Journal of Sciences, Islamic Republic of Iran 26(3): 273 - 279 (2015) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

A Projected Alternating Least square Approach for Computation of Nonnegative Matrix Factorization

M. Rezghi[*1], and M. Yousefi[2]

[1] *Department of Computer Science, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Islamic Republic of Iran*
[2] *Department of Applied Mathematics, Faculty of Sciences, Sahand University of Technology, Tabriz, Islamic Republic of Iran*

Received: 23 July 2014 / Revised: 16 July 2015 / Accepted: 8 September 2015

Abstract

Nonnegative matrix factorization (NMF) is a common method in data mining that have been used in different applications as a dimension reduction, classification or clustering method. Methods in alternating least square (ALS) approach usually used to solve this non-convex minimization problem. At each step of ALS algorithms two convex least square problems should be solved, which causes high computational cost. In this paper, based on the properties of norms and orthogonal transformations we propose a framework to project NMF's convex sub-problems to smaller problems. This projection reduces the time of finding NMF factors. Also every method on ALS class can be used with our proposed framework.

Keywords: Nonnegative matrix factorization; Alternating least squares; Initialization; Orthogonal transformation.

Introduction

Given a nonnegative matrix $A \in \mathbb{R}^{n\times m}$ and a pre-specified positive integer $k \ll \min\{n, m\}$, non-negative matrix factorization (NMF) finds two non-negative matrices $W \in \mathbb{R}^{n\times k}$ and $H \in \mathbb{R}^{k\times m}$ such that $A \approx WH$. (1.1)

Here, every column of the data matrix A is approximated by a positive linear combination of columns of the positive matrix W. So each column of H is a reduced representation of the corresponding column of the data matrix A. The main advantage of this factorization is positivity of its factors, which gives a simple interpretation of positive data. Therefore, NMF differs from previous methods like principal component analysis (PCA) [1].

Due to this advantage, this factorization have been used in different applications such as data mining, computer vision, bioinformatics and most others as a dimension reduction, classification or clustering method [2-4].

Mathematically NMF factorization can be modeled as the following constraint non-convex minimization problem $\min_{W,H\geq 0} \|A - WH\|$, (1.2)

where $\|.\|$ denotes the Frobenius norm.

It's more than a decade *Paatero* and *Tapper* [5] proposed an algorithm for NMF named positive matrix factorization. But NMF took off after introducing multiplicative updates rules (MUR) of *Lee* and *Seung*

[*] Corresponding author: Tel: +982182884480; Fax: +9821 82883493; Email: Rezghi@modares.ac.ir

[6-7].

Recently, the main proposed NMF methods are based on the alternating nonnegative least squares (ANLS) framework. Methods like projected Quasi-Newton [8], projected gradient [9], active set [10], [11] and block principal pivoting [12] are well-known examples in this approach. An inexact form of ANLS is alternating least squares (ALS) algorithm [13] which is a simple and fast algorithm for NMF.

At each step of each method in alternating least square class, two convex nonnegative least square sub-problems should be solved, which is very time consuming for real world and large scale applications.

In this paper we propose a method that project the NMF sub-problems to small appropriate spaces and use the solution of these reduced sub-problems as approximations of NMF factors.

It's clear that solving these projected sub-problems is faster than solving original sub-problems. By choosing suitable projection (such as singular vectors of SVD), we can hope to have a good approximations of solutions of the original problem.

Since nonnegative matrix factorization is an non-convex problem, so initialization of factor matrices for NMF algorithms is very important. Initialization techniques should be able to provide good suggestions (starting points) in a reasonable time. We show that projected version of ALS can be used as an initialization method for other NMF algorithms.

Experimental results on some well-known data sets, confirm the power of proposed method in speeding up the ALS and ANSL based algorithms.

The rest of this paper is organized as follows. In the second section, we provide a brief overview of several existing NMF algorithms. In third section, a framework based on projection for computation of NMF will be presented. Fourth section, contains some experimental results on well know data sets. Finally, the conclusion is given in the last section.

Materials and Methods

In this section, we provide a brief overview of the several existing NMF algorithms and then introduce our proposed algorithm based on space projection.

1. Methods for computing NMF factorization

Although there are different methods in computation of nonnegative matric factorization but they can be categorized to three main classes. They are multiplicative update rules, alternating non-negativity constrained least squares (ANLS) and gradient approaches.

