ON THE CONVERGENCE OF ICA ALGORITHMS WITH SYMMETRIC ORTHOGONALIZATION

Alper T. Erdogan

Koc University, EE Department
Istanbul, Turkey

ABSTRACT

We study the convergence behavior of Independent Component Analysis (ICA) algorithms that are based on the contrast function maximization and that employ symmetric orthogonalization method to guarantee the orthogonality property of the search matrix. In particular, the characterization of the critical points of the corresponding optimization problem and the stationary points of the conventional gradient ascent and fixed point algorithms are obtained. As an interesting and a useful feature of the symmetrical orthogonalization method, we show that the use of symmetric orthogonalization enables the monotonic convergence for the fixed point ICA algorithms that are based on the convex contrast functions.

Index Terms— Independent Component Analysis, Blind Source Separation, Symmetric Orthogonalization, Fixed Point Algorithms, Convergence

1. INTRODUCTION

In the area of Independent Component Analysis (ICA) , and Blind Source Separation (BSS), the contrast function maximization based algorithms have attracted special attention. The fast fixed point (FastICA) algorithm introduced in [1] has played an important role in the growth of such interest.

The convergence behavior of these algorithms is of both theoretical and practical concern. Recently, Regalia & Kofidis [2] shown the monotonic convergence property of the fixed point ICA algorithms for extracting a single source from a mixture. In addition to the deflationary case, where sources (components) are sequentially obtained, the convergence behavior in the case of simultaneous extraction through a algorithm using symmetrical orthogonalization is of special interest. Oja [3] has shown that a particular set of matrices, where nonzero entries at each column has a constant magnitude, are the fixed points of the FastICA algorithm for the Kurtosis cost function, and among these fixed points, only the matrices corresponding to perfect separation condition are stable.

In this article, our goal is to generalize the characterization of fixed points in [3] for more general contrast functions and for both gradient ascent and fixed point algorithms using symmetric orthogonalization. Furthermore, as an important result, we provide the extension of the work in [2] and show that symmetric orthogonalization enables monotonic convergence to a fixed point in the simultaneous extraction of sources.

This work is supported in part by TUBITAK Career Award, Contract No:104E073.

2. ICA (BSS) SETUP

We consider the following ICA setup: the p independent components (sources) $s_1, \ldots, s_p$, which are assumed to have unity variance and zero mean, are mixed through a linear memoryless mapping

$$\underbrace{\begin{bmatrix} y_1(k) \\ \vdots \\ y_q(k) \end{bmatrix}}_{\mathbf{y}(k)} = \mathbf{H} \underbrace{\begin{bmatrix} s_1(k) \\ \vdots \\ s_p(k) \end{bmatrix}}_{\mathbf{s}(k)} \quad k = 1, \ldots, \Omega, \tag{1}$$

where

- $y_k, k = 1, \ldots, q$ are the mixtures and q is the number of mixtures (we assume overdetermined case, i.e., $q \geq p$),
- $\mathbf{H} \in \Re^{q \times p}$ is the full rank (tall or square) mixing matrix,
- Ω is the number of available samples.

We assume that the mixture sequence $\mathbf{y}(k)$ is prewhitened through the matrix $\mathbf{W}_{pre}$ such that the whitened observations

$$\mathbf{x}(k) \quad = \quad \mathbf{W}_{pre}\mathbf{y}(k) = \underbrace{\mathbf{W}_{pre}\mathbf{H}}_{\mathbf{C}}\mathbf{s} \tag{2}$$

has the covariance matrix $E(\mathbf{x}\mathbf{x}^T) = \mathbf{C}\mathbf{C}^T = \mathbf{I}$.

The goal is to obtain an orthogonal separator matrix $\mathbf{\Theta} \in \Re^{p \times p}$ such that the overall mapping from the sources to the separator outputs, which is given by

$$\mathbf{G} = \mathbf{\Theta}\mathbf{C} \tag{3}$$

is equal to the product of a permutation matrix and a diagonal matrix with unity magnitude entries, i.e.,

$$\mathbf{G} = \mathbf{P}\mathbf{\Lambda} \tag{4}$$

where the diagonal matrix $\mathbf{\Lambda}$ represents the sign ambiguity and the permutation matrix $\mathbf{P}$ represents the permutation ambiguity inherent in ICA problem.

