BLIND SIGNAL SEPARATION OF ARBITRARY MIXTURES: ADAPTIVE ALGORITHMS AND STABILITY ANALYSES

Scott C. Douglas and Xiaoan Sun

Department of Electrical Engineering
Southern Methodist University
Dallas, Texas 75275 USA

ABSTRACT

In this paper, we present novel techniques for blindly separating mixtures of instantaneously-mixed sources. The algorithms are able to separate mixtures of arbitrary non-zero-kurtosis sources without specific knowledge of and without estimating the statistics of the individually-extracted sources. Both orthogonal contrast and equivariant prewhitening/source separation implementations are provided. Local stability analyses and simulations of the proposed methods verify their abilities to separate arbitrary source mixtures.

1. INTRODUCTION

Blind signal separation (BSS) refers to a general class of approaches for extracting individual source signals from a set of linear mixtures with little to no knowledge of the underlying signals' characteristics. Typically, the mixing process is linear, such that a linear demixing transformation of the mixture components produces the desired separated signals. The goal of BSS is to determine a procedure by which to adjust the parameters of the demixing system without specific knowledge of the mixing process nor of the source signals' characteristics.

Generally, BSS approaches fall into one of two classes:

- *density-based* approaches that attempt to match as closely as possible the joint density of the transformed outputs to a candidate density whose components are independent, and
- *contrast-based* approaches that employ a cost function or contrast whose extrema over the set of candidate demixing transformations correspond to true demixing transformations.

Algorithms for density-based signal separation have been developed using the classic concepts of minimum entropy, maximum likelihood, minimum mutual information, and the like [1]–[3]. Such algorithms are most-useful when the source signals have similar densities; otherwise, a computationally-expensive procedure for estimating the marginal densities of the candidate output signals is required [3]. Other *ad hoc* procedures for adjusting the marginal density models have also been proposed [4] but are generally unsatisfactory when no knowledge of the source signals is available. In such cases, contrast-based approaches are to be preferred, as they offer much more robust estimating qualities with a minimum of computational effort.

In this paper, we develop new adaptive on-line procedures for contrast-based blind signal separation of arbitrary source mixtures. The proposed algorithms are non-trivial extensions of the the iterative kurtosis-based extraction procedures described in [5, 6] as well as the EASI algorithm described in [7]. The unique and salient features of the proposed methods are as follows:

1. All of the methods are able to separate mixtures of arbitrary non-zero-kurtosis sources.
2. The only knowledge required about the source signal distributions are the numbers of positive- and negative-kurtosis sources contained in the mixtures.
3. The methods do not attempt to estimate the statistics of the extracted sources; thus, their separation abilities do not depend on such estimation procedures.
4. The methods' complexities scale linearly with the size of the demixing matrix.

We provide detailed local stability analyses of all of the proposed orthogonal contrast-based BSS methods, through which their abilities to separate arbitrary signal mixtures is indicated. Simulations show that the proposed methods can separate mixtures of positive- and negative-kurtosis sources without specific knowledge of each source's distribution.

2. REVIEW OF CONTRAST-BASED BSS

Fig. 1 outlines the structure of contrast-based BSS. The source vector sequence $\mathbf{s}(k) = [s_1(k) \ \cdots \ s_n(k)]^T$ passes through an unknown $(n \times n)$ mixing matrix $\mathbf{A}$ before being measured as the vector sequence $\mathbf{x}(k) = [x_1(k) \ \cdots \ x_n(k)]^T$. We have assumed that the number of sources equals the number of sensors, although the algorithms in this paper can be used in the more-sensors-than-sources case as well.

Two matrices are employed in the separation process. The first matrix $\mathbf{P}(k)$ performs signal prewhitening, such that the prewhitened sequence $\mathbf{v}(k)$ defined as

$$\mathbf{v}(k) = \mathbf{P}(k)\mathbf{x}(k) = \mathbf{P}(k)\mathbf{A}\mathbf{s}(k) \qquad (1)$$

contains uncorrelated elements with unit variance, or

$$E\{\mathbf{v}(k)\mathbf{v}^T(k)\} = \mathbf{I}, \qquad (2)$$

This material is based in part upon work supported by the Texas Advanced Technology Program under Grant No. 003613-0031-1999.

