

Non-negative Independent Component Analysis Algorithm Based on 2D Givens Rotations and a Newton Optimization

Wendyam S. B. Ouedraogo, Antoine Souloumiac, Christian Jutten

▶ **To cite this version:**

Wendyam S. B. Ouedraogo, Antoine Souloumiac, Christian Jutten. Non-negative Independent Component Analysis Algorithm Based on 2D Givens Rotations and a Newton Optimization. LVA/ICA 2010 - 9th International Conference on Latent Variable Analysis and Signal Separation, Sep 2010, Saint Malo, France. pp.522-529. hal-00526058

HAL Id: hal-00526058
https://hal.science/hal-00526058

Submitted on 13 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Non-negative Independent Component Analysis algorithm based on 2D Givens rotations and a Newton optimization

Wendyam S. B. OUEDRAOGO[1,2,3⋆], Antoine SOULOUMIAC[1], and Christian JUTTEN[3⋆⋆]

[1]CEA, LIST, Laboratoire d'Outils pour l'Analyse de Données, Gif-sur-Yvette, F-91191, France ;
[2]Unité Signaux et Systèmes, National School of Engineers of Tunis, BP 37, 1002 Tunis, Tunisia
[3]GIPSA-lab, UMR 5216 CNRS, University of Grenoble, 961 rue de la Houille Blanche BP 46 F-38402 Grenoble Cedex, France
{wendyam-serge-boris.ouedraogo}@cea.fr

Abstract. *In this paper, we consider the Independent Component Analysis problem when the hidden sources are non-negative (Non-negative ICA). This problem is formulated as a non-linear cost function optimization over the special orthogonal matrix group $SO(n)$. Using Givens rotations and Newton optimization, we developed an effective axis pair rotation method for Non-negative ICA. The performance of the proposed method is compared to those designed by Plumbley and simulations on synthetic data show the efficiency of the proposed algorithm.*

Key words: Non-negative ICA, Givens rotations, Newton optimization, Complexity calculation

1 Introduction

We consider the batch mode of ICA. Lets $S = \begin{bmatrix} s_1 & s_2 & \cdots & s_n \end{bmatrix}^T$ be the n hidden sources observed through a mixing matrix $A = [a_{ij}]$, $1 \leq i \leq m$ and $1 \leq j \leq n$. The noiseless model of ICA can be written:

$$X = AS \tag{1}$$

where $X = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix}^T$. We consider a square system where $m = n$.
The task of ICA is to find A and S given X.
In "classical" ICA the sources are required to be independent and non-Gaussian. Under these conditions, many algorithms based on maximization of the source

⋆ This study is conducted in the context of co-tutelle Ph-D program between CEA List, GIPSA Lab of University of Grenoble (France) and Unité Signaux et Systèmes of National School of Engineers of Tunis (Tunisia)
⋆⋆ Christian JUTTEN is also with Institut Universitaire de France

non-Gaussianity [5][4] or independence [3] have been developed for estimating the hidden sources up to the permutation and scaling indeterminations.
Subsequently, using a priori knowledge on the sources, some constraints such as sparsity have been incorporated in ICA to favour particular types of solutions [17].
In many real world applications such as biomedical imaging, music or spectrum analysis, the sources are known to be non-negative. This a priori must be taken into account when estimating the sources.
Several authors have proposed methods for solving the equation (1) under non-negativity constraint on S and/or A. The most used approach is Non-negative Matrix Factorization (NMF) [14][6][7][13] where the estimated sources and mixing matrix are all constrainted to be non-negative. However, the non-negativity alone is not sufficient to guarantee the uniqueness of the solution [8][9][10]. So depending on the application some constraints such as sparseness and/or smoothness also been incorporated in NMF to improve the parts based representation and reduce the range of admissible solutions [18].
For estimating the sources and/or mixing matrix under non-negativity constraint, another approach uses the a priori knowledges of the variables distribution to design a Bayesian method [19][20]. This approach needs however to make a "good" choice of A and S prior density and can be time-consuming.
Slightly relaxing the non-negativity constraint, Plumbley introduced Non-negative Independent Component Analysis [1][2] for solving (1) under non-negativity constraint on S, A being positive or of mixed sign. This approach requires the sources to be *non-negative* ($Pr(s_i < 0) = 0, \forall\ 1 \leq i \leq n$), *independent* ($Pr(\prod s_i) = \prod Pr(s_i)$), and *well grounded* ($\forall\ \delta > 0,\ Pr(s_i < \delta) > 0, \forall\ 1 \leq i \leq n$).
In this paper, we use Givens rotations and Newton optimization to develop an efficient axis pair rotation method for non-negative ICA.
The rest of the paper is organized as follows. Section 2 recapitulates Non-negative ICA problem and its formulation as a non-linear cost function optimization. In section 3, we describe the proposed axis pair rotation method. The computational complexity is evaluated in section 4 and we compare it with the geodesic search method designed by Plumbley. Section 5 discusses the simulation results and finally section 6 presents the conclusions.