The algorithms to achieve this goal can be derived from an optimization setting of the form

$$\begin{array}{ll} \text{optimize} & \mathcal{J}(\mathbf{\Theta}) \\ \text{subject to} & \mathbf{\Theta}\mathbf{\Theta}^T = \mathbf{I}, \end{array} \tag{5}$$

which is typically posed as a maximization problem.

In the deflation approach [4], each row is obtained separately where the above optimization problem is divided into a sequence of sub-optimization problems for each row:

$$\begin{array}{ll} \text{optimize} & \mathcal{J}_k(\mathbf{\Theta}_{k,:}) \\ \text{subject to} & \|\mathbf{\Theta}_{k,:}\|_2 = 1, \end{array} \tag{6}$$

During this process, once a row of $\mathbf{\Theta}$ is obtained, the corresponding source signal is subtracted from the mixture. The deflation approach has the advantage that the individual sub-optimization problems have a unity-2-norm constraint which is much easier to handle compared to the original orthogonal matrix constraint in (5). However, a clear disadvantage of the deflation process is the error accumulation problem, where estimation inaccuracies in the earlier optimization stages cause growing errors for the later stages.

Among the alternative approaches to solve the optimization problem in (5), a class of algorithms simultaneously train all the rows of $\mathbf{\Theta}$ using a two step procedure:

$$\begin{array}{rcll} \underline{\mathbf{\Theta}}^{(k+1)} & = & f(\mathbf{\Theta}^{(k)}) & (i) \\ \mathbf{\Theta}^{(k)} & = & \mathcal{M}_O(\underline{\mathbf{\Theta}}^{(k+1)}) & (ii) \end{array} \tag{7}$$

where

- the Step-(i) is the gradient update, where the two typical choices are
  - *The Conventional Gradient Ascent:*

$$f(\mathbf{\Theta}^{(k)}) = \mathbf{\Theta}^{(k)} + \mu^{(k)} \mathbf{\Delta}^{(k)}_{\mathbf{\Theta}^{(k)}} \tag{8}$$

    with $\mathbf{\Delta}^{(k)}_{\mathbf{\Theta}^{(k)}}$ is the gradient of the cost function with respect to $\mathbf{\Theta}^{(k)}$.
  - *Fixed Point Update:*

$$f(\mathbf{\Theta}^{(k)}) = \mathbf{\Delta}^{(k)}_{\mathbf{\Theta}^{(k)}} \tag{9}$$

- the Step-(ii) is the mapping to the set of orthogonal matrices, which is typically implemented using
  - Gram-Schmidt Orthogonalization
  - Projection to the closest orthogonal Matrix (according to Frobenius norm sense): In this case, if the singular value decomposition of $\underline{\mathbf{\Theta}}^{(k)}$ is given by,

$$\underline{\mathbf{\Theta}}^{(k)} = \underline{\mathbf{U}}^{(k)} \underline{\mathbf{\Sigma}}^{(k)} {\underline{\mathbf{V}}^{(k)}}^T \tag{10}$$

    then the minimum distance orthogonal mapping is given by

$$\mathcal{M}_O(\underline{\mathbf{\Theta}}^{(k)}) = \underline{\mathbf{U}}^{(k)} {\underline{\mathbf{V}}^{(k)}}^T. \tag{11}$$

    Note that, alternatively and equivalently, $\mathcal{M}$ can be written as

$$\mathcal{M}(\underline{\mathbf{\Theta}}^{(k)})_O = (\underline{\mathbf{\Theta}}^{(k)} {\underline{\mathbf{\Theta}}^{(k)}}^T)^{-1/2} \underline{\mathbf{\Theta}}^{(k)} \tag{12}$$

    where the matrix-square root above is the symmetric square root. This approach is also referred as symmetric orthogonalization.

In the next section, we investigate some convergence related properties of ICA (BSS) algorithms employing symmetric (minimum distance) orthogonalization.

3. CONVERGENCE PROPERTIES OF SYMMETRICAL ICA ALGORITHMS

In analyzing the convergence behavior of symmetrical ICA algorithms, we first look at the characterization of the potential convergence (or stationary) points. Following that, we'll investigate the monotonic convergence property of the fixed point ICA algorithm with symmetric orthogonalization. Throughout the article, we assume that the cost function $\mathcal{J}$ in (5) is differentiable with full-rank gradient over the set of orthogonal matrices.The case of rank deficient gradient can be similarly approached with a more careful treatment.

