Signal Separation using Non-negative Matrix Factorization Based on R1-norm

W. kider and M. E. Abd El Aziz[*]

Department of Mathematics, Faculty of Science, Zagazig University, Zagazig Egypt, P.O. Box 44519.
[*]abd_el_aziz_m@yahoo.com

Abstract: Nonnegative Matrix Factorization (NMF) based methods have found use in the context of blind source separation, semi-supervised, and unsupervised learning. These techniques require the use of a suitable cost function to determine the optimal factorization, and most work has focused on the use of least square formulation which is prone to large noise and outliers. In this paper we developed robust NMF algorithm using R_1 -norm which exhibit stability and robustness w.r.t. large noises. This algorithm is as efficient as the algorithms for least square formulations, avoiding the significant computational complexities routinely associated with R_1 -norm formulations. The experimental show that R_1 -NMF can effectively separate the observed that contain outliers better than standard NMF.

[W. kider and M. E. Abd El Aziz. **Signal Separation using Non-negative Matrix Factorization Based on R1-norm.** *Life Sci J* 2012;9(4):703-707] (ISSN:1097-8135). http://www.lifesciencesite.com. 110

Keywords: Blind signal separation, non-negative matrix factorization, R_1 -norm**,** R_1 -NMF.,

1. Introduction

During the past decades, blind source separation (BSS) has become a hot topic in the neural network community (Hyvarinen,1999), signal processing community (Cichocki ,2006), etc. The aim of BSS is to recover the latent sources, without knowing the exact mixing channel. As BSS method needs only the observations, it is quite attractive for signal recovery and system identification. The technology of BSS has many underlying applications, such as signal encryption (Chen,2008), micro-array data analysis (Stadlthanner ,2007), and so on.

On the other hand, many real sources are nonnegative, such as the natural images (Guillamet,2003, Berry,2007, Spratling ,2006) and the microarray data (Stadlthanner,2007))and data analysis, e.g., text analysis (Dhillon ,2007), Brunet,2000) Also, in BSS-based signal cryptosystem, to obtain better decryption accuracy, the plaintext signals are often preprocessed to be nonnegative before encryption (Chen ,2008). Therefore, in these cases, BSS may be solved by the widely used nonnegative matrix factorization (NMF) scheme, which is a powerful tool for data representation. The aim of NMF is to decompose a given dataset (observations) into a mixing matrix and a feature dataset (sources), which are both nonnegative.

Generally speaking, NMF does not rely on the statistical features of the sources, such as independence, nonstationarity, etc. But to solve BSS by NMF practically, it often requires some constraints to conquer the non-uniqueness of the factorization (Amari,2006, Shahnaz,2006). In fact, the constraints are widely discussed in different applications of NMF, where the volume constraint shows great potential to generate a unique result.

The rest of this paper is organized as follow. Section 2, describes the basic BSS and NMF model. Section 3, presents the rotational invariant l_1 - norm , section 4, illustrated the R_1 -NMF algorithm that used to separation , section 5, the performance measures that used, section 6 , the experimental results and discusses the points of our method compared to the standard NMF. Finally, section 7 concludes the paper.

2. BSS AND NMF

The simplest linear model of BSS is:

$$\mathbf{Y} = \mathbf{AX} + \mathbf{V} \qquad (1)$$

Where $\mathbf{Y} = [y_{it}] \in \mathbf{R}^{I\times T}$ is a matrix of observations, $\mathbf{A} = [a_{ij}] \in \mathbf{R}^{I\times J}$ is an unknown mixing matrix, $\mathbf{X} = [x_{jt}] \in \mathbf{R}^{J\times T}$ is a matrix of unknown hidden components or sources, and $\mathbf{V} \in \mathbf{R}^{I\times T}$ is a matrix of additive noise. The objective is to estimate $\mathbf{A}$ and $\mathbf{X}$ rely only on the observed $\mathbf{Y}$. Practical BSS algorithms often need some prior knowledge and assumptions about the sources, such as independence (Hyvarinen, 1999).