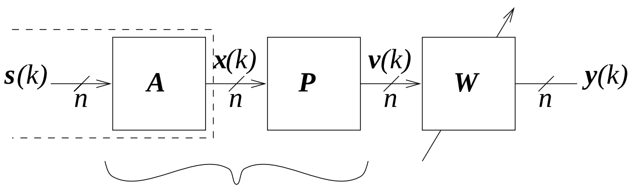

Fig. 1: Block diagram of contrast-based BSS.

where $E\{\cdot\}$ denotes statistical expectation. If $\mathbf{P}(k)$ is ideally estimated, the prewhitened mixing matrix defined as

$$\Gamma = \mathbf{PA} \quad (3)$$

is Hermitian, such that $\Gamma\Gamma^T = \Gamma^T\Gamma = \mathbf{I}$. Generally, $\mathbf{P}$ can be estimated using numerically-robust procedures such as the singular-value-decomposition (SVD), and adaptive procedures for principal component analysis can be modified to perform the prewhitening task. See [8] for more details in PCA methods.

The second matrix within this system, denoted as $\mathbf{W}(k)$, is adapted using a contrast-based BSS algorithm to solve the following constrained optimization problem:

$$\text{maximize} \quad \mathcal{J}(\mathbf{W}(k)) = \sum_{i=1}^{n} E\{\phi_i(y_i(k))\} \quad (4)$$

$$\text{such that} \quad \mathbf{W}(k)\mathbf{W}^T(k) = \mathbf{I}, \quad (5)$$

where

$$\mathbf{y}(k) = \mathbf{W}(k)\mathbf{v}(k) \quad (6)$$

and $\phi_i(y)$ is the ith contrast function. This procedure is successful if $\mathbf{W}(k)$ converges to the combined system matrix solution

$$\mathbf{C}(k) = \mathbf{W}(k)\Gamma \longrightarrow \Phi\mathbf{J}, \quad (7)$$

where Φ is an $(n\times n)$ permutation matrix and $\mathbf{J}$ is a diagonal matrix of ± 1's.

While numerous algorithms could be considered, we shall focus on gradient approaches to contrast-based BSS in this paper. Gradient methods are computationally-simple and usually require less information about the contrast function surface as compared to other methods. In addition, gradient BSS techniques can easily extended to the single- and multichannel blind deconvolution tasks [9].

The choices of the contrast functions $\phi_i(y)$ in (4) play a critical role in the success of contrast-based BSS. Convergence of $\mathbf{W}(k)$ to (4) depends on the interaction of the source probability density functions (p.d.f.'s) and the chosen $\phi_i(y)$. For example, choosing

$$\phi_i(y) = \overline{\phi}_i(y) = \text{sgn}[\kappa_i]|y|^4 \quad (8)$$

for all $1 \leq i \leq n$, where $\kappa_i = E\{|y_i(k)|^4\} - 3$ is the kurtosis of the ith extracted unit-variance signal guarantees the extraction of all non-zero-kurtosis sources in $\mathbf{s}(k)$ [5, 10, 11, 12]. Note that this property is not generally shared by other even-symmetric functions $\phi_i(y)$; see [13] for a more-detailed discussion of these issues. For this reason, we shall focus on methods that use kurtosis-based criteria in what follows.

It is possible to combine the prewhitening and orthogonal separation matrices into one matrix $\mathbf{B}(k)$ as

$$\mathbf{B}(k) = \mathbf{W}(k)\mathbf{P}(k). \quad (9)$$

Contrast-based BSS is then formulated as

$$\text{maximize} \quad \mathcal{J}(\mathbf{B}(k)) = \sum_{i=1}^{n} E\{\phi_i(y_i(k))\} \quad (10)$$

$$\text{such that} \quad E\{\mathbf{y}(k)\mathbf{y}^T(k)\} = \mathbf{I}. \quad (11)$$

Algorithms for performing this joint optimization in an iterative fashion were pioneered in [7]. In this paper, we shall focus on the orthogonalized contrast formulation in (4)–(5) for purposes of derivation. Then, we show how the resulting algorithms can be combined with simple prewhitening procedures to iteratively solve (10)–(11).

3. ANALYSIS

3.1. Analysis of an Ordered Rotation Update

The novel contrast-based methods described in this paper are based on the work in [6], in which the following ordered-rotation update was first proposed:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mathbf{D}_\beta\left\{\text{tri}[\mathbf{W}(k)\mathbf{W}^T(k)]\mathbf{f}(\mathbf{y}(k))\mathbf{v}^T(k) - \text{tri}[\mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k))]\mathbf{W}(k)\right\}, \quad (12)$$

where

$$\mathbf{W}(k) = [\mathbf{w}_1(k) \cdots \mathbf{w}_n(k)]^T \quad (13)$$

$$\mathbf{f}(\mathbf{y}) = [f_1(y_1) \cdots f_n(y_n)]^T \quad (14)$$

$$f_i(y) = \frac{\partial\phi_i(y)}{\partial y} \quad (15)$$

$$\left[\text{tri}[\mathbf{W}(k)\mathbf{W}^T(k)]\right] = \begin{cases} \mathbf{w}_i^T(k)\mathbf{w}_j(k) & \text{if } i \leq j \\ 0 & \text{if } i > j \end{cases} \quad (16)$$

and $\mathbf{D}_\beta$ is a diagonal matrix with step size entries β_j. While an analysis of (12) was provided in [6], the results do not clearly delineate the separation capabilities of this procedure.

