

On the Convergence of ICA Algorithms With Symmetric Orthogonalization

Alper T. Erdogan

Abstract—Independent component analysis (ICA) problem is often posed as the maximization/minimization of an objective/cost function under a unitary constraint, which presumes the whitening of the observed mixtures. The parallel adaptive algorithms corresponding to this optimization setting, where all the separators are jointly trained, are typically implemented by a gradient-based update of the separation matrix followed by the so-called symmetrical orthogonalization procedure to impose the unitary constraint. This article addresses the convergence analysis of such algorithms, which has been considered as a difficult task due to the complication caused by the minimum-(Frobenius or induced 2-norm) distance mapping step. We first provide a general characterization of the stationary points corresponding to these algorithms. Furthermore, we show that fixed point algorithms employing symmetrical orthogonalization are monotonically convergent for convex objective functions. We later generalize this convergence result for nonconvex objective functions. At the last part of the article, we concentrate on the kurtosis objective function as a special case. We provide a new set of critical points based on Householder reflection and we also provide the analysis for the minima/maxima/saddle-point classification of these critical points.

Index Terms—Blind source separation, convergence, fixed point algorithms, independent component analysis (ICA), symmetric orthogonalization.

I. INTRODUCTION

IN the area of independent component analysis (ICA) and blind source separation, several approaches and connected algorithms have been proposed [1]. Among these, the contrast function optimization based approaches have attracted particular attention, especially after the introduction of the FastICA algorithm [2]. The algorithms in this field are derived from search methods aiming to maximize a given contrast function that reflects a measure of independence and/or non-Gaussianity. We can organize them under two major categories as follows.

- *Deflationary algorithms*, where each separator corresponding to an individual independent component (or source) is obtained separately. Therefore, the overall algorithm is a combination of consecutive subalgorithms, referred to as one-unit algorithms. Each subalgorithm is typically posed as the maximization of an objective function of the corresponding separator output under the unit

2-norm constraint on the separator vector. The algorithm for each subsection has usually two steps (see [1], [3]):

- a gradient based update on the separation vector, ignoring the constraint, followed by;
- the enforcement of the unit 2-norm constraint by the normalization of the resulting vector.

In the deflationary algorithm, once the execution of a subalgorithm is finished, the corresponding extracted source is removed from the mixture.

- *Parallel algorithms*, where all separators are simultaneously adapted. In this case the problem is usually formulated as the maximization of an objective function of the output vector, i.e., the collection of all outputs, under the unitary constraint on the separator matrix (see, for example, [1] and [4]). A typical parallel algorithm also has two steps:

- a gradient-based update of the separation matrix, ignoring the constraint, followed by;
- the enforcement of the unitariness constraint. This step is most usually implemented by either a Gram–Schmidt-based QR factorization or a polar-decomposition based minimum distance (in Frobenius norm or induced 2-norm sense) unitary mapping. The later is usually referred as *symmetric orthogonalization* as the orthogonalization procedure is invariant with respect to the specific ordering of the columns of the separation matrix.

The major convenience of the deflationary algorithm is that it leads to a more tractable analysis (see, for example, [5]). This stems from the fact that the unit 2-norm constraint of the one-unit approach is easier to tackle than the unitary constraint of the parallel approach. However, the error propagation and growth is a major concern in the deflationary algorithms as the estimation errors generated in the previous steps cause growing estimation errors in the following steps. This drawback of the deflationary approach would cause the favoring of parallel algorithms especially if their convergence analyses are properly executed. In fact, the goal of this article is to contribute to the convergence analysis of the parallel algorithms.

In the past, a major result about the convergence of the one-unit algorithms was provided in [2], where it is shown that the one-unit FastICA algorithm has cubic convergence for the special choice of Kurtosis objective function. In [3], Regalia and Kofidis have shown the monotonic convergence of the one-unit fixed point algorithms for a larger set of objective functions. In [6], Hao and Huper used a dynamical systems framework for the local convergence analysis of FastICA as an extension of [3].

For the convergence analysis of the symmetric ICA algorithms, Oja proposed a convergence analysis for the symmetric

Manuscript received June 05, 2008; accepted December 15, 2008. First published February 10, 2009; current version published May 15, 2009. The associate editor coordinating the review of this paper and approving it for publication was Dr. Deniz Erdogmus. This work is supported in part by TUBITAK Career Award, Contract 104E073 and Turkish Academy of Sciences GEBIP Program.

The author is with the Electrical Engineering Department, Koc University, Istanbul 34450, Turkey (e-mail: alper@stanfordalumni.org).

Digital Object Identifier 10.1109/TSP.2009.2015114

FastICA algorithm based on a Kurtosis objective function [7], which also appeared in [8]. Furthermore, in [9], both one-unit and symmetric ICA algorithms are shown to approach the Cramér–Rao lower bound for certain scenarios.

In this article, we extend the monotonic convergence results for the one unit algorithm provided in [3] to the parallel ICA algorithm using symmetric orthogonalization. Furthermore, we provide a convenient characterization of critical (stationary) points of both gradient ascent and fixed point parallel algorithms using symmetric orthogonalization. Based on this characterization, we extend the set of stationary points reported in [7], [8] for the Kurtosis objective function. The extension consists of the Householder transformation matrices with the hyperplane of reflection defined by the normal vectors whose nonzero entries have equivalent magnitude. We also provide the classification of these critical points as maximum, minimum or saddle points. Our initial results were reported in the conference article [10], where we provide some partial results for real valued sources.

The organization of the article is as follows. In Section II, the ICA setup for one-unit and symmetric ICA algorithms and the corresponding notation used in the article are introduced. In Section III, we provide the characterizations of the critical points of the symmetric ICA problem and the stationary points of the gradient ascent and fixed point parallel ICA algorithms based on symmetric orthogonalization. The convergence properties of these algorithms are analyzed in Section IV, where we first prove the monotonic convergence of the fixed point algorithms based on convex objective functions and then we provide the approach guaranteeing monotonic convergence for non-convex objective functions. In Section V, the Kurtosis objective function is analyzed as a special case, where we show that some Householder transformation based orthogonal matrices form critical points and we provide their classification. Finally, Section VI is the conclusion.

II. ICA SETUP

We consider the following setup for the ICA problem.

- There are p independent components (or sources): $s_1, \dots, s_p \in \mathcal{C}$ where we define

$$\mathbf{s} = [s_1 \ s_2 \ \dots \ s_p]^T.$$

Without loss of generality, we assume that the original sources have zero mean and unit variance, and therefore

$$\mathbf{R}_s = E(\mathbf{s}\mathbf{s}^H) = \mathbf{I}.$$

- These independent components are mixed through the memoryless channel $\mathbf{H} \in \mathcal{C}^{q \times p}$ to obtain q mixtures: y_1, y_2, \dots, y_q . Therefore, the linear mapping between the independent components and mixtures is given by

$$\mathbf{y} = \mathbf{H}\mathbf{s}$$

where

$$\mathbf{y} = [y_1 \ y_2 \ \dots \ y_q]^T.$$

Here, we assume that $q \geq p$, i.e., the number of mixtures is greater than or equal to the number of independent components. This scenario corresponds to the (overcomplete or square) instantaneous blind source separation problem.

- The goal is to adaptively obtain separator $\mathbf{W} \in \mathcal{C}^{q \times p}$ from the realizations of mixtures \mathbf{y} such that the separator outputs

$$\mathbf{z} = \mathbf{W}^T \mathbf{y}$$

are equal to or an approximation of the scaled and permuted version of the original sources.

- We assume that the observed mixtures are preprocessed by a whitening transformation such that the whitened observation vector given by

$$\mathbf{x} = \mathbf{W}_{\text{pre}}^T \mathbf{y}$$

where $\mathbf{W}_{\text{pre}} \in \mathcal{C}^{q \times p}$, has the covariance

$$\mathbf{R}_x = E(\mathbf{x}\mathbf{x}^H) = \mathbf{I}.$$

Therefore, the effective mapping between the sources and whitened observations, which is given by

$$\mathbf{C} = \mathbf{W}_{\text{pre}}^T \mathbf{H}$$

is a $p \times p$ unitary matrix, i.e.,

$$\mathbf{C}\mathbf{C}^H = \mathbf{C}^H\mathbf{C} = \mathbf{I}.$$

Therefore, the goal of two-stage ICA approaches that use prewhitening is to adaptively obtain a unitary matrix $\mathbf{\Theta} \in \mathcal{C}^{p \times p}$ such that the overall map between the sources and the separator outputs given by

$$\mathbf{\Psi} = \mathbf{\Theta}^T \mathbf{C} \tag{1}$$

satisfies

$$\mathbf{\Psi} = \mathbf{P}\mathbf{\Lambda} \tag{2}$$

where $\mathbf{P} \in \mathfrak{R}^{p \times p}$ is a permutation matrix (reflecting inherent permutation ambiguity) and $\mathbf{\Lambda} \in \mathcal{C}^{p \times p}$ is a diagonal matrix with complex entries located on the unit circle (reflecting inherent phase ambiguity).

- Based on the information above, the overall separator matrix is written as

$$\mathbf{W} = \mathbf{W}_{\text{pre}} \mathbf{\Theta}$$

i.e., the product of the whitening matrix and the unitary separator matrix, and therefore, the separator outputs can be written as

$$\begin{aligned} \mathbf{z} &= \mathbf{W}^T \mathbf{y} = \mathbf{\Theta}^T \mathbf{W}_{\text{pre}}^T \mathbf{y} \\ &= \mathbf{\Theta}^T \mathbf{x}. \end{aligned}$$

- In the contrast function maximization-based approach, the problem of obtaining $\mathbf{\Theta}$ is posed as the optimization problem given by

$$\begin{aligned} &\text{maximize} \quad \mathcal{J}(\mathbf{\Theta}) \\ &\text{subject to} \quad \mathbf{\Theta}\mathbf{\Theta}^H = \mathbf{I}. \end{aligned} \tag{3}$$

The objective function \mathcal{J} is a measure of mutual independence and/or non-Gaussianity of outputs. As an example of objective functions, we can give (multiuser) Kurtosis [11], [12]:

$$\mathcal{J}(\Theta) = \sum_{k=1}^p |\mathcal{K}(\Theta_{:,k}^T \mathbf{x})| \quad (4)$$

where $\mathcal{K}(\cdot)$ is the kurtosis of its argument which is defined as

$$\mathcal{K}(a) = E(|a|^4) - 2(E(|a|^2))^2 - |E(a^2)|^2$$

and $\Theta_{:,k}$ is the k th column of Θ . Note that as in the example given by (4), the objective function would be a purely real function of its complex argument Θ . Therefore, it would not satisfy the Cauchy–Riemann conditions which implies that it is not differentiable in the usual complex derivative sense. However, we can still define the complex gradient and Hessian based on “Wirtinger calculus” (see e.g., [13]–[16], [12], [17] for details), which will be assumed for the rest of the discussion. We assume that \mathcal{J} is a smooth function of

$$\begin{bmatrix} \Re\{\mathbf{t}\} \\ \Im\{\mathbf{t}\} \end{bmatrix}$$

where

$$\mathbf{t} = \text{vec}(\Theta) = [\Theta_{:,1}^T \quad \Theta_{:,2}^T \quad \dots \quad \Theta_{:,p}^T]^T. \quad (5)$$