One of the most commonly utilized NMF algorithms developed by *Lee* and *Seung* [6], [7] is based on multiplicative update rules of W and H as follows:

$$H_{bj} \leftarrow H_{bj} \frac{(W^T A)_{bj}}{(W^T WH)_{bj}}, \quad (2.1)$$

where $1 \le b \le k$ and $1 \le j \le m$. Also for $1 \le i \le n$ and $1 \le a \le k$ we have

$$W_{ia} \leftarrow W_{ia} \frac{(AH^T)_{ia}}{(WHH^T)_{ia}}. \quad (2.2)$$

These rules are a variation of the gradient descent method. Many various algorithms have been developed for solving Eq. (1.2) after *Lee* and *Seung*'s popular algorithm.

The original NMF problem of Eq. (1.2) is not convex with respect to both variables W and H at the same time. But this with respect to one of factors is convex and easy to solve. Methods in Alternative nonnegative least square class (ANLS), use this fact in computation of NMF factors. So methods in the second ANLS approach, reformulate the non-convex problem to two convex sub-problems:

$$\min_{H \ge 0} \| A - WH \|^2, \quad (2.3)$$

With fixed W, and

$$\min_{W \ge 0} \| A^T - H^T W^T \|^2, \quad (2.4)$$

where H is fixed. This framework which is illustrated in the Algorithm 1, alternatively fixes one matrix factor and improves the other until a convergence criterion is satisfied. Different algorithms to solve these subproblems give different algorithms in ANLS class.

***Algorithm 1.* ANLS**

1: input: $W \in \mathbb{R}^{n \times k}$
2: output: $W \in \mathbb{R}^{n \times k}$ and $H \in \mathbb{R}^{k \times m}$
3: initialize $W \ge 0$
4: while some condition holds do
5: solve $\min_{H \ge 0} \| A - WH \|^2$
6: solve $\min_{W \ge 0} \| A^T - H^T W^T \|^2$
7: end while

ANLS can be considered as a 2-block coordinate descent method [14-15]. According to presented result in [16], any limit point of the sequence $\{W, H\}$ generated by ANLS is a stationary point of Eq. (1.2). Achieving a stationary point is a goal that should be mentioned in most algorithms of NMF.

It's important to know how we can solve sub-problems Eq. (2.3) and Eq. (2.4). There are several algorithms that solve sub-problems exactly. Kim and Park in [10] described applying a fast combinatorial NNLS (FC-NNLS) algorithm suggested by Benthem and Keenan [17] in solving constrain sub-problems Eq. (2.3) and Eq. (2.4). In FC-NNLS all the columns of current solution with the same passive sets are identified and collected together. FC-NNLS computes the appropriate pseudo-inverses and solves an unconstrained least squares problem for the passive variables. We address [17] for more details of FC-NNLS algorithm. But a limitation of the active set method is that typically only one variable is exchanged between working sets. Recently, Kim and Park have applied block principal pivoting method that allows the exchanges of multiple variables with a goal of finding the optimal active and passive sets faster [12].

Projected gradient methods is another approach to solve convex sub-problems Eq. (2.3) and Eq. (2.4). In basic gradient based algorithm, these sub-problems are considered as unconstraint least squares problem. In each iteration of this algorithm calculating gradient, choosing step size and projecting the update on the non-negative space should be done. Another variant of this approach can be found in in [9].

In inexact form of ANLS i.e. that will be denoted by ALS, the sub-problems Eq. (2.3) and Eq. (2.4) are solved as an unconstrained least squares problem. To enforce non-negativity in the solutions of normal equations, every negative element is set to zero. The ALS (inexact ANLS) illustrates in Algorithm 2. Although this inexact algorithm doesn't produce better approximation errors but it spends significantly less time. To get the other algorithms or their variants, see references [10], [13] and the references therein.

Algorithm 2. Inexact ANLS (ALS)

1: input: $W \in \mathbb{R}^{n\times k}$

2:Output: $W \in \mathbb{R}^{n\times k}$ and $H \in \mathbb{R}^{k\times m}$

3: Initialize $W \geq 0$

4: While some condition holds do

5: Solve $\min_{H\geq 0} \| \mathrm{A} - \mathrm{WH} \|^2$ and

set all negative elements in H to 0.

6: Solve $\min_{W\geq 0} \| \mathrm{A}^T - \mathrm{H}^T \mathrm{W}^T \|^2$ and

set all negative elements in W to 0

7: End while

1.2 ALS Approach as an Initialization Method for ANLS Algorithms

Since in ALS algorithms only unconstraint least square problems should be solved, they can be implemented very fast. But their precision is less than methods based on ANLS approach.