Recently, the behaviors of a different set of prewhitened BSS methods has been analyzed using an ordinary differential equation (ODE) approach [14, 15]. Here, we perform a similar analysis of (12) by considering its corresponding averaged ODE

$$\frac{d\mathbf{W}}{dt} = \mathbf{D}_\beta\left\{\text{tri}[\mathbf{W}\mathbf{W}^T]E\{\mathbf{f}(\mathbf{y})\mathbf{v}^T\} - \text{tri}[E\{\mathbf{y}\mathbf{f}^T(\mathbf{y})\}]\mathbf{W}\right\} \quad (17)$$

to determine the exact conditions on β_i, $f_i(y)$, and the distribution of the extracted outputs that guarantee a stable separating solution. The local stability properties of (17) are described by the following theorem and proof.

Theorem 1: *Stability of (17) About a Separating Solution.*

1.1 For small perturbations about $\mathbf{W} = \Phi\mathbf{J}\Gamma^T$ that maintain decorrelation and separation of the system outputs such that $\mathbf{y} = \Phi\mathbf{J}\mathbf{s}$, (17) is locally-stable in the prewhitened BSS task for values of β_i satisfying

$$\beta_i\left[E\{y_i f_i(y_i)\} - E\{f_i'(y_i)\}\right] > 0. \quad (18)$$

1.2 For small perturbations about $\mathbf{W} = \Phi\mathbf{J}\Gamma^T$ that only alter the magnitudes of the decorrelated and separated system outputs, (17) is marginally stable in the prewhitened BSS task for all $\beta \neq 0$.

Proof: We may consider the differential update in the combined system matrix

$$\mathbf{C} = \mathbf{W}\Gamma \tag{19}$$

by post-multiplying (17) on both sides by Γ and inserting $\Gamma\Gamma^T$ between $\mathbf{W}$ and $\mathbf{W}^T$ on the RHS of this update. This operation produces

$$\frac{d\mathbf{C}}{dt} = \mathbf{D}_\beta \big\{ \mathrm{tri}[\mathbf{C}\mathbf{C}^T] E\{\mathbf{f}(\mathbf{C}\mathbf{s})\mathbf{s}^T\} - \mathrm{tri}[\mathbf{C}E\{\mathbf{s}\mathbf{f}^T(\mathbf{C}\mathbf{s})\}]\mathbf{C} \big\} \tag{20}$$

Assuming even-symmetric source distributions and contrast functions, we only need to consider local stability about

$$\mathbf{C} \quad = \quad (\mathbf{I} + \Delta)\Phi, \tag{21}$$

where the magnitudes of the elements of the perturbation matrix Δ satisfy $|\Delta_{ij}(t)| \ll n^{-1}$ at time $t = 0$. Employing a Taylor series expansion for $\mathbf{f}(\mathbf{C}\mathbf{s})$, we obtain

$$\mathbf{f}(\mathbf{C}\mathbf{s}) \quad = \quad \mathbf{f}(\mathbf{y}) + \mathbf{F}'(\mathbf{y})\Delta\mathbf{y} + \mathcal{O}(\{\Delta_{ij}\Delta_{kl}\}), \tag{22}$$

where $\mathbf{y} = \Phi\mathbf{J}\mathbf{s}$, $\mathbf{F}'(\mathbf{y})$ is a diagonal matrix whose ith diagonal entry is $f_i'(y_i)$ and $\mathcal{O}(\{\Delta_{ij}\Delta_{kl}\})$ denotes terms of second- and higher-order in the elements of Δ. Substitution of (21) and (22) into (20) gives

$$\begin{aligned} \frac{d\Delta}{dt} \quad = \quad & \mathbf{D}_\mu \big[(\mathrm{tri}[\Delta + \Delta^T]\Lambda(\mathbf{y}) + E\{\mathbf{F}'(\mathbf{y})\Delta\mathbf{y}\mathbf{y}^T\} \\ & - \mathrm{tri}[\Delta\Lambda(\mathbf{y}) + E\{\mathbf{y}\mathbf{y}^T\Delta^T\mathbf{F}'(\mathbf{y})\}] - \Lambda(\mathbf{y})\Delta \big] \\ & + \mathcal{O}(\{\Delta_{ij}\Delta_{kl}\}), \end{aligned} \tag{23}$$