2 Non-negative Independent Component Analysis

Under the independence and well grounded assumptions, the non-negative hidden sources S can be estimated by whitening the observations X and rotating the whitened data to fit them on the positive orthant.
In fact, let Z be the whitened observations, $Z = VX = VAS$ where V is a whitening matrix. Assuming that the sources are unit variance or transformed to be so, the covariance matrix of Z is given by $C_Z = I_n = (VA)(VA)^T$, then VA is an orthonormal matrix.
Let $Y = WZ$ and W be a rotation matrix ($W^TW = WW^T = I_n$ and $\det W = 1$).

$$Y = WVAS = US \text{ where } U = WVA$$

Plumbley showed [1] that U is a permutation matrix if and only if Y is positive (i.e. each element of Y is positive) with probability 1.
It is then sufficient to find a rotation matrix W so that the components of $Y = WZ$ are positive.
We consider the following negativeness measure criterion defined in [2] :

$$J(W) = \frac{1}{2}\left\|Z - W^T Y^+\right\|_F^2$$

where $Y^+ = \left[y_1^+ y_2^+ \cdots y_n^+\right]^T$, $y_i^+ = \max(0, y_i)$ and $\|\|_F$ is the Frobenius norm.

$$J(W) = \frac{1}{2}\left\|Z - W^T Y^+\right\|_F^2 = \frac{1}{2}\left\|W^T Y - W^T Y^+\right\|_F^2 = \frac{1}{2}\left\|Y - Y^+\right\|_F^2 = \frac{1}{2}\left\|Y^-\right\|_F^2$$

where $Y^- = \left[y_1^- y_2^- \cdots y_n^-\right]^T$ and $y_i^- = \min(0, y_i)$.
One can prove that $J(W) = 0 \Leftrightarrow Y^- = 0 \Leftrightarrow Y$ is positive with probability 1.
In pratical algorithm, the task of Non-negative ICA is to find a rotation matrix W that minimizes the criterion J. This is equivalent to solve the optimization problem (2) on the group of rotation matrices $SO(n)$:

$$W^* = \arg \min_{W \in SO(n)} J(W) \tag{2}$$

Several methods such as non-negative PCA [11][12], axis pair method [11] or geodesic search [15][16] have been proposed for solving (2). In the next section we propose an efficient axis pair rotation method for solving the optimization problem.

3 Givens parametization and Newton optimization for Non-negative ICA

When the sources are independent and well grounded, the task of solving the Non-negative Independent Component Analysis problem reduces to finding the rotation matrix W^* which minimizes the criterion J (resolving equation (2)).
Noting that any general n-dimensional rotation can be written as a product of *Givens rotations* $G(i_l, j_l, \theta_l)$ where

$$G(i_l, j_l, \theta_l) = \begin{pmatrix} 1 \ldots & 0 & \ldots & 0 & \ldots 0 \\ \vdots \ddots & \vdots & & \vdots & \vdots \\ 0 \ldots & \cos(\theta_l) & \cdots & \sin(\theta_l) & \ldots 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 \ldots & -\sin(\theta_l) & \cdots & \cos(\theta_l) & \ldots 0 \\ \vdots & \vdots & & \vdots & \ddots \vdots \\ 0 \ldots & 0 & \ldots & 0 & \ldots 1 \end{pmatrix} \begin{matrix} \\ \\ i_l \\ \\ j_l \\ \\ \end{matrix} \tag{3}$$

the task of computing the optimal rotation W^* is iteratively performed by several sweep of the $\frac{n(n-1)}{2}$ rotations, each rotation $G(i,j,\theta^k_{i,j})$ decreasing the criterion for the axis pair (i,j), $1 \leq i < j \leq n$ at sweep k. The whole rotation W^* is performed by multiplying the individual one.

$$W^* = \prod_k \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} G(i,j,\theta^k_{i,j})$$

Note that the rotation $G(i,j,\theta^k_{i,j})$ not commute and the product is write from rigth to left.