3.1. Critical Points of ICA Algorithms with Symmetric Orthogonalization

We start with statement of a fact related to the critical points of the optimization problem in (5):

Theorem 1. *Let $\mathbf{\Theta}_*$ be a local maximum of the problem (5). Then there exists an $\mathbf{S} \in \mathcal{C}^{p \times p}$ such that*

$$\mathbf{\Delta}_{\mathbf{\Theta}_*} = \mathbf{\Theta}_* \mathbf{S} \text{ and } \mathbf{S} = \mathbf{S}^T. \tag{13}$$

Proof: According to the Proposition 4.7.3 in [5], if $\mathbf{\Theta}_*$ is a local maxima of (5), then the gradient $\mathbf{\Delta}_{\mathbf{\Theta}_*}$ should be a member of the polar cone of the tangent cone of the constraint set at $\mathbf{\Theta}_*$, which is the vector space given by $\{\mathbf{\Theta}_* \mathbf{S} \mid \mathbf{S} = \mathbf{S}^T\}$. □

Using the linear relation in (3), we can rewrite the condition in (13), in terms of the overall mapping $\mathbf{G}$ as

$$\mathbf{\Delta}_{\mathbf{G}_*} = \mathbf{G}_* \mathbf{S}' \text{ and } \mathbf{S}' = {\mathbf{S}'}^T, \tag{14}$$

where $\mathbf{S}' = \mathbf{C}^T\mathbf{S}\mathbf{C}$.

A similar characterization can be provided for the stationary points of the conventional gradient ascent (i.e., (8)) based ICA algorithm with symmetric orthogonalization:

Theorem 2. *An orthogonal matrix $\mathbf{\Theta}_*$ is the stationary point of the ICA algorithm with update rule given in (8) using symmetric orthogonalization if and only if there exists an $\mathbf{S} \in \Re^{p\times p}$ for which the condition in (13) holds and $\lambda_{min}(\mathbf{S})\mu^{(k)} \geq -1\ \forall k$.*

Proof: $\mathbf{\Theta}_*$ is the stationary point of the algorithm if and only if the $\mathbf{\Theta}_*$ is mapped back to itself after the two step update in (7), which leads to

$$\begin{aligned} \mathbf{\Theta}_* &= \mathcal{M}_O(\mathbf{\Theta}_* + \mu^{(k)}\mathbf{\Delta}_{\mathbf{\Theta}_*}) && (15)\\ &= \mathcal{M}_O(\mathbf{\Theta}_*(\mathbf{I} + \mu^{(k)}{\mathbf{\Theta}_*}^T\mathbf{\Delta}_{\mathbf{\Theta}_*})) && (16)\\ &= \mathbf{\Theta}_*\mathcal{M}_O(\mathbf{I} + \mu^{(k)}{\mathbf{\Theta}_*}^T\mathbf{\Delta}_{\mathbf{\Theta}_*}) && (17)\end{aligned}$$

where the last equality follows from the definition of the mapping $\mathcal{M}_O$ defined by equations (10)-(11) and from the fact that $\mathbf{\Theta}_*$ is an orthogonal matrix.

As a result, from (17) we obtain,

$$\mathcal{M}_O(\mathbf{I} + \mu^{(k)}{\mathbf{\Theta}_*}^T\mathbf{\Delta}_{\mathbf{\Theta}_*}) = \mathbf{I}, \quad (18)$$

which implies that $\mathbf{I} + \mu^{(k)}{\mathbf{\Theta}_*}^T\mathbf{\Delta}_{\mathbf{\Theta}_*}$ has a singular value decomposition given by

$$\mathbf{I} + \mu^{(k)}{\mathbf{\Theta}_*}^T\mathbf{\Delta}_{\mathbf{\Theta}_*} = \mathbf{U}_*\mathbf{\Sigma}_*{\mathbf{U}_*}^T, \quad (19)$$

where $\mathbf{U}_*$ is a orthogonal matrix and $\mathbf{\Sigma}_*$ is a nonnegative diagonal matrix. The condition in (19) also implies that $\mathbf{I} + \mu^{(k)}{\mathbf{\Theta}_*}^T\mathbf{\Delta}_{\mathbf{\Theta}_*}$ is a Hermitian positive-definite matrix, which further implies that ${\mathbf{\Theta}_*}^T\mathbf{\Delta}_{\mathbf{\Theta}_*}$ is equal to a Hermitian matrix $\mathbf{S}$ for which $\lambda_{min}(\mathbf{S})\mu^{(k)} + 1 \geq 0$. □

As a result, the condition for the stationary point for the symmetric ICA algorithm using conventional gradient ascent with appropriate step size is equivalent to the necessary condition for the local maxima of the corresponding problem.