When $\mathbf{Y}$, $\mathbf{X}$, and $\mathbf{A}$ are nonnegative, then Eq.(1) is a typical perfect NMF model . Therefore, NMF algorithms can be used to solve BSS (Hyvarinen 1999). Although there is the scaling indeterminacy of columns of $\mathbf{A}$ in NMF/BSS, it does not affect the results essentially (Cichocki,2006). Thus, the columns of $\mathbf{A}$ can be assumed to have unit length. This problem can be solved by choice a suitable cost function and perform alternating minimization similar to Expectation Maximization (EM) approach (Cichocki,2006).

There are many possibilities for defining the cost function D(**Y** ||**AX**), and many procedures for performing its alternating minimization, which lead to several kinds NMF algorithms as: multiplicative, projected gradient, and fixed point (Seung,1999, Cichocki,2006,and Dhillon,2005).

The most widely known adaptive multiplicative algorithm for NMF is based on the squared Euclidean distance (expressed as the squared Frobenius norm) that defines as:

$$D_F(\mathbf{Y}\,\|\,\mathbf{AX}) = \frac{1}{2}\|\,\mathbf{Y}-\mathbf{AX}\,\|_F^2 \quad (2)$$

$$s.t. \;\; a_{ij} \ge 0, x_{it} \ge 0, \forall i, j, t$$

Using a gradient descent approach for cost function Eq. (2) and switching alternatively between the two sets of parameters; we obtain the simple multiplicative update formulas:

$$a_{ij} \leftarrow a_{ij}\frac{[\mathbf{YX}^{\mathrm{T}}]_{ij}}{[\mathbf{AXX}^{\mathrm{T}}]_{ij}+\varepsilon} \quad (3)$$

$$x_{jt} \leftarrow x_{jt}\frac{[\mathbf{A}^{\mathrm{T}}\mathbf{Y}]_{jt}}{[\mathbf{A}^{\mathrm{T}}\mathbf{AX}]_{jt}+\varepsilon} \quad (4)$$

The above algorithm Eq. (3)- Eq. (4), called often the Lee-Seung NMF algorithm (Lee,2001) in which this algorithm is sensitive to the presence of outliers and to avoiding this we will used Rotational l_1- norm (R_1 -norm) as illustrated in the following section.

3. Rotational l_1 - norm

The R_1 -norm of a matrix was first introduced in (Ding, 2006) as rotational invariant l_1 - norm and also used for multi-task learning (Argyriou,2007, Obozinski, 2006) and tensor factorization (Huang, 2008). It is defined as:

$$\|\mathbf{X}\|_{R1} = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} x_{ij}^2\right)^{\frac{1}{2}} \quad (5)$$

While the Frobenius and l_1 -norms are defined as:

$$\|\mathbf{X}\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}^2\right)^{\frac{1}{2}}, \;\|\mathrm{X}\|_{l1} = \sum_{i=1}^{m}\sum_{j=1}^{n}|x_{ij}| \quad (6)$$

In the Euclidean space, the Frobenius norm has a fundamental property rotational invariance. In comparison, the R_1 norm has the following properties: (1) triangle inequality; (2) rotational invariance, as emphasized in (Ding ,2006). Clearly, the Frobenius norm is determined by the sum of the squared elements. In this case, the squared large elements dominate the sum. Consequently, the Frobenius norm is sensitive to outliers. In comparison, the R_1 norm is determined by the sum of elements without being squared. Thus, the R_1 norm is less sensitive to outliers than the Frobenius norm (Ding, 2006). Note that R_1 -norm is different from l_1 sparsification: in sparsification l_1 is a constraint to the objective function while in R_1 is on the main objective function itself (also we used it as constraint to the objective function).

The problem can be solved using a simple yet efficient algorithm called R_1 -NMF (R_1 -norm, Nonnegative Matrix Factorization).