where $\Lambda(\mathbf{y})$ as a diagonal matrix whose (i,i)th entry is $E\{y_i f_i(y_i)\}$. We now evaluate the expectations in (23), noting that each $f_i(y)$ is an odd function and the marginal density of each s_i is symmetric. This operation yields a set of n^2 coupled equations. For $i > j$, we have

$$\frac{d\Delta_{ij}}{dt} \quad = \quad -\beta_i \left[E\{y_i f_i(y_i)\} - E\{y_j^2 f_i'(y_i)\} \right] \Delta_{ij} \tag{24}$$

$$\begin{aligned} \frac{d\Delta_{ji}}{dt} \quad = \quad & -\beta_j \left[E\{y_j f_j(y_j)\} - E\{y_i^2 f_j'(y_j)\} \right] \Delta_{ji} \\ & + \beta_j \left[E\{y_i f_i(y_i)\} - E\{y_j^2 f_i'(y_i)\} \right] \Delta_{ij}, \end{aligned} \tag{25}$$

and for $1 \leq i \leq n$, we have

$$\frac{d\Delta_{ii}}{dt} \quad = \quad 0, \tag{26}$$

where terms of second- and higher-order have been neglected.

From (24)–(25), the behavior of Δ_{ij} and Δ_{ji} are pairwise coupled for all $1 \leq i < j \leq n$. We can combine (24) and (25) into one matrix differential equation

$$\frac{d}{dt} \begin{bmatrix} \Delta_{ij} \\ \Delta_{ji} \end{bmatrix} \quad = \quad - \begin{bmatrix} \beta_i \kappa_i & 0 \\ -\beta_j \kappa_i & \beta_j \kappa_j \end{bmatrix} \begin{bmatrix} \Delta_{ij} \\ \Delta_{ji} \end{bmatrix}, \tag{27}$$

where we have defined the quantity

$$\kappa_i \quad = \quad E\{y_i f_i(y_i)\} - E\{f_i'(y_i)\}. \tag{28}$$

Clearly, the transition matrix on the RHS of (27) has negative eigenvalues when (18) is satisfied. In such cases, the off-diagonal elements of Δ are all locally stable about zero, and this condition corresponds to separated and decorrelated output signals, verifying Theorem 1.1. Moreover, we see from (26) that the diagonal entries of Δ are marginally-stable for scale-only perturbations of the system outputs. Thus, Theorem 1.2 is proven □.

Remark: The stability condition in (18) is different from that for the separation schemes described in [7, 14, 15]. In particular, the methods described in [14, 15] guarantee local stability for separation if and only if

$$\beta\,[\kappa_i + \kappa_j] \quad > \quad 0 \tag{29}$$

for identical learning parameters $\beta = \beta_i$. While (29) is less stringent than (18) for identical β_i in the former case, we shall show that BSS methods based on the scheme in (12) can be designed to provide separation of arbitrary mixtures, in which (29) does not hold for a single fixed $\beta = \beta_i$.

3.2. Self-Stabilized Ordered-Rotation Methods

Although Theorem 1.1 indicates that the update in (12) is a potentially-useful BSS procedure, the inability of this algorithm to preserve the unit variances of the output signals as indicated by Theorem 1.2 poses numerical difficulties. Simulations show that the lengths of the rows of $\mathbf{W}(k)$ tend to grow without bound when (12) is used, causing eventual failure of the procedure. Fortunately, slight modifications of the updates can yield *self-stabilizing* algorithms that implicitly maintain the orthonormality of $\mathbf{W}(k)$ over time. Similar procedures have been used for other marginally-stable prewhitened BSS procedures [14, 15] as well as for principal and minor subspace analysis methods [16].

The first self-stabilized method is a minor generalization of the ordered-rotation KuicNet procedure for extracting positive-kurtosis sources in [6] and is

$$\begin{aligned} \mathbf{W}(k+1) \quad = \quad & \mathbf{W}(k) + \mathbf{D}_\beta \big\{ \mathbf{f}(\mathbf{y}(k))\mathbf{v}^T(k) \\ & - \mathrm{tri}[\mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k))]\mathbf{W}(k) \big\}. \end{aligned} \tag{30}$$

The following theorem describes the types of mixtures for which this algorithm is a valid BSS procedure, in which its corresponding ODE is

$$\frac{d\mathbf{W}}{dt} = \mathbf{D}_\beta \big\{ E\{\mathbf{f}(\mathbf{y})\mathbf{v}^T\} - \mathrm{tri}[E\{\mathbf{y}\mathbf{f}^T(\mathbf{y})\}]\mathbf{W} \big\}. \tag{31}$$