Equivalently, \mathcal{J} is a smooth function (with respect to Wirtinger calculus) of the conjugate coordinates

$$\mathbf{t}_c = \begin{bmatrix} \mathbf{t} \\ \bar{\mathbf{t}} \end{bmatrix} \quad (6)$$

which implies that the partial derivative of \mathcal{J} with respect to \mathbf{t} exists assuming that conjugate coordinate $\bar{\mathbf{t}}$ is constant, and vice versa. We define the gradient operator as

$$\nabla_{\Theta} \mathcal{J}(\Theta) = \text{reshape} \left(\frac{\partial \mathcal{J}}{\partial \bar{\mathbf{t}}} \right)$$

where reshape is the reverse of the vec operator in (5), i.e.,

$$\text{reshape} \left(\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_{p^2} \end{bmatrix} \right) = \begin{bmatrix} t_1 & t_{p+1} & \dots & t_{p(p-1)+1} \\ t_2 & t_{p+2} & \dots & t_{p(p-1)+2} \\ \vdots & \vdots & \vdots & \vdots \\ t_p & t_{2p} & \dots & t_{p^2} \end{bmatrix}.$$

- In the deflationary approach, the problem in (3) is divided into a sequence of p subproblems corresponding to each column of Θ :

$$\begin{aligned} & \text{maximize} && \mathcal{J}_s(\Theta_{:,k}) \\ & \text{subject to} && \|\Theta_{:,k}\|_2 = 1. \end{aligned}$$

Each subproblem is solved through the one-unit algorithm. Right after the execution of the one-unit algorithm, the extracted source is removed from the mixture.

- In the parallel approach, all columns of Θ are trained simultaneously. For this purpose, the problem in (3) is directly solved through an algorithm which typically consists of 2-step iterations given by

$$\begin{aligned} \underline{\Theta}^{(i+1)} &= \Phi(\Theta^{(i)}, \nabla_{\Theta} \mathcal{J}(\Theta^{(i)})), & \text{(i)} \\ \Theta^{(i+1)} &= \mathcal{M}_U(\underline{\Theta}^{(i+1)}), & \text{(ii)} \end{aligned} \quad (7)$$

where we have the following.

- In (7.i), the mapping Φ represents the gradient based update. Among the different options, two common ones are

* *Gradient Ascent Update:*

$$\Phi(\Theta^{(i)}, \nabla_{\Theta} \mathcal{J}(\Theta^{(i)})) = \Theta^{(i)} + \mu^{(i)} \nabla_{\Theta} \mathcal{J}(\Theta^{(i)}) \quad (8)$$

where $\mu^{(i)}$ is the nonnegative step size.

* *Fixed Point Update:*

$$\Phi(\Theta^{(i)}, \nabla_{\Theta} \mathcal{J}(\Theta^{(i)})) = \nabla_{\Theta} \mathcal{J}(\Theta^{(i)}). \quad (9)$$

- In (7.ii) \mathcal{M}_U represents the mapping to the set of unitary matrices. There are two popular approaches for this mapping:

* *QR Factorization Based:* Gram–Schmidt procedure is applied to the updated matrix to convert its columns to orthonormal vectors.

* *Symmetric Orthogonalization (based on Minimum Distance Unitary Mapping (MDUM)):* In this case, the argument is mapped to the closest unitary matrix in both Frobenius norm and induced-2-norm sense: Given Υ is a nonsingular square matrix with a singular value decomposition (SVD)

$$\Upsilon = \mathbf{U}_{\Upsilon} \Sigma_{\Upsilon} \mathbf{V}_{\Upsilon}^H \quad (10)$$

then the closest unitary matrix (in Frobenius norm sense) to Υ is unique and is given by

$$\mathcal{M}_U^{\text{MDUM}}(\Upsilon) = \mathbf{U}_{\Upsilon} \mathbf{V}_{\Upsilon}^H. \quad (11)$$

Note that, if we write

$$\begin{aligned} \Upsilon &= \mathbf{U}_{\Upsilon} \Sigma_{\Upsilon} \mathbf{V}_{\Upsilon}^H = \underbrace{\mathbf{U}_{\Upsilon} \Sigma_{\Upsilon} \mathbf{U}_{\Upsilon}^H}_{\mathbf{\Pi}} \underbrace{\mathbf{U}_{\Upsilon} \mathbf{V}_{\Upsilon}^H}_{\mathbf{Q}} \\ &= \mathbf{\Pi} \mathbf{Q} \end{aligned} \quad (12)$$

where $\mathbf{\Pi}$ is positive definite and \mathbf{Q} is unitary, the expression in (12) is referred as the polar decomposition of Υ . Therefore, the MDUM maps a matrix to the unitary matrix obtained from its polar decomposition.

An alternative, and probably the most common representation for the MDUM symmetric orthogonalization, is given by

$$\mathcal{M}_U^{\text{MDUM}}(\Upsilon) = (\Upsilon \Upsilon^H)^{-1/2} \Upsilon \quad (13)$$

where the square root matrix in the above expression refers to the symmetrical square-root. Note that the mappings in (11) and (13) are identical and produce the same result.

¹ $\bar{\mathbf{x}}$ notation stands for the vector obtained by taking complex conjugate of elements of \mathbf{x} .

The following lemmas list some important properties of MDUM symmetrical orthogonalization that will be used in the following sections.

Lemma 1: For a full rank square matrix \mathbf{A} and unitary matrix Θ , $\mathcal{M}_U^{\text{MDUM}}(\Theta\mathbf{A}) = \Theta\mathcal{M}_U^{\text{MDUM}}(\mathbf{A})$.

Proof: Proof directly follows from the fact that if $\mathbf{U}_A\Sigma_A\mathbf{V}_A^H$ is a SVD of \mathbf{A} then $\Theta\mathbf{U}_A\Sigma_A\mathbf{V}_A^H$ is an SVD of $\Theta\mathbf{A}$. Therefore, $\mathcal{M}_U^{\text{MDUM}}(\Theta\mathbf{A}) = \Theta\mathbf{U}_A\mathbf{V}_A^H = \Theta\mathcal{M}_U^{\text{MDUM}}(\mathbf{A})$. ■

Lemma 2: For a full rank square matrix \mathbf{A} , $\mathcal{M}_U^{\text{MDUM}}(\mathbf{A}) = \mathbf{I}$ if and only if \mathbf{A} is a positive-definite matrix.

Proof: Let $\mathbf{U}_A\Sigma_A\mathbf{V}_A^H$ represent an SVD of \mathbf{A} : if direction: if \mathbf{A} is a positive definite matrix then singular values and singular vectors of \mathbf{A} coincide with the eigenvalues and eigenvectors of \mathbf{A} respectively (with proper magnitude ordering). Therefore, an SVD of \mathbf{A} can be written as $\mathbf{U}_A\Sigma_A\mathbf{U}_A^H$ where Σ_A contains the eigenvalues of \mathbf{A} on the diagonal which are also the singular values. As a consequence, $\mathcal{M}_U^{\text{MDUM}}(\mathbf{A}) = \mathbf{U}_A\mathbf{U}_A^H = \mathbf{I}$. only if direction: if $\mathcal{M}_U^{\text{MDUM}}(\mathbf{A}) = \mathbf{I}$ then equivalently $\mathbf{U}_A\mathbf{V}_A^H = \mathbf{I}$. Therefore, $\mathbf{U}_A = \mathbf{V}_A$. Which further implies, $\mathbf{A} = \mathbf{U}_A\Sigma_A\mathbf{U}_A^H$. Therefore, since Σ_A is a diagonal matrix with positive entries (for a full rank matrix), \mathbf{A} is a Hermitian positive-definite matrix. ■

- We should note that there are other alternative ICA approaches based on the optimization over general matrix manifolds. Readers are referred to [18] and references there in, for details on such algorithms.

III. CRITICAL POINTS OF SYMMETRICAL ICA ALGORITHMS

In the analysis of convergence behavior of algorithms, it is important to characterize “to which points” the algorithm converges as well as “how” it converges. Therefore, in this section we’ll obtain a general characterization of the critical points of the symmetric ICA algorithms based on symmetric (MDUM) orthogonalization. We’ll assume that the gradient of the objective function is of full rank over the set of unitary matrices.

The following theorem provides the description of the critical points of the optimization problem in (3) regardless of the chosen iterative algorithm.

Theorem 1: Let Θ_* be a local maximum of the problem (3). Then there exists an $\mathbf{S} \in \mathcal{C}^{p \times p}$ such that

$$\nabla_{\Theta}\mathcal{J}(\Theta_*) = \Theta_*\mathbf{S} \text{ and } \mathbf{S} = \mathbf{S}^H. \quad (14)$$

Proof: According to the [19, Prop. 4.7.3], if Θ_* is a local maxima of (3), then the gradient $\nabla_{\Theta}\mathcal{J}(\Theta_*)$ should be a member of the polar cone of the tangent cone of the constraint set at Θ_* , which is the vector space given by $\{\Theta_*\mathbf{S} \mid \mathbf{S} = \mathbf{S}^H\}$. ■

According to this theorem, if Θ_* is a critical point of the problem in (3) then the gradient of the objective function \mathcal{J} is in the normal space [4] of the constraint set.

Using the linear relation in (1), we can rewrite the condition in (14), in terms of the overall mapping Ψ as

$$\nabla_{\Psi}\mathcal{J}(\Psi_*) = \Psi_*\mathbf{S}' \text{ and } \mathbf{S}' = \mathbf{S}'^H \quad (15)$$

where $\Psi_* = \Theta_*^T\mathbf{C}$ and $\mathbf{S}' = \mathbf{C}^H\mathbf{S}\mathbf{C}$.

A separator matrix is referred as the fixed point or the stationary point of an algorithm if the algorithm iterations starting from this point remains at the same point. If the algorithm reaches to one of these stationary points it becomes “trapped” at that point. The following theorem provides the fixed point characterization of the symmetric ICA algorithms using gradient ascent update and the symmetric (MDUM) orthogonalization.

Theorem 2: For the ICA algorithm with update rule given in (8) using $\mathcal{M}_U^{\text{MDUM}}$ symmetric orthogonalization, where the step sizes in (8) are sufficiently small such that $\underline{\Theta}$ in (7) is a full rank matrix, a unitary matrix Θ_* is the stationary point if and only if there exists an $\mathbf{S} \in \mathcal{C}^{p \times p}$ for which the condition in (14) holds.