On the other hand the solution and convergence provided by NMF algorithms usually are sensitive to the initialization of W and H. So, by having a good initialization we have faster convergence to an improved local minimum.

For solving this problem one can use an ANLS algorithms several times and set its best solution as NMF factorization result. But this is very time consuming. Rapid implementation of ALS algorithms enable us to use them as an initialization algorithm for ANLS algorithms. Recently, a robust initialization based on ALS algorithms is illustrated in [14] that can be seen in Algorithm 3.

In this algorithm we use one simple and rapid NMF algorithms such ALS based methods.

Here we run these algorithms several time with few iteration and different initial values.

Then use the best solution of these implementations as an initialization of powerful ANLS algorithms.

Algorithm 3. ALS-based initialization

1: Input: $W_0 \in \mathbb{R}^{n\times k}$ and $H_0 \in \mathbb{R}^{k\times m}$

2: Output: W and H

3: for $j = 1 : \mathrm{c}$ do

4: Initialize randomly W_0 and H_0

5: $\{\mathrm{W}_j, \mathrm{H}_j\} = \mathrm{ALS}(\mathrm{A}, \mathrm{W}_0, \mathrm{H}_0, \mathrm{d})$

7: end for

8: $j_{\min} = \arg\min_{1\leq j\leq c} \|\| \mathrm{A} - \mathrm{W}_j \mathrm{H}_j \|$

9: $W = W_{j_{\min}}$ and $H = H_{j_{\min}}$

2.1 A Projection based ALS algorithms for computation of NMF

In this section we propose a fast method for computing nonnegative matrix factorization, based on projection of its sub-problems into appropriate subspaces.

Let $\mathrm{Q} \in \mathbb{R}^{n\times n}$ be an orthogonal matrix. Since Frobenius norm is invariant under orthogonal multiplication, for arbitrary matrix $A \in \mathbb{R}^{n\times m}$, we have $\| \mathrm{A} \|^2 = \| \mathrm{Q}^T \mathrm{A} \|^2$. (3.1)

Now, if $Q=[Q_1,Q_2]$ where $Q_1 \in \mathbb{R}^{n\times l}$ and $Q_2 \in \mathbb{R}^{n\times(n-l)}$, for $l<n$, equation (3.1) becomes

$$\|A\|^2=\|Q_1^T A\|^2+\|Q_2^T A\|^2. \quad (3.2)$$

For some appropriate transforms such as singular matrices in SVD decomposition [18], Cosine transformation and some other transformations, the first term $\|Q_1^T A\|^2$ could be a good approximation of $\|A\|^2$, even for small dimension l. For example, consider the data matrix A as the CBCL data set of the size 361×2429 [19]. Here we use the first left singular vectors of A as a projection matrix. Figure 1 show the relative error between $\|A\|^2$ and $\|Q_1^T A\|^2$ defined as $Re(l)=(\|A\|^2-\|Q_1^T A\|^2)/\|A\|^2$,

for different values of l. This figure demonstrates that even for small vales of the index l, $\|Q_1^T A\|^2$ is an appropriate approximation of $\|A\|^2$ .Due to this fact we can substitute the sub-problems of NMF algorithms with small appropriate projected sub-problems and hope that its solution could approximate the solution of the original sub-problems.

Now consider the first sub-problem

$$\min \|WH-A\|^2.$$

By appropriate projection matrix Q_1 this minimization problem can be substituted with the following closed and smaller minimization problem

$$\min \|\overline{W}\,H-\overline{A}\|^2,$$

where,

Algorithm 4. Transformed ALS (TALS)

1: Input:
  Data matrix $A\in\mathbb{R}^{n\times m}$
  Initial matrix $W\in\mathbb{R}^{n\times k}$
  Projection matrices $Q_1\in\mathbb{R}^{n\times l_1}$ and $P_1\in\mathbb{R}^{m\times l_2}$.
2: Output: $W\in\mathbb{R}^{n\times k}$ and $H\in\mathbb{R}^{k\times m}$
3: Compute $\overline{A}=Q_1^T A$ and $\overline{\overline{A}}=P_1^T A^T$
4: While some condition holds do
  a) $\overline{W}=Q_1^T W$
  b) solve $\min\|\overline{W}\,H-\overline{A}\|$
  c) compute $\overline{\overline{H}}=P_1^T H^T$
  d) solve $\min\|\overline{\overline{H}}\,W^T-\overline{\overline{A}}\|$
10: end while

