Nonnegative Compression for Semi-Nonnegative Independent Component Analysis
Lu Wang, Amar Kachenoura, Laurent Albera, Ahmad Karfoul, Hua Zhong Shu, Lotfi Senhadji

To cite this version:

HAL Id: hal-01012136
https://hal.science/hal-01012136v2
Submitted on 29 Oct 2014
Nonnegative Compression for Semi-Nonnegative Independent Component Analysis

Lu Wang*,†§, Amar Kachenoura*†§, Laurent Albera*†§¶, Ahmad Karfoul†, Hua Zhong Shu†§ and Lotfi Senhadji*†§

*INSERM, UMR 1099, Rennes, F-35000, France
†Université de Rennes 1, LTSI, Rennes, F-35000, France
‡LIST, Southeast University, 2 Sipailou, 210096, Nanjing, China
§Centre de Recherche en Information Biomédicale sino-français (CRIBs), Rennes, France
¶INRIA, Centre Inria Rennes - Bretagne Atlantique, 35042 Rennes, France.
∥AL-Baath University, Mechanical and Electrical Engineering, PB. 2244, Homs, Syria.
Email: wanglyu1986@hotmail.com

Abstract—In many Independent Component Analysis (ICA) problems the mixing matrix is nonnegative while the sources are unconstrained, giving rise to what we call hereafter the Semi-Nonnegative ICA (SeNICA) problems. Exploiting the nonnegativity property can improve the ICA result. Besides, in some practical applications, the dimension of the observation space must be reduced. However, the classical dimension compression procedure, such as prewhitening, breaks the nonnegativity property of the compressed mixing matrix. Simulation results show its fast convergence property. An illustration of Blind nonnegativity of the compressed mixing matrix. Here we propose a nonnegative compression method that guarantees the nonnegativity of the compressed mixing matrix. In this paper, we introduce a new nonnegative compression method, which guarantees the nonnegativity of the compressed mixing matrix. Simulation results show its fast convergence property. An illustration of Blind nonnegativity of the compressed mixing matrix. Here we propose a nonnegative compression method that guarantees the nonnegativity of the compressed mixing matrix. This procedure is known as the prewhitening of \( x \) and \( W^T \) is called prewhitening matrix. However, such a compression method can not guarantee the nonnegativity of the compressed mixing matrix \( \bar{A} = W^T A \), because generally the unitary matrix \( W \) is not nonnegative. Now, in some practical situations, exploiting the nonnegativity property of the mixing matrix can improve the ICA result [1]–[3], [6]–[8]. Therefore, the objective of this paper is to propose a nonnegative compression method that guarantees the nonnegativity of the compressed mixing matrix \( \bar{A} \). To this end, note that span(\( W \)) = span(\( A \)), where the columns of \( A \) form a basis of nonnegative vectors of the signal subspace. Hence, there exists a nonsingular matrix \( B \in \mathbb{R}^{P \times P} \) such that the columns of matrix \( \bar{W} = WB \) form a basis of nonnegative vectors of the signal subspace span(\( A \)). It means that we can linearly transform the unitary matrix \( W \) into a nonnegative matrix \( \bar{W} \) multiplying it by \( B \) on its right. Now, since \( B \) is full rank, it admits the following QR matrix factorization:

\[
B = QRA
\]

where the columns of \( W \in \mathbb{R}^{N \times P} \) are the scaled eigenvectors corresponding to the \( P \) largest eigenvalues of the covariance matrix \( E[xx^T] - E[x]E[x^T] \). This procedure is known as the prewhitening of \( x \) and \( W^T \) is called prewhitening matrix. However, such a compression method can not guarantee the nonnegativity of the compressed mixing matrix \( \bar{A} = W^T A \), because generally the unitary matrix \( W \) is not nonnegative. Now, in some practical situations, exploiting the nonnegativity property of the mixing matrix can improve the ICA result [1]–[3], [6]–[8]. Therefore, the objective of this paper is to propose a nonnegative compression method that guarantees the nonnegativity of the compressed mixing matrix \( \bar{A} \). To this end, note that span(\( W \)) = span(\( A \)), where the columns of \( A \) form a basis of nonnegative vectors of the signal subspace. Hence, there exists a nonsingular matrix \( B \in \mathbb{R}^{P \times P} \) such that the columns of matrix \( \bar{W} = WB \) form a basis of nonnegative vectors of the signal subspace span(\( A \)). It means that we can linearly transform the unitary matrix \( W \) into a nonnegative matrix \( \bar{W} \) multiplying it by \( B \) on its right. Now, since \( B \) is full rank, it admits the following QR matrix factorization:

\[
B = QRA
\]

where \( Q \in \mathbb{R}^{P \times P} \) is a unitary matrix, where \( R \in \mathbb{R}^{P \times P} \) is a (unit) upper triangular matrix whose diagonal elements are equal to one and where \( \Lambda \in \mathbb{R}^{P \times P} \) is a nonsingular nonnegative diagonal matrix. Due to the fact that a multiplication of \( WQ \) by \( \Lambda \) does not change the subspace span(\( WQ \)) or the sign of the of the column vectors of span(\( WQR \)), looking for \( B \) of the form \( B = QR \) is sufficient. On the other hand, every unitary matrix can be factorized as a product of Givens matrices.

Definition 1. A Givens matrix \( Q^{(i,j)}(\theta_{i,j}) \), with \( i < j \), is equal to an identity matrix except the \( (i,i) \)-th, \( (j,j) \)-th and \( (i,j) \)-th entries, which are equal to \( \cos(\theta_{i,j}) \), \( \cos(\theta_{i,j}) \), \( -\sin(\theta_{i,j}) \) and \( \sin(\theta_{i,j}) \), respectively.

In addition, every unitary matrix can be factorized as a product of elementary upper triangular matrices.

\[
\bar{x} = W^T x = (W^T A)s = \bar{A}s
\]

where \( \bar{x} \) is the compressed mixing matrix.
A. Step 1: estimation of \( Q^{(i,j)}(\theta_{i,j}) \)

The Givens rotation matrix \( Q^{(i,j)}(\theta_{i,j}) \) transforms each pair of columns \((i, j)\) of \( W \) as follows:

\[
[v_i, v_j] = [w_i, w_j] \begin{bmatrix} \cos(\theta_{i,j}) & -\sin(\theta_{i,j}) \\ \sin(\theta_{i,j}) & \cos(\theta_{i,j}) \end{bmatrix}
\]

where \( v_i \) and \( w_i \) are the \( i \)-th columns of \( V \) and \( W \), respectively. Now, let us consider the following negativity measure defined in [9] and [10]:

\[
J(\theta_{i,j}) = \frac{1}{2} \sum_{n=1}^{N} (V_{n,i}^2 \mathbb{1}_{V_{n,i} < 0} + V_{n,j}^2 \mathbb{1}_{V_{n,j} < 0})
\]

where \( \mathbb{1}_{\alpha < 0} = \begin{cases} 1, \text{ if } \alpha < 0 \\ 0, \text{ otherwise} \end{cases} \), \( v_{n,i} \) and \( v_{n,j} \) are the \((n, i)\)-th and \((n, j)\)-th elements of \( V \), respectively. The purpose is to find an angle, \( \theta_{i,j} \), which minimizes the total sum of squares of negative elements of \( V \). The global optimum of \( J(\theta_{i,j}) \) is difficult to obtain analytically due to the existence of the Heaviside-step-like function \( \mathbb{1}_{\alpha < 0} \). We propose to compute \( \theta_{i,j} \) iteratively by a Newton’s method. For a given iteration \((it)\), let us consider the second order Taylor expansion, \( J_T(\theta_{i,j}) \), of \( J(\theta_{i,j}) \) around \( \theta_{i,j}^{(it)} \):

\[
J_T(\theta_{i,j}^{(it)} + \Delta \theta) = J(\theta_{i,j}^{(it)}) + \frac{dJ(\theta_{i,j}^{(it)})}{d\theta_{i,j}^{(it)}} \Delta \theta + \frac{d^2J(\theta_{i,j}^{(it)})}{2d\theta_{i,j}^{(it)}^2} (\Delta \theta)^2
\]