Proof: Θ_* is the stationary point of the algorithm if and only if the Θ_* is mapped back to itself after the two step update in (7), which can be mathematically stated as

$$\begin{aligned} \Theta_* &= \mathcal{M}_U^{\text{MDUM}}\left(\Theta_* + \mu^{(k)}\nabla_{\Theta}\mathcal{J}(\Theta_*)\right) \\ &= \mathcal{M}_U^{\text{MDUM}}\left(\Theta_*\left(\mathbf{I} + \mu^{(k)}\Theta_*^H\nabla_{\Theta}\mathcal{J}(\Theta_*)\right)\right) \\ &= \Theta_*\mathcal{M}_U^{\text{MDUM}}\left(\mathbf{I} + \mu^{(k)}\Theta_*^H\nabla_{\Theta}\mathcal{J}(\Theta_*)\right) \end{aligned} \quad (16)$$

where the last equality follows from Lemma 1. As a result, from (16) the condition for Θ_* to be stationary point can be rewritten as

$$\mathcal{M}_U^{\text{MDUM}}\left(\mathbf{I} + \mu^{(k)}\Theta_*^H\nabla_{\Theta}\mathcal{J}(\Theta_*)\right) = \mathbf{I}.$$

Due to the constraint on the step sizes, the argument of $\mathcal{M}_U^{\text{MDUM}}$ above is a full rank matrix. Therefore, according to Lemma 2, A unitary matrix Θ_* is a stationary point if and only if $\mathbf{I} + \mu^{(k)}\Theta_*^H\nabla_{\Theta}\mathcal{J}(\Theta_*)$ is a positive-definite matrix, which is further equivalent to the condition that $\Theta_*^H\nabla_{\Theta}\mathcal{J}(\Theta_*)$ is equal to a Hermitian matrix. ■

Note that the constraint on the step size in Theorem 2 implies, if $\Theta_*^H\nabla_{\Theta}\mathcal{J}(\Theta_*)$ is a positive-definite Hermitian matrix, then Θ_* is a stationary point for all nonnegative $\mu^{(k)}$ s. If $\Theta_*^H\nabla_{\Theta}\mathcal{J}(\Theta_*)$ is a Hermitian matrix with negative eigenvalues, then Θ_* is a stationary point for the non-negative step sizes that satisfy the condition²

$$\mu^{(k)} < \frac{1}{|\lambda_{\min}(\Theta_*^H\nabla_{\Theta}\mathcal{J}(\Theta_*))|} \forall k \geq 0.$$

Note that for the step size constraint defined above, the critical points of the optimization problem in (3) specified by Theorem 1 and the stationary points of the conventional gradient ascent, specified by Theorem 2, match.

For the symmetric algorithm based on fixed point update and symmetric (MDUM) orthogonalization, we can provide a similar characterization. The following theorem shows that the set of stationary points of the fixed point algorithm is a subset of the stationary points of the conventional gradient ascent algorithm.

Theorem 3: A unitary matrix Θ_* is a stationary point of the fixed point ICA algorithm with symmetric orthogonalization if and only if there exists a positive-definite $\mathbf{S} \in \mathcal{C}^{p \times p}$ such that the condition in (14) holds.

² $\lambda_{\min}(\mathbf{A})$ notation refers to the minimum eigenvalue of matrix \mathbf{A} .

Proof: Θ_* is a stationary point if and only if

$$\begin{aligned}\Theta_* &= \mathcal{M}_U^{\text{MDUM}}(\nabla_{\Theta} \mathcal{J}(\Theta_*)) \\ &\Leftrightarrow \Theta_*^H \mathcal{M}_U^{\text{MDUM}}(\nabla_{\Theta} \mathcal{J}(\Theta_*)) = \mathbf{I}.\end{aligned}$$

Using Lemma 1 an equivalent condition can be written as

$$\mathcal{M}_U^{\text{MDUM}}(\Theta_*^H \nabla_{\Theta} \mathcal{J}(\Theta_*)) = \mathbf{I}.$$

Due to Lemma 2 and due to the standing full rank assumption on the gradient matrix, this condition is equivalent to $\Theta_*^H \nabla_{\Theta} \mathcal{J}(\Theta_*)$ being a positive-definite matrix. ■

IV. CONVERGENCE OF SYMMETRIC ICA ALGORITHMS

A major concern about the symmetric ICA algorithms is about their convergence property. Especially in the case of symmetric (MDUM) orthogonalization-based parallel algorithms, the convergence analysis is obstructed by the complication caused by the orthogonalization step. However, in this section we'll show that we can extend the convergence results obtained for one-unit algorithms in [3] to parallel fixed-point ICA algorithms employing MDUM orthogonalization. Following a treatment similar to [3], we'll first concentrate on the convergence analysis of the convex objective functions. After that, we'll provide the modified fixed point algorithms and the corresponding convergence analysis for the nonconvex objective functions.

A. Convergence Analysis for Convex Objective Functions

When the objective function in (3) is convex, we can prove the monotonic convergence property of fixed point algorithms with MDUM orthogonalization. However, we first introduce the following lemma which would be useful in the proof of monotonic convergence:

Lemma 3: Given a full rank matrix $\mathbf{A} \in \mathcal{C}^{n \times n}$. The unique global maximum of the problem

$$\begin{aligned}\text{maximize} \quad & \Re\{\text{Tr}(\mathbf{A}^H \Theta)\} \\ \text{subject to} \quad & \Theta^H \Theta = \Theta \Theta^H = \mathbf{I}.\end{aligned}\quad (17)$$

is given by $\Theta = \mathcal{M}_U^{\text{MDUM}}(\mathbf{A})$.

Proof: Based on the Proposition 4.7.3 in [19], a local optimum Θ_* of the problem in (17) should satisfy

$$\Theta_* \mathbf{S} = \mathbf{A} \text{ for some } \mathbf{S} = \mathbf{S}^H. \quad (18)$$

Assuming \mathbf{A} has the singular value decomposition

$$\mathbf{A} = \mathbf{U}_A \Sigma_A \mathbf{V}_A^H$$

from (18), for the critical points of the problem in (17) we can write

$$\begin{aligned}\mathbf{S}^H \mathbf{S} &= \mathbf{A}^H \mathbf{A} = \mathbf{V}_A \Sigma_A \mathbf{U}_A^H \mathbf{U}_A \Sigma_A \mathbf{V}_A^H \\ &= \mathbf{V}_A \Sigma_A^2 \mathbf{V}_A^H.\end{aligned}$$

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

$$\mathbf{S} = \mathbf{V}_A \Lambda \mathbf{V}_A^H$$

where $\Lambda = \Sigma_A \mathbf{F}$ and \mathbf{F} is a diagonal matrix with 1's, and/or -1's on the diagonal. Note that the objective function value at the critical point Θ_* is equal to

$$\begin{aligned}\Re\{\text{Tr}(\mathbf{A}^H \Theta_*)\} &= \Re\{\text{Tr}(\mathbf{S})\} = \text{Tr}(\mathbf{V}_A \Sigma_A \mathbf{J} \mathbf{V}_A^H) \\ &= \text{Tr}(\Sigma_A \mathbf{F}).\end{aligned}$$

Among all critical points specified by (18), the global maximum value is achieved for $\mathbf{F} = \mathbf{I}$. Note that this case corresponds to

$$\Theta_* = \mathbf{U}_A \mathbf{V}_A^H = \mathcal{M}_U^{\text{MDUM}}(\mathbf{A}).$$

the global maximum point for the problem in (17) is obtained by projecting \mathbf{A} to the set of unitary matrices using mapping \mathcal{M}_U (MDUM). Furthermore, the value of $\Re\{\text{Tr}(\mathbf{A}^H \Theta_*)\}$ at the global maximum point is strictly greater than its values at other critical points obtained for the choices of $\mathbf{F} \neq \mathbf{I}$. ■

Based on the lemma above, we can state the following theorem about the monotonic convergence property of some fixed point algorithms:

Theorem 4: A Symmetric Fixed Point ICA algorithm with MDUM orthogonalization corresponding to the optimization setting in (3) where \mathcal{J} is a smooth convex objective function (over a convex set containing the set of unitary matrices), which is bounded on the set of unitary matrices, is monotonically convergent to one of the stationary points defined by Theorem 3.

Proof: Given \mathcal{J} is a smooth (with respect to conjugate coordinates described in Section II) convex function of Θ , for any $\Theta^{(k)}, \Theta^{(k+1)}$ pair in (7), with the update rule given in (9), we have (from the subgradient inequality in Appendix I)

$$\begin{aligned}\mathcal{J}(\Theta^{(k+1)}) &\geq \mathcal{J}(\Theta^{(k)}) \\ &+ 2\Re\left\{\text{Tr}\left(\nabla_{\Theta} \mathcal{J}(\Theta^{(k)})^H (\Theta^{(k+1)} - \Theta^{(k)})\right)\right\}.\end{aligned}\quad (19)$$

According to Lemma 3, given $\Theta^{(k)}$ is not a stationary point of the algorithm, i.e., $\Theta^{(k)} \neq \mathcal{M}_U^{\text{MDUM}}(\nabla_{\Theta} \mathcal{J}(\Theta^{(k)}))$ (due to Theorem 3), the choice

$$\Theta^{(k+1)} = \mathcal{M}_U^{\text{MDUM}}(\nabla_{\Theta} \mathcal{J}(\Theta^{(k)})),$$

guarantees that

$$\begin{aligned}\Re\left\{\text{Tr}\left(\nabla_{\Theta} \mathcal{J}(\Theta^{(k)})^H \Theta^{(k+1)}\right)\right\} \\ > \Re\left\{\text{Tr}\left(\nabla_{\Theta} \mathcal{J}(\Theta^{(k)})^H \Theta^{(k)}\right)\right\}\end{aligned}$$

Therefore, combined with (19), we obtain $\mathcal{J}(\Theta^{(k+1)}) > \mathcal{J}(\Theta^{(k)})$. This fact, together with the boundedness of \mathcal{J} , implies the convergence. ■

B. Extension for Nonconvex Objective Functions

The monotonic convergence property of the symmetric fixed point ICA algorithms, specified by Theorem 4, is valid for the convex objective functions. For a nonconvex objective function, we need to modify the fixed point update expression such that the resulting algorithm is guaranteed to be monotonically convergent. For this modification, we follow the similar steps used for the one-unit algorithm in [3].