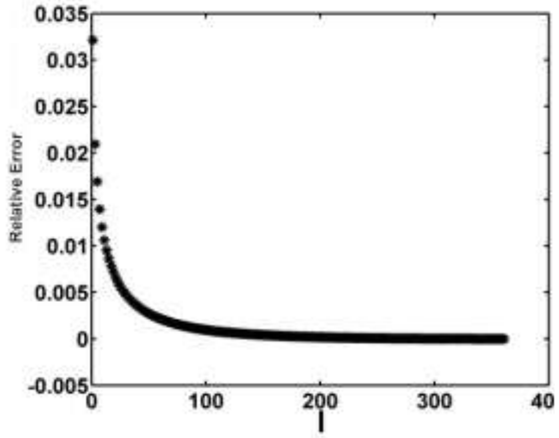


Figure 1. Relative errors of approximation of CBCL database for different values of l.

$$\overline{W}=Q_1^T W\ ,\ \overline{A}=Q_1^T A.$$

It is clear that the complexity of solving this small problem is more less than the original problem. The same process can be done for the second sub-problem. By this approach the transformed version of ALS algorithms (TALS) can be summarized in Algorithm 4.

This approach is very fast and can be used to find NMF factorization or as an initialization method for ANLS algorithms.

Results

In this section to show the performance of our proposed method, Some experimental results on some well-known datasets with nonnegative elements has been reported. Our used data sets are:

- The CBCL database [19]: The size of this data set is 361×2429 which contains 2429 face images with 19×19 pixels per image.
- The ORL database [20]: The data set is a 10304×400 matrix, which contains 400 face images of size 92×112. To reduce the computational com plexity, all the ORL images were manually aligned and cropped to size 32×32.

All experiments have been done in Matlab and computer with Intel(R) 2.10 GHz and 2GB memory.

First our proposed ALS method applied to find NMF of ORL and CBCL data sets. Here the left singular vectors of data sets used as the projection operator.

There are several methods to solve the unconstrained sub-problems in ALS method. In our experiments we used QR factorization [18] .

Figures 2 and 3 show the relative error

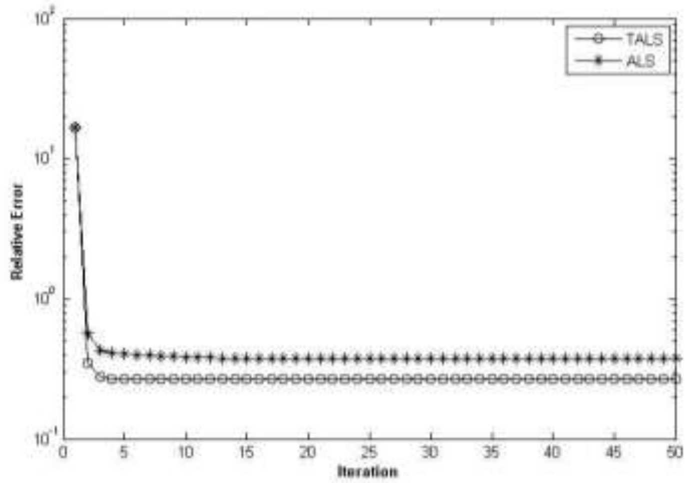


Figure 2. Relative errors of ALS and TALS algorithms for CBCL data set with $k=15$, $l_1=100$ and $l_2=100$.

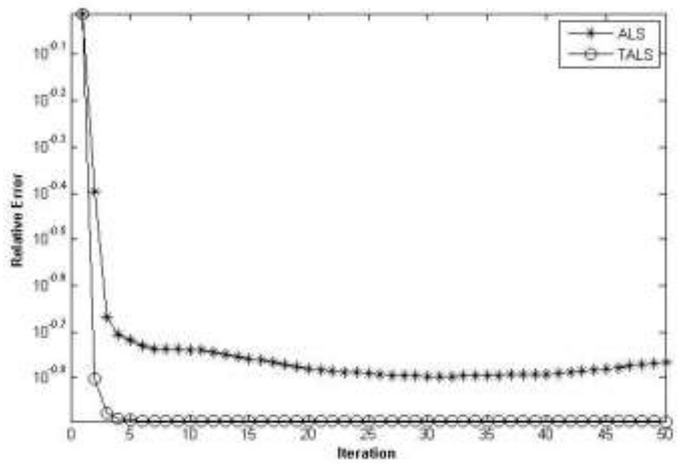


Figure 3. Relative errors of ALS and TALS algorithms for ORL data set with $k=15$, $l_1=100$ and $l_2=100$.

