2	m109	pink	volvo	white	Average
ADMM_GPSR_EUD	10.782	10.180	9.206	<b>13.142</b>	<b>13.574</b>	9.741	21.844	11.134	12.450
ADMM_BPP_EUD	10.671	10.148	9.114	12.936	13.254	9.751	22.564	11.160	12.450
<b>ADMM_HALS_EUD</b>	<b>10.883</b>	<b>10.229</b>	<b>9.487</b>	12.749	13.456	<b>9.985</b>	<b>22.584</b>	<b>11.378</b>	<b>12.593</b>
ADMM_ORI_EUD	1.072	4.538	3.704	6.383	7.960	4.755	14.472	6.742	6.203
ADMM_GPSR_KL	9.389	9.320	8.339	<b>11.906</b>	<b>12.739</b>	8.287	19.775	9.042	11.099
<b>ADMM_BPP_KL</b>	<b>9.679</b>	<b>9.350</b>	<b>8.483</b>	11.718	12.511	<b>8.621</b>	19.687	<b>9.614</b>	<b>11.208</b>
ADMM_HALS_KL	7.705	7.006	5.689	11.351	11.892	4.777	<b>20.105</b>	4.586	9.139
ADMM_ORI_KL	5.628	6.642	4.490	4.034	9.658	6.689	9.931	7.019	6.888
ADMM_GPSR_IS	6.314	5.691	4.557	7.198	8.154	4.425	<b>6.978</b>	5.671	6.123
<b>ADMM_BPP_IS</b>	<b>7.350</b>	<b>7.310</b>	<b>6.493</b>	<b>7.830</b>	<b>8.786</b>	<b>7.263</b>	6.548	<b>8.342</b>	<b>7.491</b>
ADMM_HALS_IS	6.993	5.930	4.900	7.406	8.555	4.356	5.817	7.889	6.481
ADMM_ORI_IS	4.984	4.284	3.166	5.553	6.602	2.698	5.209	2.487	4.373

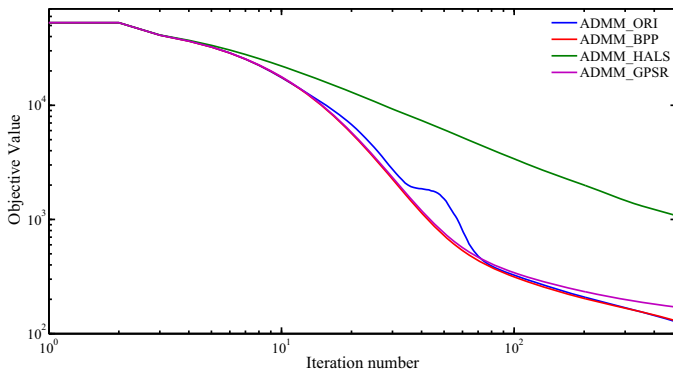


Figure 1: Comparison of the proposed ADMM algorithms with ADMM\_ORI on synthetic data: KL as cost function and  $\rho = 1$ .

$\rho = 0.1$  and the iteration number is 200. We generate noisy speech files (4 seconds in duration per file) by adding different kinds of noises. Specifically, 10 utterances are selected from the TIMIT corpus<sup>3</sup> and we manually corrupt them with 8 typical noises selected from the NOISE92 corpus [22]. We evaluate the ADMM algorithms with different cost functions under different noise conditions. Experimental results are shown in Table 1. For EUD, ADMM\_HALS achieves superior performance as compared with other ADMM algorithms. For KL-divergence and IS-divergence, ADMM\_BPP yields the best results. In particular, ADMM\_GPSR shows a good performance among EUD, KL-divergence and IS-divergence. In contrast, ADMM\_ORI demonstrates inferior performance due to the instability and low-accuracy problem.

## 5. Conclusions

In this paper, in order to guarantee the stability of ADMM while improving the accuracy, we propose a new ADMM based NMF update framework. Specifically, we develop a class of advanced regularized ADMM algorithms for NMF to achieve efficient and robust learning (update) rules by incorporating  $l_1$ -norm and the Frobenius-norm regularization. The  $l_1$ -norm promotes a uniqueness of solutions and represents that a priori on variable is Laplacian. This priori information coincides with the distribution of some real data, e.g., speech signal, which improves the accuracy of algorithm. The Frobenius-norm is a special case of Tikhonov regularization which can be adopted to