We first start by defining the Hessian of $\mathcal{J}(\Theta)$ with respect to vectorized conjugate coordinates \mathbf{t}_c (defined by (5) and (6) as [13], [14])

$$\mathbf{H}^{\mathcal{J}}(\Theta) = \begin{bmatrix} \mathbf{H}_{\mathbf{t}\mathbf{t}}^{\mathcal{J}} & \mathbf{H}_{\mathbf{t}\bar{\mathbf{t}}}^{\mathcal{J}} \\ \mathbf{H}_{\bar{\mathbf{t}}\mathbf{t}}^{\mathcal{J}} & \mathbf{H}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}^{\mathcal{J}} \end{bmatrix} \quad (20)$$

where

$$\begin{aligned} \mathbf{H}_{\mathbf{t}\mathbf{t}}^{\mathcal{J}} &= \frac{\partial}{\partial \mathbf{t}} \left(\frac{\partial \mathcal{J}}{\partial \mathbf{t}} \right)^H, \mathbf{H}_{\mathbf{t}\bar{\mathbf{t}}}^{\mathcal{J}} = \frac{\partial}{\partial \bar{\mathbf{t}}} \left(\frac{\partial \mathcal{J}}{\partial \mathbf{t}} \right)^H \\ \mathbf{H}_{\bar{\mathbf{t}}\mathbf{t}}^{\mathcal{J}} &= \frac{\partial}{\partial \mathbf{t}} \left(\frac{\partial \mathcal{J}}{\partial \bar{\mathbf{t}}} \right)^H, \mathbf{H}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}^{\mathcal{J}} = \frac{\partial}{\partial \bar{\mathbf{t}}} \left(\frac{\partial \mathcal{J}}{\partial \bar{\mathbf{t}}} \right)^H. \end{aligned}$$

Based on this definition, \mathcal{J} would be convex on a convex domain containing the set of unitary matrices if and only if $\mathbf{H}^{\mathcal{J}}$ is positive semidefinite over its domain. In the case when $\mathbf{H}^{\mathcal{J}}$ is not positive-semidefinite, we can introduce an auxiliary objective function

$$\mathcal{P}(\Theta) = \mathcal{J}(\Theta) + \gamma \text{Tr}(\Theta \Theta^H - \mathbf{I})$$

over the domain $\mathcal{D} = \{\Theta \mid \Theta^H \Theta \preceq \mathbf{I}\}$.³ We note that $\mathcal{P}(\Theta) = \mathcal{J}(\Theta)$ over the constraint set, i.e., $\{\Theta \mid \Theta^H \Theta = \mathbf{I}\}$. Here, we choose

$$\gamma \geq - \inf_{\Theta \in \mathcal{D}} \lambda_{\min}(\mathbf{H}^{\mathcal{J}}(\Theta)) \quad (21)$$

where $\lambda_{\min}(\mathbf{H}^{\mathcal{J}}(\Theta))$ stands for the minimum eigenvalue of $\mathbf{H}^{\mathcal{J}}(\Theta)$ which has negative values over \mathcal{D} due to the fact that \mathcal{J} is not convex. The choice in (21) guarantees that the Hessian of \mathcal{P} , which is given by

$$\mathbf{H}^{\mathcal{P}}(\Theta) = \mathbf{H}^{\mathcal{J}}(\Theta) + \gamma \mathbf{I}$$

is positive semidefinite over \mathcal{D} . Therefore, the resulting \mathcal{P} would be a convex function over \mathcal{D} . As a result, the corresponding fixed point algorithm would be monotonically convergent due to Theorem 4. The fixed point algorithm corresponding to this function can be written as

$$\underline{\Theta}^{(k+1)} = \gamma \Theta^{(k)} + \nabla_{\Theta} \mathcal{J}(\Theta^{(k)}) \quad (22)$$

$$\Theta^{(k+1)} = \mathcal{M}_U^{\text{MDUM}}(\underline{\Theta}^{(k+1)}) \quad (23)$$

Based on the convexity of the defined auxiliary objective function \mathcal{P} , we can write

$$\begin{aligned} &\mathcal{P}(\Theta^{(k+1)}) - \mathcal{P}(\Theta^{(k)}) \\ &\geq 2\Re \left\{ \text{Tr} \left(\left(\gamma \Theta^{(k)} + \nabla_{\Theta} \mathcal{J}(\Theta^{(k)}) \right)^H \right. \right. \\ &\quad \left. \left. \times \left(\mathcal{M}_U \left(\gamma \Theta^{(k)} + \nabla_{\Theta} \mathcal{J}(\Theta^{(k)}) \right) - \Theta^{(k)} \right) \right) \right\} \quad (24) \end{aligned}$$

which yields a lower bound on the increments of the auxiliary objective function \mathcal{P} as well as the original objective function

³The generalized inequality $\preceq: \mathbf{A} \preceq \mathbf{B}$ means $\mathbf{B} - \mathbf{A}$ is a positive semidefinite matrix.

\mathcal{J} as they have the same value over the constraint set. The following theorem asserts that this lower bound is maximized for

$$\gamma = - \inf_{\Theta \in \mathcal{D}} \lambda_{\min}(\mathbf{H}^{\mathcal{J}}(\Theta)). \quad (25)$$

Theorem 5: The lower bound expression on the right side of (24) is a monotonic nonincreasing function of γ and therefore it is maximized for the minimum value of γ in the range specified by (21).

Proof: Let $\underline{\mathbf{U}}^{(k+1)} \underline{\Sigma}^{(k+1)} \underline{\mathbf{V}}^{(k+1)}$ denote an SVD of $\underline{\Theta}^{(k+1)}$ in (22), then the lower bound in (24) can be written as

$$\begin{aligned} \mathcal{L}(\gamma) &= 2\Re \left\{ \text{Tr} \left(\left(\gamma \Theta^{(k)} + \nabla_{\Theta} \mathcal{J}(\Theta^{(k)}) \right)^H \right. \right. \\ &\quad \left. \left. \times \left(\mathcal{M}_U^{\text{MDUM}} \left(\gamma \Theta^{(k)} \right. \right. \right. \right. \\ &\quad \left. \left. \left. + \nabla_{\Theta} \mathcal{J}(\Theta^{(k)}) \right) - \Theta^{(k)} \right) \right) \right\} \\ &= 2\Re \left\{ \text{Tr} \right. \\ &\quad \left. \times \left(\underline{\mathbf{U}}^{(k+1)} \underline{\Sigma}^{(k+1)} \underline{\mathbf{V}}^{(k+1)} \underline{\mathbf{U}}^{(k+1)} \underline{\mathbf{V}}^{(k+1)} \right) \right\} \\ &\quad - 2\Re \left\{ \text{Tr} \left(\gamma \Theta^{(k)H} \Theta^{(k)} \right) \right. \\ &\quad \left. + \text{Tr} \left(\nabla_{\Theta} \mathcal{J}(\Theta^{(k)})^H \Theta^{(k)} \right) \right\} \\ &= 2\text{Tr} \left(\underline{\Sigma}^{(k+1)} \right) - 2p\gamma \\ &\quad - 2\Re \left\{ \text{Tr} \left(\nabla_{\Theta} \mathcal{J}(\Theta^{(k)})^H \Theta^{(k)} \right) \right\}. \end{aligned}$$

Therefore, the derivative of this lower bound with respect to γ can be written as

$$\frac{d\mathcal{L}(\gamma)}{d\gamma} = 2 \frac{d\text{Tr}(\underline{\Sigma}^{(k+1)})}{d\gamma} - 2p.$$

In order to derive an expression for $(d\text{Tr}(\underline{\Sigma}^{(k+1)}))/(d\gamma)$, we first write down the singular value decomposition of $\underline{\Theta}^{(k+1)}$ in the form

$$\underline{\Theta}^{(k+1)} = \sum_{m=1}^{N_d^{(k+1)}} \underline{\sigma}_m^{(k+1)} \underline{\mathbf{U}}_m^{(k+1)} \underline{\mathbf{V}}_m^{(k+1)H}$$

where

- $N_d^{(k+1)}$ is the number of distinct singular values of $\underline{\Theta}^{(k+1)}$,
- $\underline{\sigma}_m^{(k+1)}$ is the m th distinct singular value of $\underline{\Theta}^{(k+1)}$ with multiplicity $r_m^{(k+1)}$ (Note that $\sum_{m=1}^{N_d^{(k+1)}} r_m^{(k+1)} = p$),
- $\underline{\mathbf{U}}_m^{(k+1)} \in \mathcal{C}^{p \times r_m^{(k+1)}}$ and $\underline{\mathbf{V}}_m^{(k+1)} \in \mathcal{C}^{p \times r_m^{(k+1)}}$ are semiorthogonal matrices containing left and right singular vectors of $\underline{\Theta}^{(k+1)}$ corresponding to $\underline{\sigma}_m^{(k+1)}$, respectively. Note that

$$\underline{\mathbf{U}}^{(k+1)} = \begin{bmatrix} \underline{\mathbf{U}}_1^{(k+1)} & \underline{\mathbf{U}}_2^{(k+1)} & \cdots & \underline{\mathbf{U}}_{N_d^{(k+1)}}^{(k+1)} \end{bmatrix} \quad (26)$$

and

$$\underline{\mathbf{V}}^{(k+1)} = \begin{bmatrix} \underline{\mathbf{V}}_1^{(k+1)} & \underline{\mathbf{V}}_2^{(k+1)} & \cdots & \underline{\mathbf{V}}_{N_d^{(k+1)}}^{(k+1)} \end{bmatrix}.$$

Based on [20, Property 4.1], we can write the equation shown at the bottom of the page where $\lambda_n(\cdot)$ is the n th eigenvalue of its argument. Since

$$\begin{aligned} & \left| \lambda_n \left(\mathbf{\Gamma}_m^{(k+1)} \right) \right| \\ & \leq \sup_{\mathbf{q} \in \mathcal{C}^{r_m^{(k+1)}}, \|\mathbf{q}\|_2=1} \left| \mathbf{q}^H \mathbf{\Gamma}_m^{(k+1)} \mathbf{q} \right| \\ & \leq \sup_{\substack{\mathbf{q} \in \mathcal{C}^{r_m^{(k+1)}} \\ \|\mathbf{q}\|_2=1}} \left| \left(\mathbf{U}_m^{(k+1)} \mathbf{q} \right)^H \mathbf{\Theta}^{(k)} \mathbf{V}_m^{(k+1)} \mathbf{q} \right| \\ & \quad + \left| \left(\mathbf{\Theta}^{(k)} \mathbf{V}_m^{(k+1)} \mathbf{q} \right)^H \mathbf{U}_m^{(k+1)} \mathbf{q} \right| \\ & \leq \sup_{\mathbf{q} \in \mathcal{C}^{r_m^{(k+1)}}, \|\mathbf{q}\|_2=1} 2 \left\| \mathbf{U}_m^{(k+1)} \mathbf{q} \right\|_2 \left\| \mathbf{\Theta}^{(k)} \mathbf{V}_m^{(k+1)} \mathbf{q} \right\|_2 \\ & = 2 \end{aligned}$$

for all $n = 1, \dots, r_m^{(k+1)}$, we can deduce that

$$\frac{d\text{Tr} \left(\underline{\Sigma}^{(k+1)} \right)}{d\gamma} \leq p \quad (27)$$

and therefore,

$$\frac{d\mathcal{L}(\gamma)}{d\gamma} \leq 0.$$

As a result the lower bound $\mathcal{L}(\gamma)$ is a monotonic nonincreasing function of γ , and therefore, it is maximized at the minimum value of its domain specified by (21). Note that the equality in (27) is achieved if and only if $\underline{\Theta}^{(k+1)} = \mathbf{\Theta}^{(k)} \mathbf{S}$ where \mathbf{S} is a positive-definite matrix, which is the condition for $\mathbf{\Theta}^{(k)}$ to be a stationary point of the iterations in (22)–(23) as shown in the proof of Theorem 2. ■

V. KURTOSIS OBJECTIVE FUNCTION AS A SPECIAL CASE

Kurtosis is one of the most popular objective functions in blind adaptive signal processing [21], [22], [11], [23]. In this section, we'll provide a converge analysis of the multiuser kurtosis function in (4) in the light of results provided in the previous sections.