Theorem 2: *Stability of (31) About a Separating Solution.* For any small perturbations about $\mathbf{W} = \Phi\mathbf{J}\Gamma^T$, (31) is locally-stable in the prewhitened BSS task for values of β_i satisfying for all $1 \leq i \leq n$ and $i < j$ the conditions

$$\beta_i \kappa_i \quad > \quad 0 \tag{32}$$

$$\beta_j \left(\kappa_j + \kappa_i + E\{f_i'(y_i)\}] \right) \quad > \quad 0 \tag{33}$$

$$\beta_i E\{y_i f_i(y_i)\} \quad > \quad 0 \tag{34}$$

Proof: The structure of the proof largely parallels that for Theorem 1. By following a similar procedure, we determine an evolutionary equation for the perturbation matrix Δ in (21) from (31) as

$$\frac{d\Delta}{dt} = \mathbf{D}_\beta \big[E\{\mathbf{F}'(\mathbf{y})\Delta\mathbf{y}\mathbf{y}^T\} - \text{tri}[\Delta\Lambda(\mathbf{y}) + E\{\mathbf{y}\mathbf{y}^T\Delta^T\mathbf{F}'(\mathbf{y})\}] - \Lambda(\mathbf{y})\Delta\big] \tag{35}$$

The evolutions of the components Δ_{ij} and Δ_{ji} for $i < j$ are again pairwise coupled. The approximate evolutionary behavior of Δ_{ij} is described by (24), whereas

$$\frac{d\Delta_{ji}}{dt} = -\beta_j\big[E\{y_j f_(y_j)\} - E\{y_i^2 f_i'(y_j)\} + E\{y_i f_i(y_i)\}\big]\Delta_{ji} - \beta_j E\{y_j^2 f_i'(y_i)\}\Delta_{ij} \tag{36}$$

$$\frac{d\Delta_{ii}}{dt} = -2\beta_i E\{y_i f_i(y_i)\}\Delta_{ii} \tag{37}$$

describe the approximate evolutions of Δ_{ji} and Δ_{ii}.

Combining (24) and (36), one obtains the evolutionary equation for the pair $\{\Delta_{ij}, \Delta_{ji}\}$ as

$$\frac{d}{dt}\begin{bmatrix} \Delta_{ij} \\ \Delta_{ji} \end{bmatrix} = -\mathbf{H}_{ij}^{(p)} \begin{bmatrix} \Delta_{ij} \\ \Delta_{ji} \end{bmatrix} \tag{38}$$

$$\mathbf{H}_{ij}^{(p)} = \begin{bmatrix} \beta_i\kappa_i & 0 \\ \beta_j E\{f_i'(y_i)\} & \beta_j\left(\kappa_j + \kappa_i + E\{f_i'(y_i)\}\right) \end{bmatrix}, \tag{39}$$

where κ_i is as defined in (28). Because $\mathbf{H}_{ij}^{(p)}$ is lower-triangular, the eigenvalues of $\mathbf{H}_{ij}^{(p)}$ are simply its diagonal elements, and requiring these elements to be positive produces the conditions in (32) and (33). Finally, we see that (37) is locally-convergent if the term premultiplying Δ_{ii} on the RHS of this equation is negative, which produces the condition in (34). These results prove Theorem 2 □.

The second self-stabilized method is a significant modification of the ordered-rotation KuicNet procedure for extracting negative-kurtosis sources in [6] and is given by

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mathbf{D}_\beta\big\{\text{tri}[\mathbf{W}(k)\mathbf{W}^T(k)]\times \text{tri}[\mathbf{W}(k)\mathbf{W}^T(k)]\mathbf{f}(\mathbf{y}(k))\mathbf{v}^T(k) - \text{tri}[\mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k))]\mathbf{W}(k)\big\} \tag{40}$$

The following theorem describes the types of mixtures for which this algorithm is a valid BSS procedure, in which its corresponding ODE is

$$\frac{d\mathbf{W}}{dt} = \mathbf{D}_\beta\big\{\text{tri}[\mathbf{W}\mathbf{W}^T]\text{tri}[\mathbf{W}\mathbf{W}^T]E\{\mathbf{f}(\mathbf{y})\mathbf{v}^T\} - \text{tri}[E\{\mathbf{y}\mathbf{f}^T(\mathbf{y})\}]\mathbf{W}\big\} \tag{41}$$