4. An Efficient Algorithm R_1 -NMF

The problem in Eq. (1) can solved by consider the following objective function

$$\mathrm{D}_{R1}(\mathbf{Y}\,\|\,\mathbf{AX}) = \|\,\mathbf{Y}-\mathbf{AX}\|_{R1} + \lambda\|\,\mathbf{X}\|_{R1} + \mu\|\,\mathbf{A}\|_{R1} \quad (7)$$

In order to impose sparsity, we add $\|\,\mathbf{X}\,\|_{R1}$ and $\|\,\mathbf{A}\,\|_{R1}$ where $\mu, \lambda \in [0.01, 0.05]$ are non-negative regularization parameters (Phan, 2009).

By using a gradient descent approach for cost function Eq. (7) we obtain the simple multiplicative update formulas:

$$\frac{\partial \mathrm{D}_{R1}}{\partial \mathbf{X}} = 2\mathbf{CX} + \mathbf{A}^T\mathbf{\Lambda} = 0 \quad (9)$$

Where $\mathbf{C}$ is a diagonal matrix with the i-th diagonal element as:

$$c_{ii} = \frac{1}{2\,\|\,\mathrm{x}_i\,\|} \quad (10)$$

By left multiplying the two sides of Eq. (9) by $\mathbf{AC}^{-1}$ then we obtain:

$$\mathbf{\Lambda} = -2(\mathbf{AC}^{-1}\mathbf{A}^{\mathrm{T}})^{-1}\mathbf{Y} \quad (11)$$

By substitute Eq. (11) into Eq. (9) then:

$$\mathbf{X} = \mathbf{C}^{-1}\mathbf{A}^{\mathrm{T}}(\mathbf{AC}^{-1}\mathbf{A}^{\mathrm{T}})^{-1}\mathbf{Y} \quad (12)$$

The above update rules can be derived by using general multiplicative heuristic formulas as:

$$\mathbf{X} \leftarrow \mathbf{X} \otimes \left(\frac{\mathbf{C}^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{Y}}{\mathbf{AC}^{-1}\mathbf{A}^{\mathrm{T}}+\varepsilon}\right)^{[\omega]} \quad (13)$$

Where $\otimes$ is hadamard (component-wise) product and ω is over-relaxation positive parameter [0.5,2] which used to accelerate the convergence (Cichocki,2006) .By the same rule we can define the update for $\mathbf{A}$ as follow:

$$\mathbf{A} \leftarrow \mathbf{A} \otimes \left(\frac{\mathbf{C}^{-1}\mathbf{Y}\,\mathbf{X}^{\mathrm{T}}}{\mathbf{X}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{X}+\varepsilon}\right)^{[\omega]} \quad (14)$$

For NMF problem with sparsity constraints the above multiplicative algorithm Eq. (13)- Eq. (14) can be summarized in the following pseudo-code Algorithm R_1 -NMF.

Furthermore, it would be very interesting to apply R_1 -NMF algorithm for inverse problems in which matrix $\mathbf{A}$ is known and we need to estimate only matrix $\mathbf{X}$ for ill-conditioned and noisy data.

5. Performance Evaluation

In order to evaluate the performance and precision of R_1 -NMF algorithm we used two criteria (Phan, 2009):

1. Signal -to-Interference-Ratio (SIR).

   Is used to evaluate the ratio between the power of the true signal and the power of the estimated signal

$$SIR = 10\log_{10}\left(\frac{\|\mathbf{X}\|_2^2}{\|\mathbf{X}-\hat{\mathbf{X}}\|_2^2}\right) \qquad (15)$$

2. Peak Signal-to-Noise-ratio (PSNR).

In contrast to SIR, the PSNR estimates the ratio between a maximum possible value of the normalized signal and its root mean squared error, that is:

$$PSNR = 10\log_{10}\left(\frac{R^2 \times T}{\|\mathbf{X}-\hat{\mathbf{X}}\|_2^2}\right) \qquad (16)$$

Where R is a maximum value of the signal and T is the number of samples.