We write the Kurtosis objective as a function of the overall mapping Ψ in (1), where it can be shown that (see, for example, [11])

$$\mathcal{J}_{\text{MUK}}(\Psi) = \sum_{m=0}^p |\kappa_m| \sum_{n=0}^p |\Psi_{n,m}|^4 \quad (28)$$

where κ_m is the kurtosis of the m th source, $\Psi_{n,m}$ is the (n, m) th element of Ψ , and MUK stands for Multi User Kurtosis. We assume that all of the sources are non-Gaussian. In [11], the convergence analysis of the BSS algorithm corresponding to this objective function using Gram–Schmidt orthogonalization was provided. In what follows, we will look at the convergence analysis for the algorithm based on MDUM orthogonalization.

The gradient of the objective function in (28) can be written as

$$\nabla_{\Psi} \mathcal{J}_{\text{MUK}}(\Psi) = (2\Psi^{\odot 2} \odot \bar{\Psi}) \mathbf{K} \quad (29)$$

where

$$\mathbf{K} = \text{diag}(|\kappa_1|, |\kappa_2|, \dots, |\kappa_p|)$$

\odot represents Hadamard, i.e., elementwise, power or product operation, and $\bar{\Psi}$ stands for matrix obtained by complex conjugating the elements of Ψ . It is shown in Appendix II that the gradient in (29) is always full rank when Ψ is a unitary matrix. Hessian (as defined in (20)) of the Kurtosis objective function is given by

$$\begin{aligned} & \mathcal{H}_{\Psi}^{\mathcal{J}_{\text{MUK}}}(\Psi) \\ & = \begin{bmatrix} 4\text{diag}(\text{vec}((\Psi \odot \bar{\Psi}) \mathbf{K})) & 2\text{diag}(\text{vec}((\Psi^{\odot 2}) \mathbf{K})) \\ 2\text{diag}(\text{vec}((\bar{\Psi}^{\odot 2}) \mathbf{K})) & 4\text{diag}(\text{vec}((\Psi \odot \bar{\Psi}) \mathbf{K})) \end{bmatrix}. \end{aligned}$$

Since $\mathcal{H}_{\Psi}^{\mathcal{J}}$ is a diagonally dominant Hermitian matrix, with non-negative diagonal entries, it is positive semidefinite. Therefore, the objective function in (28) is a convex function of its argument. We also note that

$$\inf_{\Psi \in \mathcal{D}} \lambda_{\min}(\mathcal{J}(\Psi)) = 0.$$

A. Critical Points

We can now outline some specific examples of critical points based on (15) and (29) as follows.

- If Ψ satisfies the perfect separation condition in (2), then

$$\begin{aligned} \nabla_{\Psi} \mathcal{J}(\Psi) & = 2\mathbf{P}\mathbf{\Lambda}^2 \bar{\mathbf{\Lambda}} \mathbf{K} = 2\mathbf{P}\mathbf{K}\mathbf{\Lambda}\mathbf{K}\bar{\mathbf{\Lambda}} \\ & = 2\Psi \mathbf{K}\mathbf{\Lambda}\bar{\mathbf{\Lambda}}. \end{aligned}$$

Therefore, $\Psi^H \nabla_{\Psi} \mathcal{J}(\Psi) = 2\mathbf{K}\mathbf{\Lambda}\bar{\mathbf{\Lambda}}$ is a Hermitian matrix, which is also positive definite. As a result, not surprisingly, the 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.

$$\frac{d\text{Tr} \left(\underline{\Sigma}^{(k+1)} \right)}{d\gamma} = \frac{1}{2} \sum_{m=1}^{N_d^{(k+1)}} \sum_{n=1}^{r_m^{(k+1)}} \lambda_n \left(\underbrace{\mathbf{U}_m^{(k+1)H} \mathbf{\Theta}^{(k)} \mathbf{V}_m^{(k+1)} + \mathbf{V}_m^{(k+1)H} \mathbf{\Theta}^{(k)} \mathbf{U}_m^{(k+1)}}_{\mathbf{\Gamma}_m^{(k+1)}} \right)$$

- If Ψ has columns with nonzero entries having a constant magnitude, i.e., all the nonzero elements in column- j have magnitude equal to β_j , (the example provided in [7]), then

$$\nabla_{\Psi} \mathcal{J}(\Psi) = 2\Psi \mathbf{B}^2 \mathbf{K}$$

where $\mathbf{B} = \text{diag}(\beta_1, \beta_2, \dots, \beta_p)$. Therefore, $\Psi^H \nabla_{\Psi} \mathcal{J}(\Psi) = 2\mathbf{B}^2 \mathbf{K}$ is a Hermitian positive-definite matrix, therefore, we can make the same comments as in the previous case.

- Let Ψ_* be a matrix satisfying the critical point condition in (15). If we define

$$\Phi = \Psi_* \Pi$$

where Π is a perfect separation matrix, then based on (29), we can write

$$\nabla_{\Psi} \mathcal{J}(\Phi) = \nabla_{\Psi} \mathcal{J}(\Psi_*) \Pi.$$

Therefore

$$\Phi^H \nabla_{\Psi} \mathcal{J}(\Phi) = \Pi^H \Psi_*^H \nabla_{\Psi} \mathcal{J}(\Psi_*) \Pi$$

would be a Hermitian matrix with the same inertia as $\Psi_*^H \nabla_{\Psi} \mathcal{J}(\Psi_*)$ and in consequence, Ψ_* and Φ would share the same critical/stationary point property. Similarly, defining $\Xi = \Pi \Psi_*$ would yield

$$\Xi^H \nabla_{\Psi} \mathcal{J}(\Xi) = \Psi_*^H \nabla_{\Psi} \mathcal{J}(\Psi_*),$$

and therefore, Ξ would also share the same critical/stationary point property as Ψ_* .

- As a more interesting example, consider the case where

$$\Psi = \frac{1}{6} \begin{bmatrix} 2 & -2 & 2 & 2 & 4 & 2 \\ 2 & 4 & 2 & 2 & -2 & 2 \\ 4 & 2 & -2 & -2 & 2 & -2 \\ -2 & 2 & -2 & -2 & 2 & 4 \\ -2 & 2 & 4 & -2 & 2 & -2 \\ -2 & 2 & -2 & 4 & 2 & -2 \end{bmatrix} \quad (30)$$

and $\mathbf{K} = \mathbf{I}$. If we look at the product $\Psi^H \nabla_{\Psi} \mathcal{J}(\Psi)$, it is equal to

$$\frac{2}{9} \begin{bmatrix} 7 & 2 & -2 & -2 & 2 & -2 \\ 2 & 7 & 2 & 2 & -2 & 2 \\ -2 & 2 & 7 & -2 & 2 & -2 \\ -2 & 2 & -2 & 7 & 2 & -2 \\ 2 & -2 & 2 & 2 & 7 & 2 \\ -2 & 2 & -2 & -2 & -2 & 7 \end{bmatrix} \quad (31)$$

which is clearly a symmetric matrix, and therefore, Ψ in (30) is a critical point. Furthermore, since the smallest eigenvalue of the matrix in (31) is equal to $-(2/3)$, it would be a stationary point for the conventional gradient ascent algorithm if $\mu^{(k)} < (3/2)$. Due to the fact that the symmetric matrix in (31) isn't positive definite, Ψ in (30) is not a stationary point of the fixed point algorithm. In fact, if

we apply one iteration of fixed point update (with MDUM symmetric orthogonalization) we obtain

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

as new Ψ , which is a perfect separation point.

- The last example can be obtained as a permuted version of more general set of critical points, for $\mathbf{K} = \kappa \mathbf{I}$ which is a typical scenario in separation of digital communications signals, given by

$$\Psi = \mathbf{I} - 2\mathbf{v}\mathbf{v}^H \quad (32)$$

where $\mathbf{v} \in \mathcal{C}^{p \times 1}$ has L nonzero entries with magnitude $(1)/(\sqrt{L})$. The matrices in this category correspond to Householder transformations where the axis of reflection has the normal vector whose nonzero elements have the same magnitude. It is interesting to note here that such vectors are the critical points for the one-unit Kurtosis maximization algorithm. For this case

$$\begin{aligned} \nabla_{\Psi} \mathcal{J}(\Psi) &= 2\kappa \Psi^{\odot 2} \odot \bar{\Psi} = 2\kappa \Psi \odot \Psi \odot \bar{\Psi} \\ &= 2\kappa \Psi \odot (\mathbf{I} - 2\mathbf{v}\mathbf{v}^H) \odot (\mathbf{I} - 2\bar{\mathbf{v}}\bar{\mathbf{v}}^T) \\ &= 2\kappa \Psi \odot (\mathbf{I} - 4\text{diag}(\mathbf{a}) + 4\mathbf{a}\mathbf{a}^T) \\ &= 2\kappa \mathbf{I} + 12 \left(\frac{2-L}{L} \right) \text{diag}(\mathbf{a}) - \frac{16}{L^2} \mathbf{v}\mathbf{v}^H \end{aligned}$$

where \mathbf{a} is the vector obtained by replacing nonzero entries of \mathbf{v} with $1/L$. As a result, the product

$$\begin{aligned} \Psi^H \nabla_{\Psi} \mathcal{J}(\Psi) &= \kappa (\mathbf{I} - 2\mathbf{v}\mathbf{v}^H) \\ &\times \left(2\mathbf{I} + 12 \left(\frac{2-L}{L} \right) \text{diag}(\mathbf{a}) - \frac{16}{L^2} \mathbf{v}\mathbf{v}^H \right) \\ &= \kappa \left(2\mathbf{I} + 12 \left(\frac{2-L}{L} \right) \text{diag}(\mathbf{a}) \right. \\ &\quad \left. - 2 \frac{2L^2 - 12L + 16}{L^2} \mathbf{v}\mathbf{v}^H \right) \\ &= \mathbf{S}_v \end{aligned}$$

Note that $\mathbf{S}_v = \mathbf{S}_v^H$, and therefore, Ψ in (32) is a critical point. The eigenvalues of \mathbf{S}_v are given by

$$\lambda_n = \begin{cases} 2\kappa \frac{-L^2 + 6L - 4}{L^2} & n = 1 \\ 2\kappa \frac{L^2 - 6L + 12}{L^2} & 2 \leq n \leq L \\ 2\kappa & L + 1 \leq n \leq p. \end{cases}$$

All eigenvalues, except λ_1 are always positive. λ_1 is positive for $0 \leq L \leq 5$ and it is negative for $L \geq 6$. Therefore, the critical point in (32) is a stationary point of the fixed point symmetrical ICA algorithm only when $0 \leq L \leq 5$.

The following theorem provides the minima-maxima-saddle point characterization of the critical points in (32).

Theorem 6: The critical points in (32), for $\mathbf{K} = \kappa \mathbf{I}$, are classified as follows.