For the fixed point ICA algorithm employing symmetric orthogonalization, the stationary point condition is more restrictive, as stated by the following theorem:

Theorem 3. *An orthogonal matrix $\mathbf{\Theta}_*$ is a stationary point of the fixed point ICA algorithm with symmetric orthogonalization if and only if there exists a positive-definite $\mathbf{S} \in \Re^{p\times p}$ such that the condition in (13) holds.*

Proof: If $\mathbf{\Theta}_*$ is the stationary point then the following condition must hold:

$$\mathbf{\Theta}_* = \mathcal{M}_O(\mathbf{\Delta}_{\mathbf{\Theta}_*}). \quad (20)$$

Assuming that $\mathbf{\Delta}_{\mathbf{\Theta}_*}$ has a singular value decomposition

$$\mathbf{\Delta}_{\mathbf{\Theta}_*} = \mathbf{U}_{\mathbf{\Delta}}\mathbf{\Sigma}_{\mathbf{\Delta}}\mathbf{V}_{\mathbf{\Delta}}^T, \quad (21)$$

then

$$\mathbf{\Theta}_* = \mathbf{U}_{\mathbf{\Delta}}\mathbf{V}_{\mathbf{\Delta}}^T. \quad (22)$$

Therefore, we have

$$\begin{aligned} {\mathbf{\Theta}_*}^T\mathbf{\Delta}_{\mathbf{\Theta}_*} &= \mathbf{V}_{\mathbf{\Delta}}\mathbf{U}_{\mathbf{\Delta}}^T\mathbf{U}_{\mathbf{\Delta}}\mathbf{\Sigma}_{\mathbf{\Delta}}\mathbf{V}_{\mathbf{\Delta}}^T && (23)\\ &= \mathbf{V}_{\mathbf{\Delta}}\mathbf{\Sigma}_{\mathbf{\Delta}}\mathbf{V}_{\mathbf{\Delta}}^T = \mathbf{S} && (24)\end{aligned}$$

where $\mathbf{S}$ is a positive-definite matrix. □

3.2. Special Case of Kurtosis Maximization

In order to illustrate the results of the previous section, we concentrate on the popular Kurtosis maximization for which the cost function in (5) can be written as

$$\mathcal{J}(\mathbf{\Theta}) = \sum_{k=1}^{p} kurt(z_k) \quad (25)$$

where z_k's are the separator outputs and $kurt(.)$ is the kurtosis of its argument. We assume that all components (sources) have positive kurtosis values κ_k. The gradient of this cost function with respect to $\mathbf{G}$ is given by

$$\mathbf{\Delta}_{\mathbf{G}} = 4\mathbf{G}^{\odot 3}\mathbf{K} \quad (26)$$

where $\mathbf{K} = diag(\kappa_1, \kappa_2, \ldots, \kappa_p)$ and $\mathbf{G}^{\odot n}$ stands for n^{th} Hadamard power of $\mathbf{G}$.

We can now outline some specific examples of critical points based on (14) and (26):

Example 1: If $\mathbf{G}$ satisfies the perfect separation condition in (4), then

$$\begin{aligned} \mathbf{\Delta}_{\mathbf{G}} &= 4\mathbf{G}^{\odot 3}\mathbf{K} = 4\mathbf{P}\mathbf{\Lambda}^3\mathbf{K} = 4\mathbf{P}\mathbf{\Lambda}\mathbf{K}\mathbf{\Lambda}^2 && (27)\\ &= 4\mathbf{G}\mathbf{K}\mathbf{\Lambda}^2 && (28)\end{aligned}$$

Therefore, $\mathbf{G}^T\mathbf{\Delta}_{\mathbf{G}} = 4\mathbf{K}\mathbf{\Lambda}^2$ is a Hermitian matrix, which is also positive definite. Therefore not surprisingly, perfect separation matrices are the critical points of the optimization problem, the stationary points for the symmetrical gradient ascent algorithm for any positive $\mu^{(k)}$ and the stationary points for the symmetrical fixed point algorithm.

Example 2: If $\mathbf{G}$ has columns with nonzero entries having a constant magnitude β_j (the example provided in [3]), then

$$\mathbf{\Delta}_{\mathbf{G}} = 4\mathbf{G}^{\odot 3}\mathbf{K} = 4\mathbf{G}\mathbf{B}^2\mathbf{K} \quad (29)$$

where $\mathbf{B} = diag(\beta_1, \beta_2, \ldots, \beta_p)$. Therefore, $\mathbf{G}^T\mathbf{\Delta}_{\mathbf{G}} = 4\mathbf{B}^2\mathbf{K}$ is a Hermitian positive-definite matrix, therefore, we can make the same comments as in the previous case.