6. Simulations and Experimental Results:

We have conducted extensive simulations with experiments designed specifically to address, how robust is the R_1 -NMF algorithm to noisy mixtures under multiplicative Gaussian noise, additive Gaussian noise. The multiplicative R_1 -NMF algorithm and standard NMF algorithm have been extensive tested on many difficult benchmarks. Illustrative examples are provided to give insight into the multi-layer techniques (Phan, 2009).

a. Example 1

In this example the simulations were performed on the "X_5smooth" dataset (Phan,2009)which contain three sparse and smooth nonnegative signals .We considered the matrix $\mathbf{X}$ with three nonnegative sources (truncating their length to the first 1000 samples) shown in Fig.1. We mixed these sources with a random mixing matrix $\mathbf{A}$ of dimension 3×3, whose elements were drawn independently from a uniform random distribution in the unit interval as displayed in Fig. 1.

After using R_1 -NMF to separate the mixture signals, the result of estimated signals shown in Fig. 2 and the result of standard NMF in Fig. 3. The performance was evaluated with SIR is 36.1622dB and 23.5268dB for R_1 -NMF and standard NMF respectively.

b. *Example* 2

In this example we applied our R_1 -NMF algorithm on a datasets (Phan,2009) (such as "Speech4", "SP_Ex1_Signals"," Speech8" and "X_10rand_sparse"). In which we take three sources from each dataset and we mix these sources with random mixture matrix with noise (5dB), the performance of our algorithm shown in Fig. 4, Fig. 5and Fig. 6 and Table 1.

Where in Fig. 5 the estimated speech (from "Speech4" dataset) by R_1 -NMF and its original in which there are scaling and permutation in the estimated speech, but this Figure cannot tell us the performance of our algorithm so, we can illustrated this performance through using Hinton histogram in which it used to compute and visualize the correlation matrix $\mathbf{G}$ of two matrices $\mathbf{X}$ and $\hat{\mathbf{X}}$, for evaluating the performance of uncorrelated sources (Phan,2009) (if the sources are estimated perfect then $\mathbf{G}$ become diagonal matrix).From Fig. 6 we can see that our algorithm is better than standard NMF.

Table 1 illustrated the performance of our algorithm in which SIR and PSNR is better than standard NMF but it take long time than standard NMF.

Algorithm R_1 -NMF

Input : $\mathbf{Y} \in \mathbf{R}_+^{I\times T}$:input data, J :rank of approximation,
ω :over-relaxation, and λ, μ sparsity degrees

Output: $\mathbf{A} \in \mathbf{R}_+^{I\times J}$ and $\mathbf{X} \in \mathbf{R}_+^{J\times T}$ such that cost function (7) is minimized.

```
begin
    initialization for A and X
    repeat                                   /* update X and A */
        update X by (13)
        update A by (14)
        foreach a_j of A do a_j ← a_j / ||a_j||_p   /* normalize to ℓp unit length */
    until a stopping criterion is met        /* convergence condition */
End
```

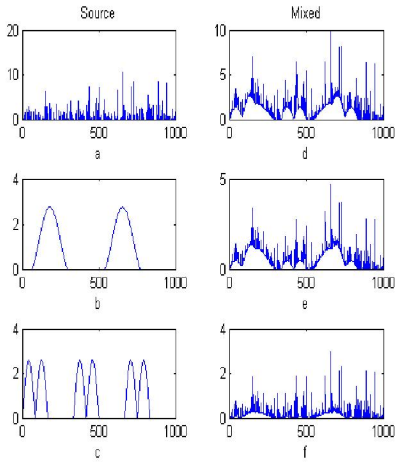


Fig 1: illustration of simulation experiments with three nonnegative sources and their typical mixtures using a randomly generated (uniformly distributed) mixing matrix (a, b, c) sources (nonnegative components) and (d, e, f) mixtures Signals.