- *Global Maxima:* $L = 1$ and $L = 2$ corresponds to Global Maxima for all $p \geq 2$,
 - *Global Minima:* $L = p = 4$ corresponds to Global Minima,
 - *Saddle Points:* All remaining cases are saddle points.
- Proof:* The proof is in Appendix III. ■

VI. CONCLUSION

In this article, various analytical results related to the convergence behavior of the ICA/BSS algorithms based on symmetric orthogonalization are presented. In particular,

- a convenient characterization of the stationary points of fixed point (gradient ascent) algorithms is obtained where it is shown that a point is stationary if and only if the corresponding gradient matrix is equal to the separator matrix multiplied by a positive-definite (Hermitian) matrix;
- in the convex objective function case, the monotonic convergence property of the corresponding fixed point algorithms with symmetric orthogonalization is proven;
- the modification of the fixed point algorithm to guarantee monotonic convergence in the case of nonconvex objective functions is proposed;
- an analysis of stationary points of the kurtosis objective function is provided based on the general characterization introduced previously;
- a rigorous study of the stability behavior of a new family of stationary points for kurtosis objective function (which are based on Householder transformation) is provided;
- the full rank property of the Kurtosis objective function (over the set of unitary matrices) is shown.

APPENDIX I

SUBGRADIENT INEQUALITY FOR REAL VALUED FUNCTION OF COMPLEX MATRICES

Let $\mathcal{J}(\Theta)$ is a real valued function of $\Theta \in \mathcal{C}^{p \times p}$, which is also a smooth convex function of the real vector $\mathbf{t} = \begin{bmatrix} \Re\{\mathbf{t}\} \\ \Im\{\mathbf{t}\} \end{bmatrix}$, where $\mathbf{t} = \text{vec}(\Theta)$. Defining $\mathbf{t}^{(1)} = \text{vec}(\Theta^{(1)})$, $\mathbf{t}^{(2)} = \text{vec}(\Theta^{(2)})$

$$\mathbf{t}^{(k)} = \begin{bmatrix} \Re\{\mathbf{t}^{(k)}\} \\ \Im\{\mathbf{t}^{(k)}\} \end{bmatrix} \quad k = 1, 2$$

where $\Theta_1, \Theta_2 \in \mathcal{C}^{p \times p}$, the usual subgradient inequality for \mathcal{J} , which is a convex real function of real vector \mathbf{t} can be written as [24]

$$\mathcal{J}(\mathbf{t}^{(2)}) \geq \mathcal{J}(\mathbf{t}^{(1)}) + \nabla_{\mathbf{t}} \mathcal{J}(\mathbf{t}^{(1)})^T (\mathbf{t}^{(2)} - \mathbf{t}^{(1)}).$$

Noting that

$$\begin{aligned} \nabla_{\mathbf{t}} \mathcal{J} &= \begin{bmatrix} \nabla_{\Re\{\mathbf{t}\}} \mathcal{J} \\ \nabla_{\Im\{\mathbf{t}\}} \mathcal{J} \end{bmatrix} \quad \text{and (from [14])} \\ \nabla_{\mathbf{t}} \mathcal{J} &= \frac{1}{2} (\nabla_{\Re\{\mathbf{t}\}} \mathcal{J} + i \nabla_{\Im\{\mathbf{t}\}} \mathcal{J}) \end{aligned}$$

we can write

$$\begin{aligned} \nabla_{\mathbf{t}} \mathcal{J}(\mathbf{t}^{(1)})^T (\mathbf{t}^{(2)} - \mathbf{t}^{(1)}) \\ = 2\Re \left\{ \nabla_{\mathbf{t}} \mathcal{J}(\mathbf{t}^{(1)})^H (\mathbf{t}^{(2)} - \mathbf{t}^{(1)}) \right\}. \end{aligned}$$

Using the equality

$$\begin{aligned} \nabla_{\mathbf{t}} \mathcal{J}(\mathbf{t}^{(1)})^H (\mathbf{t}^{(2)} - \mathbf{t}^{(1)}) \\ = \text{Tr} \left(\nabla_{\Theta} \mathcal{J}(\Theta^{(1)})^H (\Theta^{(2)} - \Theta^{(1)}) \right) \end{aligned}$$

we obtain the desired form of the subgradient inequality

$$\begin{aligned} \mathcal{J}(\Theta^{(2)}) \geq \mathcal{J}(\Theta^{(1)}) \\ + 2\Re \left\{ \text{Tr} \left(\nabla_{\Theta} \mathcal{J}(\Theta^{(1)})^H (\Theta^{(2)} - \Theta^{(1)}) \right) \right\}. \end{aligned}$$

APPENDIX II

FULL RANK PROPERTY OF KURTOSIS OBJECTIVE FUNCTION GRADIENT

Given a unitary matrix $\Psi \in \mathcal{C}^{p \times p}$, we can write

$$\Psi = \sum_{k=0}^p \lambda_k \mathbf{q}_k \mathbf{q}_k^H$$

where $\lambda_k \in \mathcal{C}$'s and $\mathbf{q}_k \in \mathcal{C}^{p \times 1}$ s are the eigenvalues and the corresponding eigenvectors of Ψ . Based on this expansion, the gradient of the Kurtosis objective function in (29) can be written as

$$\begin{aligned} \nabla_{\Psi} \mathcal{J}(\Psi) &= (2\Psi \odot \Psi \odot \bar{\Psi}) \mathbf{K} \\ &= \left(2 \sum_{k=1}^p \sum_{l=1}^p \sum_{m=1}^p \lambda_k \lambda_l \bar{\lambda}_m \mathbf{v}_{klm} \mathbf{v}_{klm}^H \right) \mathbf{K} \\ &= \mathbf{V} \mathbf{D} \mathbf{V}^H \mathbf{K} \end{aligned} \quad (33)$$

where $\mathbf{v}_{klm} = \mathbf{q}_k \odot \mathbf{q}_l \odot \bar{\mathbf{q}}_m$

$$\mathbf{V} = [\mathbf{v}_{111} \quad \mathbf{v}_{112} \quad \dots \quad \mathbf{v}_{ppp}]$$

and $\mathbf{D} = \text{diag}(2\lambda_1 |\lambda_1|^2, 2\lambda_1^2 \bar{\lambda}_2, \dots, 2|\lambda_p|^2 \lambda_p)$. Therefore, for the range of $\nabla_{\Psi} \mathcal{J}(\Psi)$, we can write

$$\begin{aligned} \mathcal{R}(\nabla_{\Psi} \mathcal{J}(\Psi)) &= \mathcal{R}(\mathbf{V}) \\ &= \text{Span}\{\mathbf{q}_1 \odot \mathbf{q}_1 \odot \bar{\mathbf{q}}_1 \\ &\quad \mathbf{q}_1 \odot \mathbf{q}_1 \odot \bar{\mathbf{q}}_2, \dots, \mathbf{q}_p \odot \mathbf{q}_p \odot \bar{\mathbf{q}}_p\}. \end{aligned}$$

We show that \mathbf{V} is a full-rank matrix, based on the approach introduced in [25] and [26]. We first define the mapping $g : \mathcal{C}^{p \times 1} \mapsto \mathcal{C}^{p \times 1}$ as

$$g(\mathbf{x}) = \mathbf{x} \odot \mathbf{x} \odot \bar{\mathbf{x}}.$$

Then the set in (34) would be equivalent to the range space of the mapping $f : \mathcal{C}^{p \times 1} \mapsto \mathcal{C}^{p \times 1}$ defined as $f(\mathbf{x}) = g(\mathbf{Q}\mathbf{x})$ where

$$\mathbf{Q} = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \dots \quad \mathbf{q}_p].$$

Since g is an onto mapping, and \mathbf{Q} is an invertible (unitary) matrix, the span of \mathbf{f} is equal to $\mathcal{C}^{p \times 1}$. Therefore, \mathbf{V} in (33) has full rank, which further implies that $\nabla_{\Psi} \mathcal{J}(\Psi)$ is a full-rank matrix.

APPENDIX III
PROOF OF THEOREM 6

To simplify the expressions, without loss of generality, we will assume that $\kappa = 1$.

Global Maxima and Global Minima: Without loss of generality, assuming that the first L entries of the \mathbf{v} vector is nonzero, we can write a matrix given by (32) more explicitly as shown in (34) at the bottom of the page, where $\Theta_{kl} = \angle v_k - \angle v_l$, where \angle is the phase operator.

- When $L = 1$, the matrix in (34) is a diagonal matrix. Therefore, it corresponds to a perfect separation matrix, or equivalently, a global maximum.
- When $L = 2$, first two diagonal entries are equal to zero and the corresponding matrix has only one nonzero entry per row (or per column). Therefore, this case also corresponds to a global maximum.
- When $L = p = 4$, the matrix in (34) would be equivalent to

$$\Psi = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2}e^{j\Theta_{12}} & -\frac{1}{2}e^{j\Theta_{13}} & -\frac{1}{2}e^{j\Theta_{14}} \\ -\frac{1}{2}e^{-j\Theta_{12}} & \frac{1}{2} & -\frac{1}{2}e^{j\Theta_{23}} & -\frac{1}{2}e^{j\Theta_{24}} \\ -\frac{1}{2}e^{-j\Theta_{13}} & -\frac{1}{2}e^{-j\Theta_{23}} & \frac{1}{2} & -\frac{1}{2}e^{j\Theta_{34}} \\ -\frac{1}{2}e^{-j\Theta_{14}} & -\frac{1}{2}e^{-j\Theta_{24}} & -\frac{1}{2}e^{-j\Theta_{34}} & \frac{1}{2} \end{bmatrix} \quad (35)$$

which is a matrix with equal magnitude ($(1/\sqrt{p})$) entries that corresponds to a global minimum of the objective function.

Saddle Points: In order to show that the remaining cases correspond to saddle points, we start by defining

$$\Omega = \Psi \odot \bar{\Psi} = \mathbf{I} - 4\text{diag}(\mathbf{a}) + 4\mathbf{a}\mathbf{a}^T$$

Based on this definition, we can write

$$\begin{aligned} \mathcal{J}_{\text{MUK}}(\Psi) &= \|\Omega\|_F^2 = \text{Tr}(\Omega^H \Omega) \\ &= \text{Tr}((\mathbf{I} - 4\text{diag}(\mathbf{a}) + 4\mathbf{a}\mathbf{a}^T)^T \\ &\quad \times (\mathbf{I} - 4\text{diag}(\mathbf{a}) + 4\mathbf{a}\mathbf{a}^T)). \end{aligned}$$

After some simplifications, assuming L nonzero elements, we obtain

$$\mathcal{J}_{\text{MUK}}(\Psi) = p - 8 + 24\|\mathbf{a}\|_2^2 - 32\|\mathbf{a}\|_3^3 + 16\|\mathbf{a}\|_4^4 \quad (36)$$

where $\|\mathbf{a}\|_N = (\sum_{k=1}^p |a_k|^N)^{1/N}$, which can be further simplified to

$$\mathcal{J}_{\text{MUK}}(\Psi_c) = p - 8 + 8 \left(3\frac{1}{L} - 2\frac{1}{L^2} \right).$$

Based on the expression in (36) we can perform the stability analysis as follows.