Theorem 3: *Stability of (41) About a Separating Solution.* For any small perturbation about $\mathbf{W} = \Phi\mathbf{J}\Gamma^T$, (41) is locally-stable in the prewhitened BSS task for values of β_i satisfying for all $1 \le i \le n$ and $i < j$ the conditions

$$\beta_i\kappa_i > 0 \tag{42}$$

$$\beta_j\left(\kappa_j - \kappa_i - E\{f_i'(y_i)\}]\right) > 0 \tag{43}$$

$$\beta_i E\{y_i f_i(y_i)\} < 0 \tag{44}$$

Proof: The proof of this theorem is similar to that of Theorem 2. The evolutionary equation for Δ derived from (41) yields a pairwise joint update for $\{\Delta_{ij}, \Delta_{ji}\}$, $i < j$ as

$$\frac{d}{dt}\begin{bmatrix} \Delta_{ij} \\ \Delta_{ji} \end{bmatrix} = -\mathbf{H}_{ij}^{(m)} \begin{bmatrix} \Delta_{ij} \\ \Delta_{ji} \end{bmatrix} \tag{45}$$

$$\mathbf{H}_{ij}^{(m)} = \begin{bmatrix} \beta_i\kappa_i & 0 \\ \begin{Bmatrix} -\beta_j(-2\kappa_i \\ -E\{f_i'(y_i)\}) \end{Bmatrix} & \beta_j\left(\kappa_j - \kappa_i - E\{f_i'(y_i)\}\right) \end{bmatrix} \tag{46}$$

Requiring both diagonal elements of $\mathbf{H}_{ij}^{(m)}$ to be positive produces the conditions in (42) and (43). Finally, the evolutionary behavior of Δ_{ii} obeys

$$\frac{d\Delta_{ii}}{dt} = 2\beta_i E\{y_i f_i(y_i)\}\Delta_{ii}. \tag{47}$$

We require (44) for this equation to converge. Taken together, these results prove the theorem □.

4. COMBINED BSS FOR ARBITRARY MIXTURES

The analytical results of the previous section shall now be used to develop algorithms that blindly extract signals from arbitrary mixtures. These algorithms are iterative procedures that require the following knowledge about the source signals within the mixtures:

- the source vector $\mathbf{s}(k)$ contains m negative-kurtosis and p positive-kurtosis sources, where $m + p = n$, and
- the values of m and p are known.

Extensions of these methods to the case of unknown numbers of positive- and negative-kurtosis source mixtures are the subject of current work.

4.1. Self-Stabilized Algorithm for Prewhitened BSS

Consider the ordered-rotation BSS procedure in (12) for the common nonlinearity choice $f_i(y) = y^3$ for all $1 \le i \le n$. The results of Theorem 1.1 suggest that one should choose

$$\text{sgn}[\beta_i] = \text{sgn}[\kappa_i] \tag{48}$$

for this algorithm, where $\kappa_i = E\{|y_i(k)|^4\} - 3$ is the kurtosis of $y_i(k)$. In practice, one does not know the sign of the kurtosis of the output signal $y_i(k)$ before the system has converged. Taking the viewpoint espoused in [5], however, we need not be limited in this way. If the number of negative- and positive-kurtosis sources is known *a priori*, then we can

$$\text{choose } \beta_i < 0 \text{ if } i \le m \tag{49}$$

$$\text{choose } \beta_i > 0 \text{ if } i > m \tag{50}$$

in (12), and the algorithm will extract the negative-kurtosis sources in the first m system outputs and the p positive-kurtosis sources in the remaining system outputs. The stability conditions of Theorem 1.1 will be satisfied at a separating solution with these choices due to the unique extrema of the kurtosis contrast for prewhitened BSS [5].

Table 1: Contrast-Based BSS Algorithm for Prewhitened Mixtures

Equation	*MACs*
initialize: $\mathbf{W}(0)$ unitary	
for $k \geq 0$ do	
$\mathbf{z}_0(k) = \mathbf{0}$, $\mathbf{q}_0(k) = \mathbf{0}$	
for $i = 1$ to m do	
$y_i(k) = \mathbf{w}_i^T(k)\mathbf{v}(k)$	n^2
$\epsilon_i(k) = \beta_i y_i^3(k)$	$3n$
$\mathbf{z}_i(k) = \mathbf{z}_{i-1}(k) + \epsilon_i(k)\mathbf{w}_i(k)$	n^2
if $i \leq m$	
$\psi_i(k) = \mathbf{w}_i^T(k)\mathbf{z}_i(k)$	mn
$\mathbf{q}_i(k) = \mathbf{q}_{i-1}(k) + \psi_i(k)\mathbf{w}_i(k)$	mn
$\zeta_i(k) = \mathbf{w}_i^T(k)\mathbf{q}_i(k)$	mn
$\mathbf{w}_i(k+1) = \mathbf{w}_i(k) + \zeta_i(k)\mathbf{v}(k) - y_i(k)\mathbf{z}_i(k)$	$2mn$
else	
$\mathbf{w}_i(k+1) = \mathbf{w}_i(k) + \epsilon_i(k)\mathbf{v}(k) - y_i(k)\mathbf{z}_i(k)$	$2pn$
end	
end	
Total: $(4n + 3m + 3)n$	

