Spectrochimica Acta Part A 86 (2012) 315–319

Contents lists available at SciVerse ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

A novel initialization method for nonnegative matrix factorization and its application in component recognition with three-dimensional fluorescence spectra

Shaohui Yu[a,b], Yujun Zhang[a,*], Wenqing Liu[a], Nanjing Zhao[a], Xue Xiao[a], Gaofang Yin[a]

[a] Key Laboratory of Environmental Optics & Technology, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China
[b] Departments of Mathematics, Hefei Normal University, Hefei 230061, China

ARTICLE INFO

Article history:
Received 23 August 2011
Received in revised form 10 October 2011
Accepted 16 October 2011

Keywords:
Nonnegative matrix factorization
k-means clustering
Singular value decomposition
Independent component analysis
Three-dimensional fluorescence spectra

ABSTRACT

Nonnegative matrix factorization has been widely used in many areas and has been applied for component recognition with three dimensional fluorescence spectra recently. However, nonnegative matrix factorization is a nonconvex programming in the iteration process, thus the solution is dependent on the initial values and consequently not unique. Up to now, an effective global convergent algorithm is still absent. In this work, we propose an initialization scheme based on independent component analysis. Compared with other initialization schemes, the optimal solution of nonnegative matrix factorization based on independent component analysis is much better and it is demonstrated by typical experiments of component recognition with three-dimensional fluorescence spectra.

1. Introduction

Since it was first proposed in 1999 [1], nonnegative matrix factorization (NMF) has attracted more and more attention [2–11]. Now it has been widely used in many areas and very recently, it has been applied to realize component recognition with three-dimensional fluorescence spectra in the study of multi-component mixtures [12]. After decades of development, there have been many algorithms proposed for NMF, such as multiplicative iterative algorithms [2], projected gradient algorithms [3,4] and second-order algorithms [5], etc. With these algorithms, enhanced convergence rates are achieved. However, since NMF is a nonconvex programming, the non-unique property of the solution frequently results in local optima in many algorithms. In order to obtain the global optima, some researchers have tried to improve the initial solutions. For example, Lee and Seung [2] proposed a random initialization method; Wild proposed two initialization methods based on spherical k-means clustering [6] and on structure initialization [7] and Koren and Carmel proposed the SPCA initialization method [8]. Langville et al. compared random initialization, centroid initialization, SVD-centroid initialization, random acol initialization, random C initialization and Co-occurrence initialization [9]. Then, Zheng et al. compared random initialization, PCA initialization, fuzzy clustering initialization and Garbor wavelets initialization [10]. Boutsidis and Gallopoulos proposed NNDSVD initialization based on singular value decomposition [11]. All these previous initialization methods can be classified into two groups, with one based on first-order statistics and the other based on second-order statistics. The initial values obtained by these initialization methods are often far away from the global optima and consequently eventually may lead to inaccurate results. For example, when it is used for component recognition of fluorescence spectra [12], sometimes it will obtain the wrong components [12]. Thus, a much more effective initialization scheme for NMF is still highly required.

Independent component analysis (ICA) is a new technology which was first proposed at the end of 20th century and developed together with blind source separation. Up to now, a set of models and algorithms have been developed for ICA. For example, Comon proposed the concept of ICA and constructed a mathematical model for it in 1994 [13]. Separation algorithms of ICA were presented by Bell and Sejnowski in 1995, which attracted great attention [14]. Hyvarinen presented fast fixed-point algorithm which has high convergence rate and becomes a classical algorithm of ICA [15]. As a statistics method, ICA can divide mixing signals into statistical independent signals according to which we can find out the hidden information in the mixing signals. Because the signals extracted from ICA are related to the characteristics of practical problems, ICA has been applied in many fields in recent years with the development and improvement of its own theory, for example, in biomedical signal processing [16–18], speech separation [14], face recognition [19], and financial data processing [20],

* Corresponding author. Tel.: +86 551 5593147.
E-mail address: yjzhang@aiofm.ac.cn (Y. Zhang).