- When $2 < L < p$:
 - Assuming one nonzero component of \mathbf{a} increased by a small positive amount ϵ while another nonzero component of \mathbf{a} decreased by the same amount (preserving $\|\mathbf{v}\|_2 = 1$), then the objective function at the perturbed point is given by

$$\begin{aligned} \mathcal{J}_{\text{MUK}}(\Psi') &= p - 8 + 24 \left(\frac{L-2}{L^2} + \left(\frac{1}{L} - \epsilon \right)^2 + \left(\frac{1}{L} + \epsilon \right)^2 \right) \\ &\quad - 32 \left(\frac{L-2}{L^3} + \left(\frac{1}{L} + \epsilon \right)^3 + \left(\frac{1}{L} - \epsilon \right)^3 \right) \\ &\quad + 16 \left(\frac{L-2}{L^2} + \left(\frac{1}{L} - \epsilon \right)^2 + \left(\frac{1}{L} + \epsilon \right)^2 \right)^2 \\ &= \mathcal{J}_{\text{MUK}}(\Psi_c) + \frac{48L - 128}{L} \epsilon^2 + 64\epsilon^4. \end{aligned}$$

Therefore, the prescribed perturbation leads to an increase in the MUK objective function. Therefore, the corresponding critical point is not a maximum point.

- Assuming that one nonzero component of \mathbf{a} increased by a small positive amount ϵ while a zero component of \mathbf{a} increased by the same amount, then the objective function at the perturbed point is given by

$$\begin{aligned} \mathcal{J}_{\text{MUK}}(\Psi') &= p - 8 + 24 \left(\frac{L-1}{L^2} + \left(\frac{1}{L} - \epsilon \right)^2 + \epsilon^2 \right) \\ &\quad - 32 \left(\frac{L-1}{L^3} + \left(\frac{1}{L} + \epsilon \right)^3 + \epsilon^3 \right) \\ &\quad + 16 \left(\frac{L-1}{L^2} + \left(\frac{1}{L} - \epsilon \right)^2 + \epsilon^2 \right)^2 \end{aligned}$$

$$\begin{aligned} \Psi &= \mathbf{I} - 2\mathbf{v}\mathbf{v}^H \\ &= \begin{bmatrix} (1 - \frac{2}{L}) & -\frac{2}{L}e^{j\Theta_{12}} & \dots & -\frac{2}{L}e^{j\Theta_{1L}} & 0 & 0 & \dots & 0 \\ -\frac{2}{L}e^{-j\Theta_{12}} & (1 - \frac{2}{L}) & \dots & -\frac{2}{L}e^{j\Theta_{2L}} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{2}{L}e^{-j\Theta_{1L}} & -\frac{2}{L}e^{-j\Theta_{2L}} & \dots & (1 - \frac{2}{L}) & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \end{bmatrix} \quad (34) \end{aligned}$$

$$\begin{aligned}
 &= \mathcal{J}_{\text{MUK}}(\Psi_c) + \frac{32 - 48L}{L^2} \epsilon \\
 &\quad + \frac{48L^2 - 32L + 64}{L^2} \epsilon^2 - \frac{128}{L} \epsilon^3 + 64\epsilon^4
 \end{aligned}$$

For $2 < L < p$ and sufficiently small ϵ , the sum of the terms with positive powers of ϵ in the above expression will always be negative. Therefore, the prescribed perturbation leads to a decrease in the objective function.

— As a result for $2 < L < p$, the corresponding Ψ_c 's are neither maxima nor minima. They are nonstable saddle points.

• The case with $2 < L = p$ needs a special treatment:

— The case with $p = 4$ is previously shown to correspond to global minima.

— For other cases, we need to analyze the perturbation in more detail, i.e., the behavior of the objective function in all neighboring points. For this purpose, we parameterize the neighbors of the critical point Ψ_c with

$$\Psi_{\epsilon, \mathbf{A}} = \Psi_c e^{\epsilon \mathbf{A}}$$

where \mathbf{A} is a skew-Hermitian matrix with Frobenius norm equal to 1, i.e., $\|\mathbf{A}\|_F = 1$, and $\epsilon \neq 0$ is a small real number. For sufficiently small ϵ

$$\|\Psi_c - \Psi_{\epsilon, \mathbf{A}}\|_F \approx \epsilon.$$

Based on the power series expansion

$$\Psi_{\epsilon, \mathbf{A}} = \Psi_c \sum_{k=0}^{\infty} \epsilon^k \frac{\mathbf{A}^k}{k!}$$

we can write the difference between critical point and its neighbor as

$$\begin{aligned}
 \Psi_{\epsilon, \mathbf{A}} - \Psi_c &= \Psi_c \sum_{k=0}^{\infty} \epsilon^k \frac{\mathbf{A}^k}{k!} - \Psi_c = \Psi_c \sum_{k=1}^{\infty} \epsilon^k \frac{\mathbf{A}^k}{k!} \\
 &= \epsilon \Psi_c \mathbf{A} + \frac{\epsilon^2}{2} \Psi_c \mathbf{A}^2 + \dots
 \end{aligned}$$

Using this expression, the value of the objective function at the neighboring point can be written as

$$\begin{aligned}
 \mathcal{J}(\Psi_{\epsilon, \mathbf{A}}) &= \mathcal{J}(\Psi_c) \\
 &+ \underbrace{2\Re\left\{\text{Tr}\left(\nabla_{\Psi} \mathcal{J}(\Psi_c)^H \left(\epsilon \Psi_c \mathbf{A} + \frac{\epsilon^2}{2} \Psi_c \mathbf{A}^2\right)\right)\right\}}_{\Delta \mathcal{J}_1(\epsilon, \mathbf{A})} \\
 &+ \underbrace{\Re\left\{\text{vec}(\epsilon \Psi_c \mathbf{A})^H \mathbf{H}_{\text{tt}}(\Psi_c) \text{vec}(\epsilon \Psi_c \mathbf{A})\right\}}_{\Delta \mathcal{J}_2(\epsilon, \mathbf{A})} \\
 &+ \underbrace{\Re\left\{\text{vec}(\epsilon \Psi_c \mathbf{A})^H \mathbf{H}_{\text{tt}}(\Psi_c) \overline{\text{vec}(\epsilon \Psi_c \mathbf{A})}\right\}}_{\Delta \mathcal{J}_3(\epsilon, \mathbf{A})} + o(\epsilon^3). \quad (37)
 \end{aligned}$$

The terms $\Delta \mathcal{J}_1(\epsilon, \mathbf{A})$, $\Delta \mathcal{J}_2(\epsilon, \mathbf{A})$, $\Delta \mathcal{J}_3(\epsilon, \mathbf{A})$ in (37) can be reduced as follows:

$$* \Delta \mathcal{J}_1(\epsilon, \mathbf{A})$$

$$\begin{aligned}
 \Delta \mathcal{J}_1(\epsilon, \mathbf{A}) &= 2\Re\left\{\text{Tr}\left(\nabla_{\Psi} \mathcal{J}(\Psi_c)^H \left(\epsilon \Psi_c \mathbf{A} + \frac{\epsilon^2}{2} \Psi_c \mathbf{A}^2\right)\right)\right\} \\
 &= 2\Re\left\{\text{Tr}\left(\nabla_{\Psi} \mathcal{J}(\Psi_c)^H \left(\frac{\epsilon^2}{2} \Psi_c \mathbf{A}^2\right)\right)\right\} \quad (38)
 \end{aligned}$$

which is due to the fact that, at the critical point Ψ_c

$$\nabla_{\Psi} \mathcal{J}(\Psi_c) = \Psi_c \mathbf{S} \quad \text{for some } \mathbf{S} = \mathbf{S}^*$$

and therefore

$$\begin{aligned}
 &\Re\left\{\text{Tr}(\nabla_{\Psi} \mathcal{J}(\Psi_c)^H (\epsilon \Psi_c \mathbf{A}))\right\} \\
 &= \Re\left\{\text{Tr}(\mathbf{S} \Psi_c^H (\epsilon \Psi_c \mathbf{A}))\right\} \\
 &= \epsilon \Re\left\{\text{Tr}(\mathbf{S} \mathbf{A})\right\} = 0.
 \end{aligned}$$

For the $L = p$ case, the gradient $\nabla_{\Psi} \mathcal{J}(\Psi_c)$ can be written as

$$\begin{aligned}
 \nabla_{\Psi} \mathcal{J}(\Psi_c) &= 2\Psi_c \odot \Psi_c \odot \bar{\Psi}_c \\
 &= 2\Psi_c \odot (\mathbf{I} - 2\mathbf{v}\mathbf{v}^H) \odot (\mathbf{I} - 2\bar{\mathbf{v}}\bar{\mathbf{v}}^T) \\
 &= 2\Psi_c \left(\mathbf{I} \left(1 - \frac{4}{p}\right) + \frac{4}{p^2} \mathbf{1}\mathbf{1}^T \right) \\
 &= \left(2 \left(1 - \frac{6}{p} + \frac{12}{p^2}\right) \mathbf{I} - \frac{16}{p^2} \mathbf{v}\mathbf{v}^H \right)
 \end{aligned}$$

Based on this expression, we can further simplify (38) as follows:

$$\begin{aligned}
 \Delta \mathcal{J}_1(\epsilon, \mathbf{A}) &= \epsilon^2 \Re\left\{\text{Tr}(\nabla_{\Psi} \mathcal{J}(\Psi_c)^H (\Psi_c \mathbf{A}^2))\right\} \\
 &= \epsilon^2 \Re\left\{\text{Tr}\left(\left(2 \left(1 - \frac{6}{p} + \frac{12}{p^2}\right) \mathbf{I} - \frac{16}{p^2} \mathbf{v}\mathbf{v}^H\right) (\mathbf{I} - 2\mathbf{v}\mathbf{v}^H) \mathbf{A}^2\right)\right\} \\
 &= \epsilon^2 \Re\left\{\text{Tr}\left(\left(2 \left(1 - \frac{6}{p} + \frac{12}{p^2}\right) \mathbf{I} - \left(4 - \frac{24}{p} + \frac{32}{p^2}\right) \mathbf{v}\mathbf{v}^H\right) \mathbf{A}^2\right)\right\} \\
 &= \epsilon^2 \left(2 \frac{p^2 - 6p + 12}{p^2} \Re\left\{\text{Tr}(\mathbf{A}^2)\right\} - 4 \frac{p^2 - 6p + 8}{p^2} \Re\left\{\text{Tr}(\mathbf{v}\mathbf{v}^H \mathbf{A}^2)\right\} \right) \\
 &= \epsilon^2 \left(-2 \frac{p^2 - 6p + 12}{p^2} - 4 \frac{p^2 - 6p + 8}{p^2} \mathbf{v}^H \mathbf{A}^2 \mathbf{v} \right) \\
 &= \epsilon^2 \left(-2 \frac{p^2 - 6p + 12}{p^2} + 4 \frac{p^2 - 6p + 8}{p^2} \|\mathbf{A}\mathbf{v}\|_2^2 \right) \quad (39)
 \end{aligned}$$

where we used the fact that for a skew-Hermitian \mathbf{A} with $\|\mathbf{A}\|_F = 1$, $\Re\{\text{Tr}(\mathbf{A}^2)\}$ is equal to -1 .