$\|A-WH\|/\|A\|$ for CBCL and ORL data sets when $k=15$ and dimensions of projections for the first and second subproblems are $l_1=100$ and $l_2=100$, respectively**.** Here proposed TALS method have better results in comparision with ALS.

This is occured based on denoising prpoperty of the first singular vectors. More detailes also can be seen In the Table 1. These results demonstrate that by TALS method we obtaine better results in less time in comparison with ALS method.

Also we used TALS and ALS methods as initialization methods in Algorithm 3 for ANLS algorithm and compare their quality in producing good initial points. As we mentioned before, ANLS is an approach and different algorithms in solving its sub-problems lead to different NMF solvers. So our methods can be adjusted with different algorithms.

In this paper we used active set algorithm based method for solving nonnegative sub-problems of original and projected models [10]. Also the following criteria $|f_{k+1}-f_k|/|f_k|\le\Delta$, (4.1)

where $f_k=\|A-W_kH_k\|$ and $\Delta=1e-4$, has been used as the stopping condition in the line 4 of the Algorithm 1. The inner iterations c and d in Algorithm 3 are considered as 10.

We apply ANLS algorithm based on ALS and TALS initialization method on CBCL and ORL databases which k is 49 and 25 for these data sets, respectively.

Table 2 shows the time and number of iteration of these two methods to achive stopping condition. Also the final relative error defined as $\|A-WH\|/\|A\|$,

Table 1. Comparision of ALS and TALS method for ORL and CBCL data sets.

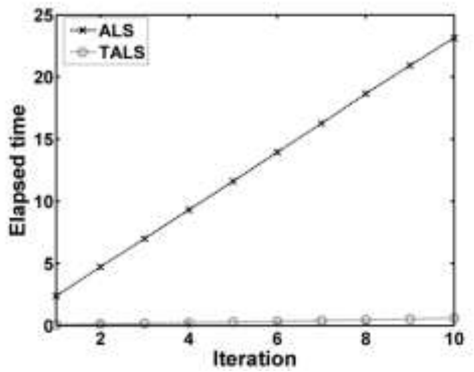
	Elapsed time		Ieration		Relative error	
Method	ALS	TALS	ALS	TALS	ALS	TALS
CBCL (r=15)	6.25	2.32	50	50	0.3795	0.2655
ORL (r=15)	3.6478	0.62	50	3	0.1578	0.1277

Table 2. Comparison of elapsed time, number of iterations and relative error between ALS-based and TALS-based initialization of ANLS via active set method ($\Delta = 10^{-4}$).

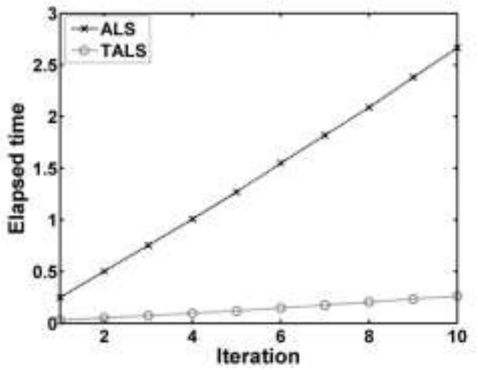
(c=10, d=10)	Elapsed time		No. iteration		Relative eror	
Initialization method	ALS	TALS	ALS	TALS	ALS	TALS
CBCL (l_1, l_2 =60)	2434.72	1811.87	175	141	0.0813	0.0811
ORL (l_1, l_2 =30)	96.23	71.23	89	68	0.1125	0.1124

(4.2) has been reported in this Table. From this table its clear that ANLS method with TALS initialization achives the same relative error as ANLS with ALS initialization. But its consuming time is very smaller than ALS based initialization.

Table 2 shows the time and number of iteration of these two methods to achieve stopping condition. Also the final relative error defined as $\|A - WH\| / \|A\|$, (4.2) has been reported in this Table. From this table it is clear that the ANLS method with TALS initialization archives the same relative error as the ANLS with ALS initialization. But its consuming time is very smaller than ALS based initialization.



(a)



(b)

Figure 4. Comparison of the initialization time between ALS and TALS on the CBCL database (a) and ORL database (b) with different iteration d=c=1,..,10.