1386-1425/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.saa.2011.10.042

etc. As an effective preprocessing method, ICA has been applied to the preprocessing of multi-dimensional data, which removes Rayleigh and Raman scattering simultaneously [21]. In this work, as an initialization method, ICA is applied to the preprocessing of the NMF. Compared with other initialization methods, the initial values based on ICA initialization method are closer to the global optima. In our experiment of component recognition with NMF, it is found that the similarity index with ICA as the initialization scheme is much higher than other initialization schemes and thus is beneficial for component recognition.

The rest of the paper is divided into three sections. In Section 2, the basic concepts of NMF and ICA are introduced. In Section 3, the detailed description and analysis of the typical experiments are given and the results of different initialization methods are shown. Finally, a short conclusion is given in Section 4.

2. Theory

2.1. NMF

The basic decomposition model of NMF is as follows

$$Y = AS + E, \tag{1}$$

where $Y \in R^{M\times N}$ is the mixing spectra signal, $A \in R^{M\times K}$ and $S \in R^{K\times N}$ are mixing matrix and source spectra signal, respectively and $E \in R^{M\times N}$ is the noise. With the known measured data Y, the mixing matrix A and the approximation of source spectra signal X computed by formula (1) can be converted to the following optimization problem [22],

$$\min D_F(Y \left\| AX \right.) = \left\| Y - AX \right\|_F^2/2 \quad s.t\, A \geq 0, X \geq 0. \tag{2}$$

This is a nonconvex programming with respect to A and X. The multiplicative iterative update of formula (2) is

$$X = X(A^TY/A^TAX) \quad A = A(YX^T/AXX^T), \tag{3}$$

and the projected gradient update is

$$X = [X - \alpha_X P_X]_+ \quad A = [A - P_A\alpha_A]_+, \tag{4}$$

where $[\ ]_+$ means that the value is forced to be nonnegative, P_A and P_X are the descent directions of A and X respectively and α_A,α_X are the learning rate.

2.2. ICA

According to the law of large number in statistics, a single spectrum signal has greater independence than mixing spectra signal. So, the aim of solving Eq. (1) by ICA is to obtain a demixing matrix W, which can produce independent component X as independent as possible by the following expression,

$$X = WY. \tag{5}$$

The independent component X can be considered as the approximation of source spectra signal S. When ICA based on maximum entropy principle is applied to extract independent component, the contrast function is as follows,

$$J_G(w) = [E\{G(w^TX)\} - E\{G(v)\}]^2, \tag{6}$$

where G is a nonquadratic function and v is a Gaussian variable of zero mean and unit variance. w is one of the M dimensional weight vector in W. According to different problems, G in contrast function has different choices [23]. Solving Eq. (6) by Newton method, the approximate Newton iteration from the kth iteration to the $k+1$th iteration is

$$w_{k+1} = E\{Yg(w_k^TY)\} - E\{g'(w_k^TY)\}w_k, \tag{7}$$

where g and g' are the first derivative and the second derivative of G respectively. To estimate several independent components, we should remove the extracted independent component from the measured mixing spectra signal Y [23] and the procedure in Eq. (7) is repeated until all the independent components are extracted. In the following, we take ICA as the initialization method of NMF and demonstrate its feasibility in component recognition with three dimensional fluorescence spectra.

3. Experiments

3.1. Experiment description

Three representative reagents (tryptophan, tyrosine and pyrocatechol) are selected. The mixtures of the three reagents with different concentrations are prepared (Table 1). Each mixture is scanned by Hitachi F-7000 fluorescence spectrophotometer with excitation wavelength EX 230–320 nm and emission wavelength EM 290–450 nm. In the measurement, the scanning intervals are taken as 4.0 nm and 2.0 nm, respectively for excitation wavelength and emission wavelength and three-way data array Y is obtained.

Table 1
Mixtures of tryptophan, tyrosine and pyrocatechol with different concentration (units: mg/l).