* $\Delta\mathcal{J}_2(\epsilon, \mathbf{A})$:

$$\begin{aligned} \Delta\mathcal{J}_2(\epsilon, \mathbf{A}) &= \epsilon^2 \text{vec}(\Psi_c \mathbf{A})^H 4 \text{diag}(\text{vec}(\Psi_c \bar{\Psi}_c)) \text{vec}(\Psi_c \mathbf{A}) \\ &= 4\epsilon^2 \|\bar{\Psi}_c \odot (\Psi_c \mathbf{A})\|_F^2 \\ &= 4\epsilon^2 \left(\left(1 - \frac{4}{p}\right) \sum_{k=1}^p |A_{kk}|^2 \right. \\ &\quad \left. + \left(\frac{16}{p} - 2\right) \sum_{k=1}^p |A_{kk}| \text{Im}\{v_k \xi_k\} \right. \\ &\quad \left. + 16 \sum_{k=1}^p |v_k|^2 |\xi_k|^2 + \frac{4}{p^2} - \frac{16}{p^2} \|\xi\|_2^2 \right) \end{aligned} \quad (40)$$

where ξ is a vector defined as $\xi = \mathbf{v}^H \mathbf{A}$ and whose k th element is denoted by ξ_k , v_k is the k th component of vector \mathbf{v} and we skipped some algebraic simplification details due to length constraints.

* $\Delta\mathcal{J}_3(\epsilon, \mathbf{A})$

$$\begin{aligned} \Delta\mathcal{J}_3(\epsilon, \mathbf{A}) &= \epsilon^2 \Re \left\{ \text{vec}(\Psi_c \mathbf{A}) 2 \text{diag}(\text{vec}(\Psi_c^{\odot 2})) \overline{\text{vec}(\Psi_c \mathbf{A})} \right\} \\ &= 2\epsilon^2 \left(\left(\frac{4}{p} - 1\right) \sum_{k=1}^p |A_{kk}|^2 \right. \\ &\quad \left. + \left(4 - \frac{16}{p}\right) \sum_{k=1}^p |A_{kk}| \text{Im}\{v_k \xi_k\} \right. \\ &\quad \left. + \left(4 - \frac{16}{p}\right) \Re \left\{ \left(\sum_{k=1}^p v_k^* \xi_k^* \right)^2 \right\} \right. \\ &\quad \left. + \frac{4}{p^2} \sum_{k=1}^p \sum_{l=1}^p e^{i2\Theta_{kl}} (A_{kl}^*)^2 \right) \end{aligned} \quad (41)$$

where $\Theta_{kl} = \angle v_k v_l^*$.

Finally, the (37), (39), (40), (41) can be combined to obtain

$$\mathcal{J}(\Psi_{\epsilon, \mathbf{A}}) - \mathcal{J}(\Psi_c) = \alpha(\mathbf{A})\epsilon^2 + o(\epsilon^3)$$

where

$$\begin{aligned} \alpha(\mathbf{A}) &= -2 \frac{p^2 - 6p + 4}{p^2} + 4 \frac{p^2 - 6p - 8}{p^2} \|\xi\|_2^2 \\ &\quad + \left(2 - \frac{8}{p}\right) \sum_{k=1}^p |A_{kk}|^2 + 16 \sum_{k=1}^p |v_k|^2 |\xi_k|^2 \\ &\quad + \left(\frac{32}{p} - 8\right) \sum_{k=1}^p (\text{Im}\{A_{kk}\} \text{Im}\{v_k \xi_k\} \\ &\quad - \Re\{(v_k^* \xi_k^*)^2\}) \\ &\quad + \frac{8}{p^2} \sum_{k=1}^p \sum_{l=1}^p e^{i2\Theta_{kl}} (A_{kl}^*)^2. \end{aligned} \quad (42)$$

Using the expression in (42), we analyze three different cases of \mathbf{A} .

* Case 1): If we choose $\mathbf{A}_1 = i\mathbf{v}\mathbf{v}^H$, then $\xi = i\mathbf{v}^H$ and therefore $\|\xi\| = 1$. The expression in (42) simplifies to

$$\alpha(\mathbf{A}_1) = 2 \frac{p^2 - 5p + 4}{p^2} = 2 \frac{(p-1)(p-4)}{p^2}$$

which is negative for $1 < p < 4$ and positive for $p > 4$.

* Case 2): If we choose $\mathbf{A}_2 = i\boldsymbol{\eta}\boldsymbol{\eta}^H$, where $\boldsymbol{\eta} = \sqrt{(p/2)} [v_2^* \quad -v_1^* \quad \mathbf{0}]^T$, then $\xi = \mathbf{0}$. In this case, we obtain

$$\alpha(\mathbf{A}_2) = -\frac{(p-4)^2}{p^2}$$

which is negative for all $p \neq 4$.

* Case 3): if we choose $\mathbf{A}_3 = i\boldsymbol{\eta}\boldsymbol{\eta}^H$, where

$$\boldsymbol{\eta} = \mathbf{f} \odot \mathbf{v},$$

and \mathbf{f} is the complex exponential vector

$$\mathbf{f} = [1 \quad e^{i\frac{2\pi}{p}} \quad e^{2i\frac{2\pi}{p}} \quad \dots \quad e^{(p-1)i\frac{2\pi}{p}}]^T$$

then $\xi = \mathbf{0}$. Since in (42)

$$\frac{8}{p^2} \sum_{k=1}^p \sum_{l=1}^p e^{i2\Theta_{kl}} (A_{kl}^*)^2 = \begin{cases} -\frac{8}{p^2} & p=2 \\ 0 & p>2 \end{cases}$$

we obtain

$$\alpha(\mathbf{A}_3) = \begin{cases} \frac{-2p^2 + 14p - 24}{p^2} = -1, & p=2 \\ \frac{-2p^2 + 14p - 16}{p^2}, & p>2 \end{cases}$$

Therefore, $\alpha(\mathbf{A}_3)$ is positive for $3 \leq p \leq 5$ and negative for $p=2$ or $p>5$.

—Based on Cases 1) and 2), we can conclude that, for $p > 4$, the perturbation in the “direction” determined by \mathbf{A}_1 causes increase in the objective function while the perturbation in the “direction” determined by \mathbf{A}_2 causes decrease.

For $p=3$, based on Cases 1) and 3), we observe that the perturbation in the “direction” determined by \mathbf{A}_1 causes decrease in the objective function while the perturbation in the “direction” determined by \mathbf{A}_3 causes an increase. As a result, we conclude that for the $2 < L = p$ case, the corresponding Householder reflection based critical points are saddle points except for $p=4$ which corresponds to global minima.

ACKNOWLEDGMENT

The author would like to thank anonymous reviewers for their comments which contributed to the quality of the article.

REFERENCES

- [1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Hoboken, NJ: Wiley, 2001.

- [2] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1483–1492, Oct. 1997.
- [3] P. A. Regalia and E. Kofidis, "Monotonic convergence of fixed-point algorithms for ICA," *IEEE Trans. Neural Netw.*, vol. 14, no. 4, pp. 943–949, July 2003.
- [4] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Mar. 2002.
- [5] A. T. Erdogan, "Globally convergent deflationary instantaneous blind source separation algorithm for digital communication signals," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2182–2192, May 2007.
- [6] H. Shen and K. Huper, "Local convergence analysis of FastICA," *Lecture Notes Comput. Sci.*, vol. 3889, pp. 893–900, Feb. 2006.
- [7] E. Oja, "Convergence of the symmetrical FastICA algorithm," in *Proc. 9th Int. Conf. Neural Inf. Process.*, Nov. 2002, vol. 3, pp. 1368–1372.
- [8] E. Oja and Z. Yuan, "The FastICA algorithm revisited: Convergence analysis," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1370–1381, Nov. 2006.
- [9] P. Tichavsky, Z. Koldovsky, and E. Oja, "Performance analysis of the FastICA algorithm and Cramer–Rao bounds for linear independent component analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1189–1203, Apr. 2006.
- [10] A. T. Erdogan, "On the convergence of ICA algorithms with symmetric orthogonalization," in *Proc. ICASSP*, Mar.–Apr. 2008, pp. 1925–1928.
- [11] C. Papadias, "Globally convergent blind source separation based on a multiuser kurtosis maximization criterion," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3508–3519, Dec. 2000.
- [12] H. Li and T. Adali, "Gradient and fixed-point complex ICA algorithms based on kurtosis maximization," in *Proc. 16th IEEE Signal Process. Soc. Workshop Machine Learning for Signal Process.*, Sep. 2006, pp. 85–90.
- [13] A. van den Bos, "Complex gradient and hessian," *Proc. IEE Vision Image Signal Processing*, vol. 141, no. 6, pp. 380–383, Dec. 1994.
- [14] K. Kreutz-Delgado, The Complex Gradient Operator and the Cr-Calculus: Ece275a-Lecture Supplement: Fall 2006.
- [15] A. Hjørnungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.
- [16] A. Hjørnungnes, D. Gesbert, and D. P. Palomar, "Unified theory of complex-valued matrix differentiation," in *Proc. IEEE ICASSP 2007*, Apr. 2007, vol. 3, pp. 345–348.
- [17] H. Li and T. Adali, "Optimization in the complex domain for nonlinear adaptive filtering," in *Proc. 40th Asilomar Conf. Signals Syst. Comput.*, 2006, pp. 263–267.
- [18] P. A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton, 2008.
- [19] D. P. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization*. New York: Athena Scientific, 2003.
- [20] T. Söderström, "Perturbation results for singular values," Uppsala Univ., Uppsala, Sweden, Tech. Rep. 1999-001, Apr. 1999.
- [21] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inf. Theory*, vol. 36, pp. 312–321, Mar. 1990.
- [22] O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Trans. Inf. Theory*, vol. 39, pp. 504–519, Mar. 1993.
- [23] P. Regalia and M. Mboup, "Undermodeled equalization: A characterization of stationary points for a family of blind criteria," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 760–770, Mar. 1999.
- [24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Stanford, CA: Stanford Univ., 2001.
- [25] G. Gorni and H. Tutaj-Gasińska, "On the entrywise powers of matrices," *Commun. Algebra*, vol. 32, pp. 495–520, 2004.
- [26] X. Sun, X. Du, and D. Liu, "On the range of a Hadamard power of a positive semidefinite matrix," *Linear Algebra Appl.*, vol. 416, pp. 868–871, 2006.



Alper T. Erdogan was born in Ankara, Turkey, in 1971. He received his B.S. degree from the Middle East Technical University, Ankara, Turkey, in 1993, and the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, in 1995 and 1999, respectively.

He was a Principal Research Engineer in Globespan-Virata Corporation (formerly Excess Bandwidth and Virata Corporations) from September 1999 to November 2001. He joined Electrical and Electronics Engineering Department of Koc University, Istanbul, Turkey, in January 2002, where he is

currently an Associate Professor. His research interests include wireless, fiber and wireline communications, adaptive signal processing, optimization, system theory and control, and information theory.

Dr. Erdogan is the recipient of TUBITAK Career Award (2005), Werner Von Siemens Excellence Award (2007) and TUBA GEBIP Outstanding Young Scientist Award (2008). He currently serves as an Associate Editor for IEEE TRANSACTION ON SIGNAL PROCESSING.