where \( \theta_{i,j}^{(it)} \) is the solution at the \( it \)-th iteration, \( \Delta \theta = \theta_{i,j} - \theta_{i,j}^{(it)} \), \( dJ(\theta_{i,j}^{(it)})/d\theta_{i,j}^{(it)} \) and \( d^2J(\theta_{i,j}^{(it)})/d\theta_{i,j}^{(it)}^2 \) are the first and second order derivatives of (6) with respect to \((w.r.t.) \theta_{i,j}^{(it)} \), respectively, which are given by [10]:

\[
dJ(\theta_{i,j}^{(it)})/d\theta_{i,j}^{(it)} = \sum_{n=1}^{N} V_{n,i}^2 \mathbb{1}_{V_{n,i} < 0} \mathbb{1}_{V_{n,j} > 0} - V_{n,i}^2 \mathbb{1}_{V_{n,i} > 0} \mathbb{1}_{V_{n,j} < 0} \quad \quad (8)
\]

and:

\[
d^2J(\theta_{i,j}^{(it)})/d\theta_{i,j}^{(it)}^2 = \sum_{n=1}^{N} \left( (v_{n,i}^{(it)})^2 - (v_{n,j}^{(it)})^2 \right) \mathbb{1}_{v_{n,i}^{(it)} < 0} \mathbb{1}_{v_{n,j}^{(it)} > 0} \mathbb{1}_{v_{n,i}^{(it)} > 0} \mathbb{1}_{v_{n,j}^{(it)} < 0} \quad \quad (9)
\]
which are given by:

\[
\begin{bmatrix}
\cos(\theta_{i,j}^{(it)}) & -\sin(\theta_{i,j}^{(it)}) \\
\sin(\theta_{i,j}^{(it)}) & \cos(\theta_{i,j}^{(it)})
\end{bmatrix}
\]

(10)

Then the minimum of (7) w.r.t. \( \Delta \theta \) can be reached at:

\[
\Delta \theta = -[d^2 J(\theta_{i,j}^{(it)}) / d(\theta_{i,j}^{(it)})^2]^{-1} [d J(\theta_{i,j}^{(it)}) / d \theta_{i,j}^{(it)}]
\]

(11)

\( d^2 J(\theta_{i,j}^{(it)}) / d(\theta_{i,j}^{(it)})^2 = 0 \) means that all the elements of \( \theta_{i,j}^{(it)} \) and \( \psi_{i,j}^{(it)} \) have the same sign. In this situation it is not necessary to perform the optimization since it will not decrease the criterion [10]. The process of all the rotation angles \( \theta_{i,j} \) is called a rotation sweep.

**B. Step 2: estimation of \( R_{i,j}^{(it)}(u_{i,j}) \)**

After the rotation step, the elementary upper triangular matrix \( R_{i,j}^{(it)}(u_{i,j}) \) transforms each pair of columns \( (i, j) \) of \( V \) as follows:

\[
[\vec{w}_i, \vec{w}_j] = [v_i, v_j] \begin{bmatrix} 1 & u_{i,j} \\ 0 & 1 \end{bmatrix}
\]

(12)

where \( \vec{w}_i \) is the \( i \)-th column of \( \vec{W} \), and \( u_{i,j} \) is the shear factor. From (12), \( \vec{w}_i \) remains unchanged, the negativity measure criterion is then defined as follows:

\[
J(u_{i,j}) = \frac{1}{2} \sum_{n=1}^{N} \sum_{j=1}^{2} \frac{\bar{w}_{n,j}^2}{1 - \bar{w}_{n,j}^2} < 0
\]

(13)

Similarly, the second order Taylor expansion \( J_T(u_{i,j}) \) of \( J(u_{i,j}) \) around \( u_{i,j}^{(it)} \) can be expressed as follows:

\[
J_T(u_{i,j}^{(it)} + \Delta u) = J(u_{i,j}^{(it)}) + \frac{d J(u_{i,j}^{(it)})}{d u_{i,j}} \Delta u + \frac{d^2 J(u_{i,j}^{(it)})}{2 d u_{i,j}^2} (\Delta u)^2
\]