3.1 Computing the rotation $G(i,j,\theta^k_{i,j})$ at sweep k

For fixed (i,j), the optimal rotation $G(i,j,\theta^k_{i,j})$ is determined by the angle $\theta^k_{i,j}$. To simplify notation, we replace $\theta^k_{i,j}$ by θ so $G(i,j,\theta) = G(i,j,\theta^k_{i,j})$.
Noting that for updating Y, when multiplying by $G(i,j,\theta)$, the components of Y remain unchanged excepted for rows i and j, the computing of the optimal angle θ is done on the reduced 2D data given by (3) :

$$Y^k_{i,j} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos\theta) \end{pmatrix} \begin{pmatrix} Y(i,.) \\ Y(j,.) \end{pmatrix} \tag{4}$$

The criterion on reduced 2D data is given by :

$$J(\theta) = \frac{1}{2} \left\| (Y^k_{i,j})^- \right\|^2_F = \frac{1}{2} \sum_l \left(Y^2_{1l} \mathbb{1}_{Y_{1l}<0} + Y^2_{jl} \mathbb{1}_{Y_{jl}<0} \right) \tag{5}$$

where $\mathbb{1}_{Y_x<0} = \begin{cases} 1 \text{ if } Y_x < 0 \\ 0 \text{ otherwise} \end{cases}$.
Differentiating (4) with respect to θ and noticing that

$$Y_{1l} = Y(i,l)\cos(\theta) + Y(j,l)\sin(\theta) \Longrightarrow \frac{dY_{1l}}{d\theta} = -Y(i,l)\sin(\theta) + Y(j,l)\,cos(\theta) = Y_{2l}$$

$$Y_{2l} = -Y(i,l)\sin(\theta) + Y(j,l)\cos(\theta) \Longrightarrow \frac{dY_{2l}}{d\theta} = -Y(i,l)\cos(\theta) - Y(j,l)\sin(\theta) = -Y_{1l}$$

we get

$$\frac{dJ}{d\theta} = \sum_l Y_{1l} Y_{2l} \left[\mathbb{1}_{Y_{1l}<0} \mathbb{1}_{Y_{2l}>0} - \mathbb{1}_{Y_{1l}>0} \mathbb{1}_{Y_{2l}<0} \right] \tag{6}$$

and

$$\frac{d^2J}{d\theta^2} = \sum_l \left(Y^2_{2l} - Y^2_{1l} \right) \left[\mathbb{1}_{Y_{1l}<0} \mathbb{1}_{Y_{2l}>0} - \mathbb{1}_{Y_{1l}>0} \mathbb{1}_{Y_{2l}<0} \right] \tag{7}$$

A Newton method is used for optimizing θ leading to :

$$\theta = -\frac{dJ}{d\theta} \Big/ \frac{d^2J}{d\theta^2} \tag{8}$$

Note that $\frac{d^2J}{d\theta^2} = 0$ if all the samples are in the positive and/or the negative quadrant or all the samples are on the first and/or the second bisector. In this case it is not necessary to perform the rotation because it would not decrease the criterion.

3.2 Proposed algorithm

The Givens based parametization/Newton optimization method is described in the following algorithm :

```
Start whith whitening the data
   Z=VX
Initialization
   W = I_{n}, Y = Z, k = 1
Begin
   Repeat
      For i=1 to n-1
         For j=i+1 to n
            Compute dJ=dJ/dtheta as in (6)
            Compute d2J=d^{2}J/dtheta^{2} as in (7)
            If (d2J==0)
               continue
            Else
               theta_min=-dJ/d2J
               Form G(i,j,theta_min) as in (3)
               W = G(i,j,theta_min) W
               Y = G(i,j,theta_min) Y
            End
         End
      End
   k=k+1
   Until J(W) is less than a tolerance
End
```

4 Computational complexity

The proposed algorithm complexity is evaluated by counting the number of floating point operations (flops). One flop corresponds to one multiplication followed by one addition. We suppose that we have p samples ($p >> n$), the whole computational complexity is compute by adding individual complixity term and keeping only the term having p.

Proposed Axis pair rotation method
Note that for updating W and Y when rotating axis (i, j), it is only necessary to update the rows i and j.

1. $\frac{dJ}{d\theta} \longrightarrow O(4p)$
2. $\frac{d^2J}{d\theta^2} \longrightarrow O(2p)$
3. $\theta = \theta - \frac{dJ}{d\theta} / \frac{d^2J}{d\theta^2} \longrightarrow O(1)$
4. updating $W \longrightarrow O(4n)$
5. updating $Y \longrightarrow O(4p)$

For one sweep, the proposed algorithm has a complexity of $O(\frac{n(n-1)}{2}10p) = O(5n(n-1)p)$.