	No 1	No 2	No3	No 4
Tryptophan	0.24	0	0.11	0.18
Tyrosine	0	0.3	0.107	0.125
Pyrocatechol	0.2	0.16	0	0.15

At first, three-way data array Y is preprocessed by ICA. The extracted components X are inaccurate compared with the source spectra signals S, because the fluorescence spectra of the three components are severely overlapping. Taking the results of ICA as initial values, we can compute the solutions by NMF accurately. For comparison, the three-way data array Y is also preprocessed by random initialization [2], k-mean clustering initialization [6] and NNDSVD initialization [11], respectively. The parameter for comparison is the relative error of the approximation spectra signal X

$$R = ||Y - AX||/||Y|| \times 100\%, \tag{8}$$

and the similarity coefficient between approximation spectra signal X and the source spectra signal S is calculated by

$$C = \sum_{i=1}^{M}\sum_{j=1}^{N} x_{i,j}s_{i,j} \Big/ \left(\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} x^2}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} s^2}\right) \quad \begin{matrix} x_{i,j} \in X_n,\ s_{i,j} \in S_k, \\ n = 1,\cdots,K,\ k = 1,\cdots,K \end{matrix}. \tag{9}$$

3.2. Results and discussion

Without loss of generality, tests are performed by two representative algorithms of NMF which are multiplicative iterative updates [2] and projected gradient updates [4]. The maximum of iteration number of NMF is set as 500. In order to avoid the influence of the randomness, each initialization method runs 20 times. The relative errors of multiplicative iterative updates and projected gradient updates are shown in Fig. 1.

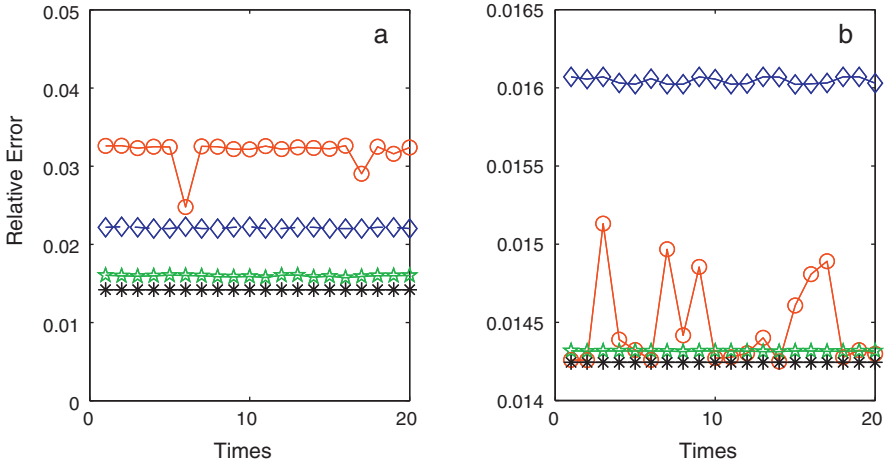

Fig. 1. The relative error of different initialization methods in 20 times: (a) multiplicative iterative updates; (b) projected gradient updates. Random initialization: '–o–', k-mean clustering: '–*–', ICA initialization: '–◇–', NNDSVD initialization: '–☆–'.