Example 3: As a more interesting example, consider the case where

$$\mathbf{G} = \frac{1}{7}\begin{bmatrix} -2 & 2 & -2 & 2 & 2 & -5 & -2 \\ -2 & 2 & -2 & 2 & 2 & 2 & 5 \\ 2 & -2 & 2 & 5 & -2 & -2 & 2 \\ 2 & 5 & 2 & -2 & -2 & -2 & 2 \\ -2 & 2 & -2 & 2 & -5 & 2 & -2 \\ -2 & 2 & 5 & 2 & 2 & 2 & -2 \\ 5 & 2 & -2 & 2 & 2 & 2 & -2 \end{bmatrix} \quad (30)$$

and $\mathbf{K} = \mathbf{I}$. If we look at the product $\mathbf{G}^T \boldsymbol{\Delta}_{\mathbf{G}}$, it is equal to

$$\frac{4}{7^3} \begin{bmatrix} 103 & 30 & -30 & 30 & 30 & 30 & -30 \\ 30 & 103 & 30 & -30 & -30 & -30 & 30 \\ -30 & 30 & 103 & 30 & 30 & 30 & -30 \\ 30 & -30 & 30 & 103 & -30 & -30 & 30 \\ 30 & -30 & 30 & -30 & 103 & -30 & 30 \\ 30 & -30 & 30 & -30 & -30 & 103 & 30 \\ -30 & 30 & -30 & 30 & 30 & 30 & 103 \end{bmatrix} \tag{31}$$

which is clearly a symmetric matrix, and therefore, $\mathbf{G}$ in (30) is a critical point. Furthermore, since the smallest eigenvalue of the matrix in (31) is equal to $\frac{-44}{49}$, it would be a stationary point for the conventional gradient ascent algorithm if $\mu^{(k)} < \frac{49}{44}$. Due to the fact that the symmetric matrix in (31) isn't positive definite, $\mathbf{G}$ in (30) is not a stationary point of the fixed point algorithm. In fact, if we apply one iteration of fixed point update (with symmetric orthogonalization) we obtain

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{32}$$

as new $\mathbf{G}$, which is a perfect separation point.

3.3. Monotonic Convergence of Fixed Point ICA Algorithms with Symmetric Orthogonalization

In order to show the monotonic convergence property of the Fixed Point ICA Algorithm employing symmetrical orthogonalization, we follow a treatment similar to the one in [2]. The basic result is summarized by the following theorem:

Theorem 4. *Fixed Point ICA algorithm with symmetrical orthogonalization corresponding to the optimization setting in (5) where $\mathcal{J}$ is a convex cost function (bounded on the set of orthogonal matrices), is monotonically convergent to one of the stationary points defined by Theorem 3.*

Proof: Given $\mathcal{J}$ is a differentiable convex function, for any $\mathbf{G}^{(k)}, \mathbf{G}^{(k+1)}$ pair, we have

$$\begin{aligned} \mathcal{J}(\mathbf{G}^{(k+1)}) &\geq \mathcal{J}(\mathbf{G}^{(k)}) \\ &\quad + \mathrm{Tr}(\boldsymbol{\Delta}_{\mathbf{G}^{(k)}}^T (\mathbf{G}^{(k+1)} - \mathbf{G}^{(k)})). \end{aligned} \tag{33}$$

If we now look at the optimization problem

$$\begin{array}{ll} \text{maximize} & \mathrm{Tr}(\boldsymbol{\Delta}_{\mathbf{G}^{(k)}}^T \mathbf{G}) \\ \text{subject to} & \mathbf{G}\mathbf{G}^T = \mathbf{I}, \end{array} \tag{34}$$

according to Proposition 4.7.3 in [5], a local optimum $\mathbf{G}_*$ of the problem in (34) should satisfy

$$\mathbf{G}_* \mathbf{S} = \boldsymbol{\Delta}_{\mathbf{G}^{(k)}} \qquad \text{for some } \mathbf{S} = \mathbf{S}^T. \tag{35}$$