In practice, the update in (12) does not maintain the scales of the extracted outputs, and for this reason, we consider an alternative self-stabilized method that combines the updates in (30) and (40) in a clever way. This combined update is

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mathbf{D}_\beta \left\{\mathbf{G}(k)\mathbf{f}(\mathbf{y}(k))\mathbf{v}^T(k) - \text{tri}[\mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k))]\mathbf{W}(k)\right\} \quad (51)$$

$$\mathbf{G}(k) = \begin{bmatrix} \text{tri}_m[\mathbf{W}(k)\mathbf{W}^T(k)]\text{tri}_m[\mathbf{W}(k)\mathbf{W}^T(k)] & \mathbf{0} \\ \mathbf{0}^T & \mathbf{I} \end{bmatrix}, \quad (52)$$

where $\text{tri}_m[\mathbf{W}(k)\mathbf{W}^T(k)]$ is an $(m \times m)$ matrix containing the first m rows and columns of $\text{tri}[\mathbf{W}(k)\mathbf{W}^T(k)]$ and $f_i(y) = y^3$ for all i. Selection of each β_i for this combined update follows (49)–(50). The next theorem extends the local stability results of Theorems 2 and 3 to (51) for which the associated ODE is omitted for brevity.

Theorem 4: *Stability of the Averaged ODE of (51) About a Separating Solution.*
For any small perturbation about $\mathbf{W} = \Phi\mathbf{J}\Gamma^T$, the averaged ODE of (51) is locally-stable in the prewhitened BSS task when extracting mixtures of m negative- and p positive-kurtosis sources for the step size selection in (49)–(50).

Proof: Combining the local stability conditions of Theorems 2 and 3 and noting that $E\{f'(y_i)\} = 3E\{y_i^2\} = 3$ results in the common set of conditions

$$\beta_i \kappa_i > 0 \quad (53)$$

$$\beta_j\left(\kappa_j + \text{sgn}(\kappa_j)[\kappa_i + 3]\right) > 0. \quad (54)$$

Since $\kappa_i > -3$ is always true, both (53) and (54) are satisfied if each β_i satisfies (48). The correspondence between (48) and (49)–(50) then follows from the extrema properties of the kurtosis contrast in [5] □.

Remark: Although it might appear that (51) requires $\mathcal{O}(n^3)$ operations to implement, we can use order-recursive calculations to implement the lower-triangular-matrix-vector multiplies. Table 1 lists a row-by-row implementation of the algorithm, in which the number of multiply/accumulates (MACs) is found to be of $\mathcal{O}(n^2)$.

Table 2: Equivariant BSS Algorithm for Arbitrary Mixtures

Equation	*MACs*
initialize: $\mathbf{B}(0)$ nonsingular	
for $k \geq 0$ do	
$\mathbf{z}_0(k) = \mathbf{0}$	
$\mathbf{y}(k) = \mathbf{B}(k)\mathbf{x}(k)$	n^2
$\mathbf{u}(k) = \mathbf{B}^T(k)\mathbf{y}(k)$	n^2
for $i = 1$ to n do	
$\epsilon_i(k) = \beta_i y_i^3(k)$	$3n$
$\delta_i(k) = \epsilon_i(k) - \mu y_i(k)$	n
$\mathbf{z}_i(k) = \mathbf{z}_{i-1}(k) + \epsilon_i(k)\mathbf{w}_i(k)$	n^2
$\mathbf{b}_i(k+1) = (1+\mu)\mathbf{b}_i(k)$ $+ \delta_i(k)\mathbf{u}(k) - y_i(k)\mathbf{z}_i(k)$	$3n^2$
end	
Total: $6n^2 + 4n$	