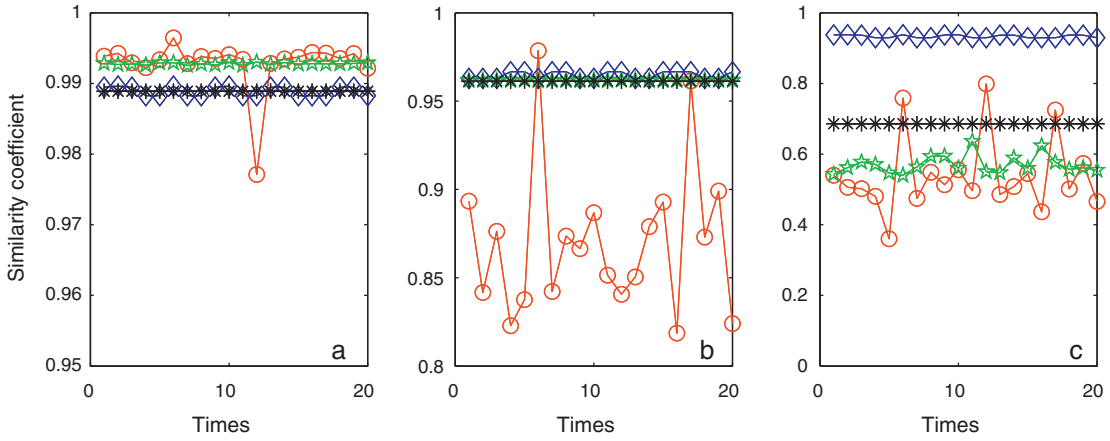

Fig. 2. The similarity coefficients between approximation spectra signals and source spectra signals for multiplicative iterative updates in 20 times. (a) First spectra signal, (b) second spectra signal, (c) third spectra signal. Random initialization: '–o–', k-mean clustering: '–*–', ICA initialization: '–◇–', NNDSVD initialization: '–☆–'.

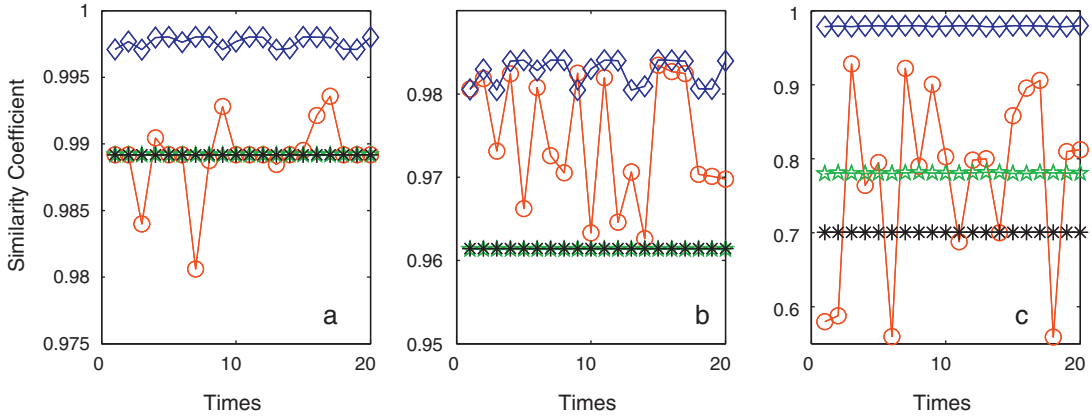

Fig. 3. The similarity coefficients between approximation spectra signals and source spectra signals for projected gradient updates in 20 times: (a) first spectra signal; (b) second spectra signal; (c) third spectra signal. Random initialization: '–o–', k-mean clustering: '–*–', ICA initialization: '–◇–', NNDSVD initialization: '–☆–'.

It can be seen from Fig. 1 that the relative errors of k-mean clustering initialization, NNDSVD initialization and ICA initialization change very little during the iterations of 20 times. This is mainly because these three initialization methods compute the same measured spectra signal each time. On the contrary, the relative error of random initialization changes greatly in the iteration of different times, as the initial value produced by random initialization method has nothing to do with the measured spectra signal. At the same time, we can see the relative error of k-mean clustering initialization is the lowest for both multiplicative iterative updates (see Fig. 1(a)) and projected gradient updates (see Fig. 1(b)). Though the relative error of ICA initialization is the highest for projected gradient updates (see Fig. 1(b)), the relative errors of four initialization methods change between 0.014 and

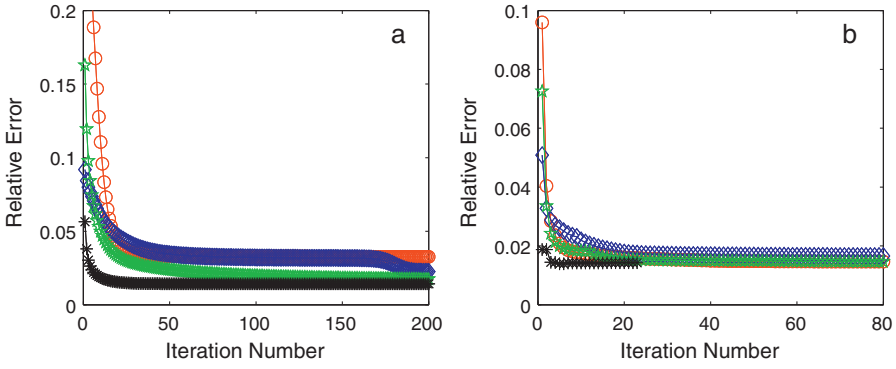

Fig. 4. The relative error with the increasing iteration number: (a) multiplicative iterative updates; (b) projected gradient updates. Random initialization: '–o–', k-mean clustering: '–*–', ICA initialization: '–◊–', NNDSVD initialization: '–☆–'.