<sup>3</sup><https://catalog.ldc.upenn.edu/LDC93S1>

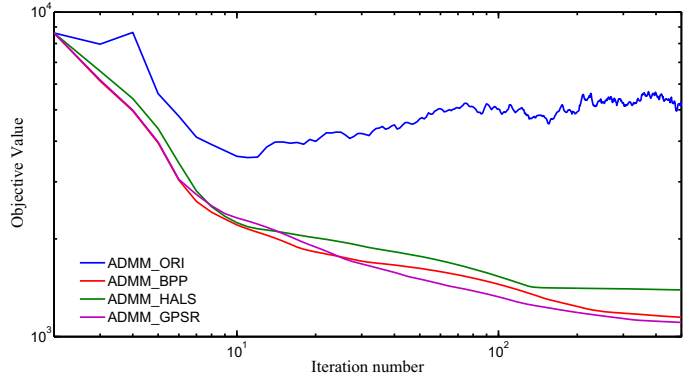


Figure 2: Comparison of the proposed ADMM algorithms with ADMM\_ORI on real spectrum data: KL as cost function and  $\rho = 0.5$ .

stabilize the algorithms especially for ADMM\_BPP and ADMM\_HALS. The proposed framework turns the updates of variable  $W$  and  $H$  into the NNLS problem. In this study, we adopt BPP, HALS and GPSR to update variables instead of solving the close-form solution. The new update rules under the proposed framework can make the solution much more stable and accelerate the convergence speed. We evaluate the ADMM algorithms using synthetic data and real data. We compare the performance of ADMM algorithms on source separation using EUD, KL-divergence and IS-divergence as cost functions. Results demonstrate that the proposed algorithms converge faster and yield more stable and accurate results than the conventional ADMM algorithm.

## 6. Acknowledgements

We thank Dennis L. Sun from Stanford University for valuable discussions. This work was partially supported by the National Natural Science Foundation of China (Grant No. 61175018) and the Seed Foundation of Innovation and Creation for Graduate Students in Northwestern Polytechnical University.

## 7. References

- [1] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA03)*, pp. 177–180, 2003.
- [2] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama, "Complex nmf: A new sparse representation for acoustic signals," *Acoustics, Speech and Signal Processing, ICASSP. IEEE International Conference on*, pp. 3437–3440, 2009.
- [3] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the  $\beta$ -divergence," *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, 2011.
- [4] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in neural information processing systems*, vol. 13, pp. 556–562, 2001.
- [5] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural computation*, vol. 19, no. 10, pp. 2756–2779, 2007.
- [6] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, pp. 586–597, 2007.
- [7] J. Kim and H. Park, "Fast nonnegative matrix factorization: An active-set-like method and comparisons," *SIAM Journal on Scientific Computing*, vol. 33, no. 6, pp. 3261–3281, 2011.
- [8] A. Cichocki and P. Anh-Huy, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 92, no. 3, pp. 708–721, 2009.
- [9] A. R. De Pierro, "On the relation between the isra and the em algorithm for positron emission tomography," *Medical Imaging, IEEE Transactions on*, vol. 12, no. 2, pp. 328–333, 1993.
- [10] Y. Cao, P. Eggermont, and S. Terebey, "Cross burg entropy maximization and its application to ringing suppression in image reconstruction," *Image Processing, IEEE Transactions on*, vol. 8, no. 2, pp. 286–292, 1999.
- [11] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.
- [12] B. King, C. Févotte, and P. Smaragdis, "Optimal cost function and magnitude power for nmf-based speech separation and music interpolation," *Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on*, pp. 1–6, 2012.
- [13] E. F. Gonzalez and Y. Zhang, "Accelerating the lee-seung algorithm for non-negative matrix factorization," *Dept. Comput. & Appl. Math., Rice Univ., Houston, TX, Tech. Rep. TR-05-02*, 2005.
- [14] T. Virtanen, A. T. Cemgil, and S. Godsill, "Bayesian extensions to non-negative matrix factorisation for audio signal modelling," *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pp. 1825–1828, 2008.
- [15] D. L. Sun, "Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence," *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, pp. 6242–6246, 2014.
- [16] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [17] D.-Y. Huang, S. Rahardja, and H. Huang, "Maximum a posteriori based adaptive algorithms," *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*, pp. 1628–1632, 2007.
- [18] M. Fuhry and L. Reichel, "A new tikhonov regularization method," *Numerical Algorithms*, vol. 59, no. 3, pp. 433–445, 2012.
- [19] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [20] J. Kim, Y. He, and H. Park, "Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework," *Journal of Global Optimization*, vol. 58, no. 2, pp. 285–319, 2014.
- [21] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 4, pp. 1462–1469, 2006.
- [22] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247 – 251, 1993.