(14)

where \( u_{i,j}^{(it)} \) is the solution at the \( i \)-th iteration, \( \Delta u = u_{i,j} - u_{i,j}^{(it)} \), \( d J(u_{i,j}^{(it)}) / d u_{i,j} \), and \( d^2 J(u_{i,j}^{(it)}) / d u_{i,j}^2 \) are the first and second order derivatives of (13) w.r.t. \( u_{i,j}^{(it)} \), respectively, which are given by:

\[
\frac{d J(u_{i,j}^{(it)})}{d u_{i,j}} = \frac{N}{n=1} \sum \bar{w}_{n,i} \bar{w}_{n,j}^2 \frac{1}{1 - \bar{w}_{n,j}^2} < 0
\]

and:

\[
\frac{d^2 J(u_{i,j}^{(it)})}{d u_{i,j}^2} = \frac{N}{n=1} \sum \bar{w}_{n,i} \bar{w}_{n,j}^2 \frac{1}{(1 - \bar{w}_{n,j}^2)^2} < 0
\]

with:

\[
[\bar{w}_{n,i}, \bar{w}_{n,j}] = [v_{n,i}, v_{n,j}] \begin{bmatrix} 1 & u_{i,j}^{(it)} \\ 0 & 1 \end{bmatrix}
\]

(17)

Then the minimum of (14) w.r.t. \( \Delta u \) can be reached at:

\[
\Delta u = -[d^2 J(u_{i,j}^{(it)} / d u_{i,j}^{(it)})^2]^{-1} [d J(u_{i,j}^{(it)}) / d u_{i,j}^{(it)}]
\]

(18)

When \( d^2 J(u_{i,j}^{(it)}) / d u_{i,j}^{(it)} = 0 \), it means that all the elements of \( \vec{w}_j^{(it)} \) have the same sign and (18) does not need to be computed. The process of all the shear factors \( u_{i,j} \) is called a shear sweep.

It is noteworthy that the second order derivatives (9) and (16) may not exist since the first order derivatives of the criteria (6) and (13) are continuous but non-differentiable at any point where \( w_{n,j} = 0 \) for some \( n,j \) [9]. In this case, we skip the estimation of the current parameter and continue to estimate the next one. There may exist a few very small negative values (less than \(-1e^{-20}\)) in the final result, which can be forced to zeros without affecting the effectiveness of the algorithm. In practice, the new NN-COMP algorithm needs several rotation sweeps followed by several shear sweeps, in order to ensure the convergence.

### III. Simulation Results

In this section, the performance of the proposed NN-COMP algorithm is evaluated following the SeNICA model. Firstly, we aim at studying the convergence property of NN-COMP. Secondly, we test the usefulness of NN-COMP as a preprocessing step of a recent published SeNICA method based on a Nonnegative Joint Diagonalization by Congruence (NJDC) method NLNUJ1D [7]. For all the following experiments, a noisy SeNICA model \( x = As + \nu \) is considered and the additive noise \( \nu \) is modeled as a zero-mean unit-variance Gaussian vector. In addition, the Signal to Noise Ratio (SNR) is defined by: \( SNR = 20 \log_{10}(\|As\|_F / \|
u\|_F) \). We repeat the experiments with 200 independent Monte Carlo (MC) trials.

#### A. Convergence test

In this experiment, let us generate: the \((N \times P)\) mixing matrix \( A \) with elements independently drawn from a uniform distribution between 0 and 1, and the source vector \( s \) with \( M = 2000 \) sample points, which are independent and uniformly distributed between \(-\sqrt{3}\) and \(\sqrt{3}\). More precisely, we consider two situations: i) in the first one the number of observations \( N \) equals 20 and the number of sources \( P \) is set to 3, and ii) in the second one \( N = 150 \) and \( P = 20 \). For both situations, we vary the SNR from 0 dB to 30 dB with a step of 5 dB. To evaluate the convergence performance of the NN-COMP, we define a negativity criterion as a ratio between the sum of the squared negative entries and the sum of all the squared components of \( \bar{W} \): \( \gamma(\bar{W}) = \|\bar{W}_{-}\|_F^2 / \|\bar{W}\|_F^2 \). Figure 2 depicts the mean values of the negativity criterion over the MC trials for different experiments. More precisely, the \( \gamma(\bar{W}) \) values calculated for each rotation sweep are plotted with blue solid line and those calculated for each shear sweep are plotted with red dash-dot line. It is interesting to show that the negativity ratio dramatically decreases thanks to the second step of NN-COMP: multiplying \( \bar{V} \) with the elementary upper triangular matrices. The proposed method seems to converge within a small number of sweeps, about 25 sweeps, whatever the number of the observations \( N \), the number of sources \( P \) and the SNR values.