0.0165, thus the difference is negligible. Relative error only shows the accuracy of the iterative updates and the relative error will become smaller and smaller by increasing iteration number. To show how much different initialization methods affect the component recognition of three-dimensional fluorescence spectra, we compare the similarity coefficient between approximation spectra signals X and source spectra signals S by different initialization methods. The similarity coefficients of three spectra signals are shown in Figs. 2 and 3, respectively.

From Fig. 2, it can be seen that the similarity coefficients of three spectra signals based on ICA initialization are quite large and one of the similarity coefficients based on k-mean clustering initialization and NNDSVD initialization is smaller than that based on ICA initialization (see Fig. 2(c)). Two of the similarity coefficients based on random initialization are quite smaller (see Fig. 2(b) and (c)), and the similarity coefficient of the first spectra signal will occasionally produce small value in multiple times (see Fig. 2(a)).

The similarity coefficients in Fig. 3 show that the algorithms based on ICA initialization result in much higher similarity coefficient and thus recovers the single spectra signals the best, which implies that the global optima is found. This is what is expected in component recognition with three dimensional fluorescence spectra since the similarity coefficient should be as close to 1 as possible if one component in the decomposed spectra (approximation spectra signals X) is identified as one in the reference spectra (source spectra signals S). By comparing Fig. 2 and Fig. 3, we can draw a conclusion that different iterative algorithms of NMF based on ICA initialization can accurately find the source spectra signal, namely, the global optima, though the descent rate is a little slower than other schemes.

Finally, the relative errors with increasing iteration number are shown in Fig. 4. It is found that the relative error magnitudes of the initial value obtained by four different initialization methods satisfy the inequality $R_{\text{rand}} > R_{\text{NNDSVD}} > R_{\text{ICA}} > R_{\text{k-mean}}$. The main reason is that random initialization has nothing to do with the measured spectra signal Y, so the relative error is the largest. Although NNDSVD initialization produces the initial value according to the measured spectra signal, the orthogonality of these initial values results in a big difference between the approximation spectra signals and the source spectra signals. k-mean clustering initialization gets the initial value by the clustering analysis of the measured spectra signal, so the relative error is the smallest. Although the initial value obtained by ICA initialization method does not give the smallest relative error, it is the closest to the source spectra signal and thus it is the best for component recognition.

4. Conclusion

It is very important to find a good initialization method so as not to get into the local optima in NMF. We analyzed the existing initialization methods in this paper. According to statistics, these initialization methods can be classified into two classes. One is based on first-order statistics of the measured spectra signals, which is represented by k-mean clustering initialization method and the other is based on second-order statistics of the measured spectra signals, which is represented by NNDSVD initialization method. The advantage of the first class is that the relative error is small and the disadvantage is that it cannot reveal the characteristics of the source spectra signals because it only averages the measured spectra signals. The disadvantage of the second class lies in the fact that the approximate spectra signals are much different from the source spectra signals due to the orthogonality of the initial values. The ICA initialization method proposed in this paper preprocesses the measured spectra signals by high-order statistics, which can extract independent components and the initial value is closer to the source spectra signals. Compared with other initialization methods, ICA initialization method produces higher similarity coefficients between the approximation spectra signals and the source spectra signals (Figs. 2 and 3). As a result, it can recover the source spectra signals very well. Although the descent rate of relative error with the ICA initialization is a little slower than other initialization methods, it behaves much better in component recognition with NMF. As a matter of fact, the ICA-based NMF does not only provide an important way for component recognition of three-dimensional fluorescence spectra, but also it may find wide application in many other areas.