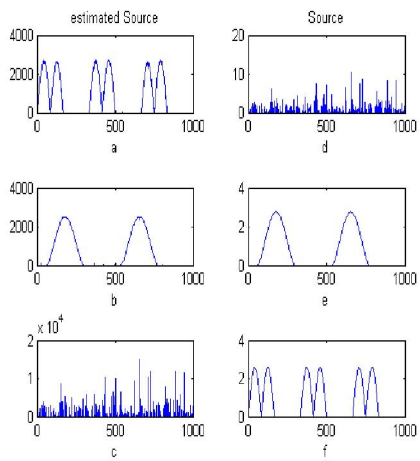
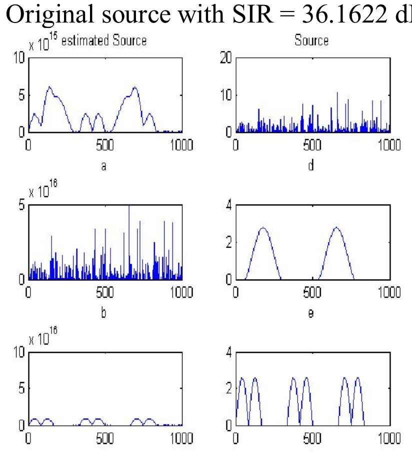


Fig 2: The estimated source by R_1 -NMF and Original source with SIR = 36.1622 dB (with scaling)



.

Fig **3:** The estimated source by standard NMF and Original source with SIR= 23.5268.

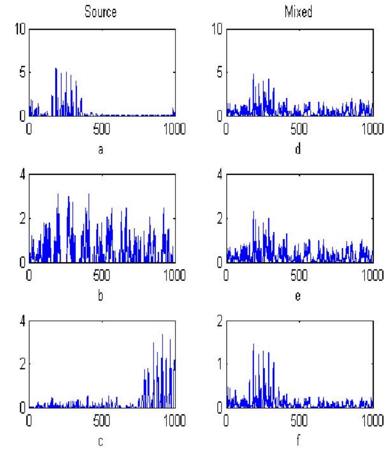


Fig 4: Original source in left and mixture source with noise (5dB) in right

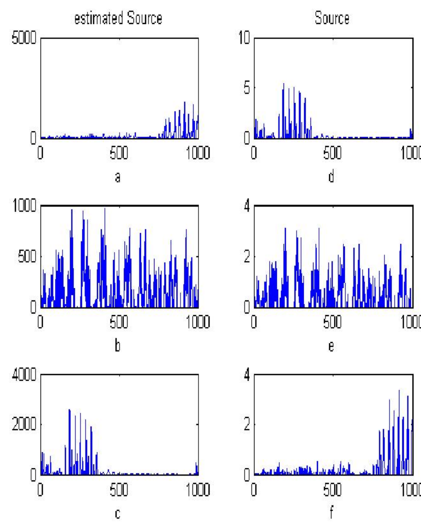


Fig5: Original source in right and estimated source by R_1 -NMF in left.

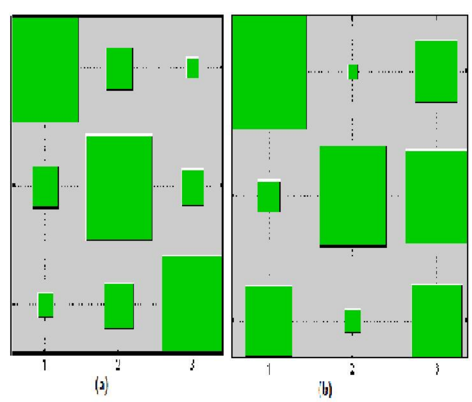


Fig 6: Hinton histogram for G (a) correlation for standard R_1 NMF, (b) correlation for NMF.

Table1: Performance of R_1 -NMF and NMF for four datasets with 4 layers (maximum 1000 iteration / layer).

Noise	R_1 -NMF			NMF		
	PSNR	SIR	Run time	PSNR	SIR	Run time
Speech4	31.53	35.82	27.56	25.15	21.65	15.52
SP_Ex1_Signals	30.33	34.77	11.91	24.83	32.69	8.410
Speech8	28.90	36.64	14.97	23.48	28.73	11.72
X_10rand_sparse	27.22	27.23	9.921	24.80	26.54	8.052

Corresponding Author:
M. E. Abd El Aziz Department of Mathematics, Faculty of Science, Zagazig University, Zagazig Egypt, P.O. Box 44519.abd_el_aziz_m@yahoo.com

7. Discussions

We have introduced the R_1 -NMF algorithm which is flexible and robust cost function and forms a basis for the development of a new class of multiplicative algorithms for NMF. This algorithm allows us to reconstruct (recover) the original signals and to estimate the mixing matrices, even when the observed data are imprecise and/or corrupted by noise. Extensive empirical studies have been performed on group of signals from NMFLAB, to demonstrate performance of our algorithm.