Also Figure 4 shows the consuming time by ALS and TALS method in initialization step of ANLS method for different outer iterations $c = d = 1,...,10$. From this figure, it is clear that the time by ALS algorithm is less than time by ALS algorithm and their difference by increasing of outer iterations becomes larger.

In the end we compare our initialization TALS based method with the well-known SVD based method [21]. The numerical results based on SVD for both of databases CBCL and ORL are reported in Table 3. For CBCL our proposed method is faster than SVD based method and for ORL data set they have almost the same time.

Table 3. Elapsed time, number of iterations and relative error for ANLS via active set method based on SVD initialization ($\Delta = 10^{-4}$).

Data set	Time	Iteration	Value
CBCL	2340.11	179	0.0809
ORL	67.53	60	0.1128

Discussion

In this paper we proposed a fast ALS type method based on projection of data to an appoperiate subspsces. It has been shown that this method can improve both approximation and time of convergence in ALS based methods. This method does not depend on subproblems solving method and so it can be adjust with every method based on alternating approach.

References

1. Turk M. and Pentland A. Eigenfaces for recognition.

Journal of cognitive neuro science. **3**: 71-86 (1991).

2. Kuang D., Yun S. and Park H. SymNMF: nonnegative low-rank approximation of a similarity matrix for graph clustering. *Journal of Global Optimizatio.* **62** : 545-574 (2015).
3. Kim H. and Park His. Probabilistic non-negative matrix factorization: theory and application to microarray data analysis. *Journal of Bioinformatics and Computational Biology.* 1**2**: 1-25 (2014).
4. Li S.Z., Hou X., Zhang H. and Cheng Q. Learning spatially localized, parts-based representation. *Proc. IEEE computer society conference on computer vision and pattern recognition.* (2001).
5. Paatero P. and Tapper U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *J. Environ- metrics*, **5**: 111-126 (1994).
6. Lee D.D. and Seung H.S. Learning the parts of objects by non-negative matrix factorization. *Nature*. **401**:788-791(1999).
7. Lee D.D. and Seung H.S. Algorithms for non-negative matrix factorization. *Advances in neural information processing*. **13**: 556-562 (2001).
8. Zdunek R. and Cichocki A. Non-negative matrix factorization with Quasi-Newton optimization. *Proc. The eighth international conference on artificial intelligence and soft computing (ICAISC)*. pages: 870-879 (2006).
9. Lin C.-J. Projected gradient methods for non-negative matrix factorization. *Neural computation*. **19**: 2756-2779 (2007).
10. Kim H. and Park H. Non-negative matrix factorization based on alternating non-negativity constrained least squares and the active set method. *SIAM journal on matrix analysis and applications*, **30**: 713-730 (2008).
11. Lawson C.L. and Hanson R.J. Solving Least Squares Problems. *Prentice Hall, Englewood Cliffs, NJ*. (1974).
12. Kim J. and Park H. Toward faster non-negative matrix factorization: A new algorithm and comparisons. *Proc. IEEE international conference on data mining (ICDM)*. pages: 353-362(2008).
13. Berry M.W., Browne M., Langville A.N., Pauca V.P. and Plemmons R.J. Algorithms and applications for approximate non-negative matrix factorization. *Computational statistics and data analysis*. 52:155-173 (2007).
14. Cichocki A., Zdunek R., Phan A.H. and Amari S.-I. Non-negative Matrix and Tensor Factorizations. Applications To Exploratory Multi-Way Data Analysis and Blind Source Separation, (2009).
15. Bertsekas D.P. Nonlinear Programming , 2nd ed, *Athena Scientific, Belmont. MA*, (1999).
16. Grippo L. and Sciandrone M. On the convergence of the block nonlinear Gauss-Seidel method under convex constraints. *Operations research letters*. **26**: 127-136 (2000).
17. Van Benthem M. and Keenan M. Fast algorithm for the solution of large-scale non-negativity constrained least squares problems. *J. Chemometrics*. **18**: 441-450 (2004).
18. Golub G.H. and Van Loan C.F. Matrix Computations, 3rd end, *Johns Hopkins University Press*. (1996).
19. MIT Center for Biological and Computation Learning CBCL Face Database. Available at: http://cbcl.mit.edu/softwaredatasets/FaceData2.html.
20. AT&T Laboratories Cambridge ORL Database of Faces. Available at: http:// www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.
21. Boutsidis C. and Gallopoulos E. SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition*.**41**: 1350-1362 (2008).