Acknowledgements

The work was supported by National 863 Program of China (2007AA061502), Anhui Provincial Natural Science Foundation (1108085J19), the Foundation of Director of Anhui Institute of Optics and Fine Mechanics (Y03AG31144) and the University Natural Science Foundation of Anhui Province (KJ2011B134).

References

[1] D.D. Lee, H.S. Seung, Learning of the parts of objects by non-negative matrix factorization, Nature 401 (1999) 788–791.
[2] D.D. Lee, H.S. Seung, Algorithms for nonnegative matrix factorization [A], in: Advances in Neural Information Processing Systems [C], MIT Press, Cambridge, 2000, pp. 556–562.
[3] J.M. Bardsley, C.R. Vogel, Nonnegatively constrained convex programming method for image reconstruction, SIAM J. Sci. Comput. 4 (2004) 1326–1343.

[4] Ch-J. Lin, Projected gradient methods for non-negative matrix factorization, Technical Report, Department of Computer Science, National Taiwan University, 2005.
[5] R. Zdunek, A. Cichocki, Non-negative matrix factorization with quasi-Newton optimization, Lecture Notes in Artificial Intelligence, 4029 (2006) 870–879.
[6] S. Wild, Seeding non-negative matrix factorizations with the spherical k-means clustering, Master of Science Thesis, University of Colorado, 2003.
[7] S. Wild, J. Curry, A. Dougherty, Improving non-negative matrix factorizations through structured initialization, Pattern Recognit. 37 (2004) 2217–2232.
[8] Y. Koren, L. Carmel, Robust linear dimensionality reduction, IEEE Trans. Visual. Comput. Graphics 10 (2004) 459–470.
[9] A.N. Langville, C.D. Meyer, R. Albright, Initializations for the nonnegative matrix factorization, in: Proceeding of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006.
[10] Z.L. Zheng, J. Yang, Y.T. Zhu, Initialization enhancer for non-negative matrix factorization, Eng. Appl. Artif. Intell. 20 (2007) 101–110.
[11] C. Boutsidis, E. Gallopoulos, SVD based initialization: a head start for nonnegative matrix factorization, Pattern Recognit. 41 (2008) 1350–1362.
[12] S.H. Yu, Y.J. Zhang, W.Q. Liu, N.J. Zhao, X. Xiao, G.F. Yin, Component recognition with three dimensional fluorescence spectra based on non-negative matrix factorization, J. Chemom., doi:10.1002/cem.1404.
[13] P. Comon, Independent component analysis, a new concept? Sig. Proc. 36 (1994) 287–314.
[14] A.J. Bell, T.J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, Neural Comput. 7 (1995) 1129–1159.
[15] A. Hyvarinen, A fast fixed-point algorithm for independent component analysis, Neural Comput. 9 (1997) 1483–1492.
[16] S. Makeig, A.J. Bell, T.P. Jung, et al., Independent component analysis of electroencephalographic data, Adv. Neural Inform. Proc. Syst. 8 (1996) 145–151.
[17] R. Vigario, V. Jousmaki, M. Hamalainenet, et al., Independent component analysis for identification of artifacts in magnetoenceph-alographic recordings, Adv. Neural Inform. Proc. Syst. 10 (1998) 229–235.
[18] M.J. McKeown, S. Makeig, G.G. Brown, et al., Analysis of fMRI data by blind separation into independent spatial components, Hum. Brain Mapp. 6 (1998) 160–188.
[19] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, Face recognition by independent component analysis, IEEE Trans. Neural Networks 13 (2002) 1450–1464.
[20] A.D. Back, A.S. Weigend, A first application of independent component analysis to extracting structure from stock returns, Int. J. Neural Syst. 8 (1997) 473–484.
[21] D.J-R. Bouveresse, H. Benabid, D.N. Rutledge, Independent component analysis as a pretreatment for parallel factor analysis to eliminate artifacts from multiway data, Anal. Chim. Acta 589 (2007) 216–224.
[22] A. Cichocki, R. Zdunek, A-H. Phan, S. Amari, Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation, Wiley, 2009.
[23] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Trans. Neural Networks 10 (1999) 626–634.