References

1. Argyriou; Evgeniou ,T.; Pontil, M. (2007)."Multi-task feature learning". NIPS, pages 41–48.
2. Amari ,S.; Cichocki ,A.; Zdunek, R. (2006)."New algorithms for non-negative matrix factorization in applications to blind source separation", Proc. IEE E International Conference on Acoustics, Speech, and Signal Processing,ICASSP2006 5Toulouse, France, 14–19, pp. 621–624.
3. Berry, M.; Browne, M.; Langville, A.; Pauca, P.; Plemmons, R. (2007)."Algorithms and applications for approximate nonnegative matrix factorization", Computational Statistics and Data Analysis 52(1) 155–173.
4. Brunet, J.; Tamayo, P; Golub, T.; Mesirov, J. (2000)." Metagenes and molecular pattern discovery using matrix factorization", PNAS 101 , 4164–4169.
5. Chen, Li.; Kwok-Tung, Lo.; Shujun, Li. (2008)." Cryptanalyzing an encryption scheme based on blind source separation," IEEE Trans. Circuits Syst. I, vol. 55, no. 4, pp. 1055–1063.
6. Cichocki, A.;Zdunek, R .; Amari ,S.(2006)."Csiszar's divergences for non-negative matrix factorization: Family of new algorithms", Springer LNCS 3889 ,32–39.
7. Dhillon, I.; Sra, S .(2005)."Generalized nonnegative matrix approximations with Bregman divergences", In Neural Information Proc. Systems (NIPS), Vancouver, Canada, 283–290.
8. Dhillon, I .; Tropp ,J. (2007)."Matrix nearness problems with Bregman divergences", SIAM Journal on Matrix Analysis and Applications 29(4) 1120–1146.
9. Ding, C.; Zhou, D.; Zha, H. (2006)."R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization". Proc. Int' Conf. Machine Learning (ICML).
10. Guillamet, D.;Vitri`a, J.; Schiele, B. (2003)." Introducing a weighted nonnegative matrix factorization for image classification", Pattern Recognition Letters 24. 2447–2454.
11. Huang ,H.; Ding, C.(2008)." Robust tensor factorization using r1 norm". CVPR, pages 1–8, 2008.
12. Hyvarinen, A.(1999)."Fast and robust fixed-point algorithms for independent component analysis", IEEE Trans. Neural New., vol. 10, no. 3, pp. 626–634.
13. Lee , S.; Seung, H. (1999)." Learning the parts of objects by nonnegative matrix factorization", Nature (401) . 788–791.
14. Lee, S. ;Choi,S.(2001)."Application of nonnegative matrix factorization to dynamic positron emission tomography", Proc. 3rd International Conference on Independent Component Analysis and Blind Signal Separation (San Diego, CA), pp. 556–562.
15. Obozinski, G.; Taskar, B.; Jordan ,M. (2006)." Multi-task feature selection". Technical report, Department of Statistics, University of California, Berkeley.
16. Phan ,A.; Cichocki, A.; Zdunek, R.; Amari,S.(2009)." Nonnegative Matrix and Tensor Factorizations"; John Wiley & Sons Ltd.: Chichester, UK.
17. Shahnaz, F.; Berry, M.;Pauca ,P .; Plemmons, R.(2006)." Document clustering using non-negative matrix factorization", Journal on Information Processing and Management 42 , 373–386.
18. Spratling, M. (2006)." Learning image components for object recognition", Journal of Machine Learning Research 7 , 793–815.
19. Stadlthanner, K. ; Lutter, D.;Theis, F.;Lang, E.;Tome, A.; Georgieva, P.; Puntonet, C. (2007)."Sparse nonnegative matrix factorization with genetic algorithms for microarray analysis", in Proc. Int. Joint Conf. Neural Netw. Orlando, FL, pp. 294–299.

9/25/2012