Geodesic search method

The computational complexity of the geodesic search method [15] designed by Plumbley is evaluated as follow :

1. $\theta = \frac{1}{2}\left\|Y^- Y^T - Y(Y^-)^T\right\|_F \longrightarrow O(n^2p)$
2. $H = \left[Y^- Y^T - Y(Y^-)^T\right] / \theta \longrightarrow O(n^2)$
3. $\frac{dJ}{dt} = -2\theta \longrightarrow O(1)$
4. $\frac{d^2J}{dt^2} = \|KoHY\|_F + < Y^-, H^2Y > \longrightarrow O(2n^2p)$
5. $t = -\arctan\left(\frac{dJ}{dt} / \frac{d^2J}{dt^2}\right) \longrightarrow O(1)$
6. $B = tH \longrightarrow O(n^2)$
7. $R = \exp(B) \longrightarrow O(n^3)$
8. $W = RW \longrightarrow O(n^3)$
9. $Y = RY \longrightarrow O(n^2p)$

The geodesic search method has a complexity of $O(4n^2p)$.

Comparing to the geodesic search approach, the axis pair rotation method is less complex especially when having small number of sources ($n \leq 5$) and the two algorithms have similar computational complexity otherwise.

References

1. Plumbley, M.: Conditions for Nonnegative Independent Component Analysis. IEEE Signal Processing Letters 9, 177–180 (2002)
2. Plumbley, M.: Algorithms for Nonnegative Independent Component Analysis. IEEE Transactions on Neural Networks 14, 534–543 (2003)
3. Comon, P.: Independent component analysis, A new concept? Signal Processing 36, 287–314 (1994)
4. Cardoso, J.F., Souloumiac, A.: Blind Beamforming for Non Gaussian Signal. IEE Proceedings-F 140, 362–370 (1993)
5. Hyvärinen, A.: Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. IEEE Transactions on Neural Networks 10, 626–634 (1999)
6. Lee, D.D., Seung, H.S.: Algorithms for Non-negative Matrix Factorization: 14th Annual Neural Information Processing Systems Conference. NIPS, Denver (2000)
7. Lin, C.J.: Projected Gradient Methods for Non-negative Matrix Factorization. Neural Computation 19, 2756–2779 (2007)
8. Laurberg, H.: Uniqueness of Non-negative Matrix Factorization. IEEE Statistical Signal Processing Workshop, 49-53 (2007)
9. Theis, F.J., Stadlthanner, K., Tanaka, T.: First Results on Uniqueness of Sparse Non-negative Matrix Factorization. EUSIPCO 2005

10. Donoho, D., Stodden, V.: When Does Non-Negative Matrix Factorization Give a Correct Decomposition into Parts? Advances in Neural Information Processing Systems 16, 1141-1148 (2003)
11. Plumbley, M., Oja, E.: A "Nonnegative PCA" Algorithm for Independent Component Analysis. IEEE Transactions on Neural Networks 15, 66–76 (2004)
12. Li, Y., Zheng, H.: Improvement for Nonnegative PCA Algorithm for Independent Component Analysis. Neural Networks and Brain, 2005. ICNN&B '05. International Conference on, 2000–2002(2005)
13. Cichocki, A., Zdunek, R., Phan, A.H., Amari, S.: Nonnegative Matrix and Tensor Factorizations Applications to Exploratory Multi-way Data Analysis and Blind Source Separation. John Wiley & Sons, Ltd (2009)
14. Paatero, P.: Least Squares Formulation of robust non-negative factor analysis. Chemometrics and Intelligent Laboratory Systems 37, 23–35 (1997)
15. Plumbley, M.: Optimization using Fourier Expansion over a Geodesic for Non-Negative ICA. Lecture notes in computer science 3195, 44–56 (2004)
16. Plumbley, M.: Geometrical methods for non-negative ICA: Manifolds, Lie groups and toral subalgebras. Neurocomputing 67, 161–197 (2005)
17. Bronstein, A.M., Bronstein, M.M., Zibulevsky, M., Zeevi, Y.Y.: Sparse ICA for Blind Separation of Transmitted and Reflected Images. Wiley Periodicals, 84–91 (2005)
18. Hoyer, P.O.: Non-negative Matrix Factorization with Sparseness Constraints. Journal of Machine Learning Research 5, 1457–1469 (2004)
19. Moussaoui, S., Brie, D., Mohammad-Djafari, A., Carteret, C.: Separation of Non-Negative Mixture of Non-Negative Sources Using a Bayesian Approach and MCMC Sampling. IEEE Transactions on Signal Processing 54, 4133-4145 (2006)
20. Duarte, L.T., Jutten, C., Moussaoui, S.: A Bayesian Nonlinear Source Separation Method for Smart Ion-Selective Electrode Arrays. IEEE Sensors Journal 9, 1763-1771 (2009)