From (35), we can write

$$\begin{aligned} \mathbf{S}^T \mathbf{S} &= \boldsymbol{\Delta}_{\mathbf{G}^{(k)}}^T \boldsymbol{\Delta}_{\mathbf{G}^{(k)}} && (36) \\ &= \mathbf{V}_{\boldsymbol{\Delta}} \boldsymbol{\Sigma}_{\boldsymbol{\Delta}} \mathbf{U}_{\boldsymbol{\Delta}}^T \mathbf{U}_{\boldsymbol{\Delta}} \boldsymbol{\Sigma}_{\boldsymbol{\Delta}} \mathbf{V}_{\boldsymbol{\Delta}}^T && (37) \\ &= \mathbf{V}_{\boldsymbol{\Delta}} \boldsymbol{\Sigma}_{\boldsymbol{\Delta}}^2 \mathbf{V}_{\boldsymbol{\Delta}}^T. && (38) \end{aligned}$$

From which we conclude that $\mathbf{S}$ is a Hermitian matrix with

$$\mathbf{S} = \mathbf{V}_{\boldsymbol{\Delta}} \boldsymbol{\Lambda} \mathbf{V}_{\boldsymbol{\Delta}}^T \tag{39}$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_{\boldsymbol{\Delta}} \mathbf{J}$ and $\mathbf{J}$ is a diagonal matrix with 1's,and/or -1s on the diagonal. Note that the cost function value at the critical point $\mathbf{G}_*$ is equal to

$$\begin{aligned} \mathrm{Tr}(\boldsymbol{\Delta}_{\mathbf{G}^{(k)}}^T \mathbf{G}_*) &= \mathrm{Tr}(\mathbf{S}) && (40) \\ &= \mathrm{Tr}(\mathbf{V}_{\boldsymbol{\Delta}} \Sigma_{\boldsymbol{\Delta}} \mathbf{J} \mathbf{V}_{\boldsymbol{\Delta}}^T) && (41) \\ &= \mathrm{Tr}(\Sigma_{\boldsymbol{\Delta}} \mathbf{J}). && (42) \end{aligned}$$

Among all critical points specified by (35), the global maximum value is achieved for $\mathbf{J} = \mathbf{I}$. Note that this case corresponds to

$$\begin{aligned} \mathbf{G}_* &= \mathbf{U}_{\boldsymbol{\Delta}} \mathbf{V}_{\boldsymbol{\Delta}}^T && (43) \\ &= \mathcal{M}_O(\boldsymbol{\Delta}_{\mathbf{G}^{(\mathbf{k})}}), && (44) \end{aligned}$$

i.e., the global maximum point for the problem in (34) is obtained by projecting the gradient $\boldsymbol{\Delta}_{\mathbf{G}^{(k)}}$ to the set of orthogonal matrices using mapping $\mathcal{M}_O$ (symmetric orthogonalization). Furthermore, the value of $\mathrm{Tr}(\boldsymbol{\Delta}_{\mathbf{G}^{(k)}}^T \mathbf{G}_*)$ at the global maximum point is strictly greater than its values at other critical points obtained for the choices of $\mathbf{J} \neq \mathbf{I}$. Therefore, given $\mathbf{G}^{(k)}$ is not a stationary point of the algorithm, i.e., $\mathbf{G}^{(k)} \neq \mathcal{M}_O(\boldsymbol{\Delta}_{\mathbf{G}^{(\mathbf{k})}})$, the choice

$$\mathbf{G}^{(k+1)} = \mathcal{M}_O(\boldsymbol{\Delta}_{\mathbf{G}^{(\mathbf{k})}}), \tag{45}$$

combined with (33), guarantees that $\mathcal{J}(\mathbf{G}^{(k+1)}) > \mathcal{J}(\mathbf{G}^{(k)})$. This fact, together with the boundedness of $\mathcal{J}$, implies the convergence. □

4. REFERENCES

[1] Aapo Hyvärinen and Erkki Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, October 1997.

[2] P. A. Regalia and E. Kofidis, "Monotonic convergence of fixed-point algorithms for ICA," *IEEE Transactions on Neural Networks*, vol. 14, no. 4, pp. 943–949, July 2003.

[3] Erikki Oja, "Convergence of the symmetrical FastICA algorithm," *Proc. of the 9th Int. Conf. on Neural Information Processing*, vol. 3, pp. 1368–1372, November 2002.

[4] P. Loubaton and Phillip Regalia, "Blind deconvolution of multivariate signals: a deflation approach," *IEEE International Conference on Communications*, vol. 2, pp. 1160–1164, May 1993.

[5] D. P. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, 2003.