#### B. BSS performance on MRS data

We have recently presented a new NJDC method in [7], called NNLUJ1D. The SeNICA method based on NNLUJ1D, namely NNLUJ1D-ICA, outperforms some classical ICA
methods on SeNICA problem. However, its numerical complexity, which is of order $O(N^4(P + M))$, significantly increases as the number of observation $N$ increases. Here, we propose to compare the behavior of the NNLUJ1D-ICA with and without using NN-COMP as a preprocessing step. For this purpose, an experiment is carried out on simulated MRS data, where the nonnegativity constraint on the mixing matrix $A$ is verified. Two realistic MRS metabolites, namely the Choline and Creatine, are generated by Lorentzian and Gaussian functions [11]. The mixing matrix $A$ is similarly generated as in the previous section. The performances of the two algorithms (NNLUJ1D-ICA and NN-COMP + NNLUJ1D-ICA) are studied as a function of the number of observations $N$, by varying $N$ from 2 to 30 with a step of 2. We test two SNR values: 5 dB and 15 dB. The performance criterion is defined as the error between the source $s$ and its estimate $\tilde{s}$, as follows [12]:

$$\alpha(s, \tilde{s}) = \frac{1}{p} \sum_{p=1}^{P} \min_{p', \tilde{s}_{p'}} \in I_p^2 d(s_p, \tilde{s}_{p'})$$

(19)

where $s_p$ and $\tilde{s}_{p'}$ are the $p$-th and $p'$-th components of $s$ and $\tilde{s}$, respectively. $I_p^2$ is defined recursively by: $I_0^2 = \{1, \cdots, P \} \times \{1, \cdots, P \}$, and $I_{p+1}^2 = I_p^2 - J_p^2$, where $J_p^2 = \arg\min_{(p, p') \in I_p^2} d(s_p, \tilde{s}_{p'})$. In addition, $d(s_p, \tilde{s}_{p'})$ is defined as the pseudo-distance between two random variables [13]:

$$d(s_p, \tilde{s}_{p'}) = 1 - (\|s_p \tilde{s}_{p'}\|^2)/(\|s_p\|^2 \|\tilde{s}_{p'}\|^2)$$

The estimation accuracy of the mixing matrix is not provided, since we just estimate the compressed mixing matrix.

Figure 3 displays an example of the separation results of the two methods (NNLUJ1D-ICA and NN-COMP + NNLUJ1D-ICA) with $N=20$ observations and a SNR of 5 dB. Both methods separate the sources quasi-perfectly. We can see that the NN-COMP algorithm does not deteriorate the separation quality ($\alpha(s, \tilde{s}) = 0.0543$ and 0.0544 for NNLUJ1D-ICA and NN-COMP + NNLUJ1D-ICA, respectively). The average curves of error $\alpha(s, \tilde{s})$, as well as that of the numerical complexities [14] in terms of floating point operations (flops) as a function of $N$ are shown in figure 4. It shows that for both methods, the increase of $N$ yields a better estimation of $s$ regardless of the tested values of SNR. The two methods perform quite similarly in terms of the source separation quality. For each $N$ value and SNR level, the errors $\alpha(s, \tilde{s})$ generated by the two methods are approximately the same. However, the numerical complexity of NNLUJ1D-ICA grows exponentially up to $10^{10}$ flops as a function of $N$. Fortunately, the burdensome numerical complexity can be reduced to less than $10^6$ flops, thanks to the NN-COMP method. It confirms the validity and the effectiveness of the proposed NN-COMP method as a preprocessing step of SeNICA algorithms.

IV. Conclusion

In this paper, we introduced a nonnegative compression method, called NN-COMP, in order to solve large scale SeNICA problems. It transforms the classical prewhitening matrix into a nonnegative matrix, which then preserves the
nonnegativity of the original mixing matrix. Simulation results show its fast convergence property. An illustration of a BSS application on MRS data confirms the validity and the efficiency of the proposed method.

REFERENCES