4.2. An Equivariant Algorithm for Combined Prewhitening and BSS

The algorithm in (51) can be easily combined with a well-known natural gradient prewhitening procedure [7, 17]. In this algorithm, $\mathbf{y}(k)$ is computed as

$$\mathbf{y}(k) = \mathbf{B}(k)\mathbf{x}(k), \quad (55)$$

where the $(n \times n)$ separation matrix $\mathbf{B}(k) = [\mathbf{b}_1(k) \cdots \mathbf{b}_n(k)]^T$ is adapted to satisfy (4)–(5) at convergence. Due to space limitations, we omit the full derivation of this approach and give only the final form of the algorithm:

$$\mathbf{B}(k+1) = (1+\mu)\mathbf{B}(k) + [\mathbf{D}_\beta\mathbf{f}(\mathbf{y}(k)) - \mu\mathbf{y}(k)]\,\mathbf{u}^T(k) - \mathbf{D}_\beta\text{tri}[\mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k))]\mathbf{B}(k) \quad (56)$$

$$\mathbf{u}(k) = \mathbf{B}^T(k)\mathbf{y}(k), \quad (57)$$

where $f_i(y) = y^3$ and the diagonal entries of $\mathbf{D}_\beta$ satisfy (49)–(50). As is the case for (51), the update in (56) can be implemented using $\mathcal{O}(n^2)$ MACs; Table 2 lists this order-recursive implementation.

Remark: The update in (56) resembles the EASI algorithm of Cardoso and Laheld [7], except that the outer product $\mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k))$ in the EASI update has been replaced with only its lower-triangular portion. The stability properties of our proposed algorithm, however, are fundamentally different from those of EASI. Unlike the approach specified in [7] where scaling of the individual nonlinearities $f_i(y)$ in $\mathbf{f}(\mathbf{y}(k))$ is suggested for stability, we instead select common step sizes β_i *across the rows* of the matrix $\mathbf{B}(k)$ in the contrast portion of the update. Such a choice is directly motivated from the previous analyses of the orthogonalized contrast approaches, and simulations verify the robustness of the proposed method.

5. SIMULATIONS

We now explore the proposed BSS methods via simulations. In all cases, we have mixed four unit-variance sources–two

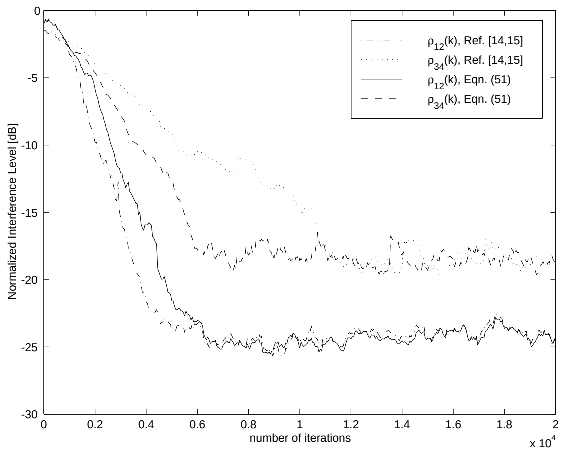

Fig. 2: Evolutions of $\rho_{ij}(k)$ for the orthogonal contrast BSS methods in the first simulation example.

uniform and two Laplacian–with the mixing matrix

$$\mathbf{A} = \begin{bmatrix} 0.7 & 0.2 & 0.3 & 0.6 \\ 0.2 & 0.8 & 0.5 & 0.4 \\ 0.5 & 0.3 & 0.1 & 0.9 \\ 0.6 & 0.4 & 0.7 & 0.8 \end{bmatrix}. \tag{58}$$

For the prewhitened methods, we have set $\mathbf{P} = \mathbf{R}^{-1}$, where $\mathbf{QR} = \mathbf{A}$ is the QR decomposition, such that $\Gamma = \mathbf{Q}$. Random initial matrices $\mathbf{W}(0)$ satisfying $\mathbf{W}(0)\mathbf{W}^T(0) = \mathbf{I}$ were employed in each case, and the performance factors

$$\rho_{ij}(k) = \sum_{l=i}^{j} \left[\frac{\sum_{t=1}^{n} c_{lt}^2(k)}{\max_{1\leq t\leq n} c_{lt}^2(k)} - 1 \right] \tag{59}$$

for $\{i = 1, j = 2\}$ and $\{i = 3, j = 4\}$ were computed for each algorithm using averages of 50 simulation runs.

Fig. 2 shows the evolutions of the performance factors for (51) as well as the positive- and negative-kurtosis source extraction methods in [14, 15]. Here, the values of $\beta = -0.001$ and $\beta = 0.00033$ for the positive- and negative-kurtosis algorithms in [14, 15] produced the same steady-state performance as (51) with $\beta_i = -0.0008$ for $i \leq m$ and $\beta_i = 0.0004$ for $i > m$. As can be seen, the proposed algorithm's performance in extracting negative-kurtosis sources is nearly as good as that of the existing method. The performance improvement of (51) over the existing method in extracting positive-kurtosis sources is significant, however, due to the former's use of a combined update.

Fig. 3 shows the evolutions of the performance factors for the equivariant prewhitening/BSS algorithm in (56), where $\mu = 0.002$, $\beta_i = -0.002$ for $i \leq m$, and $\beta_i = 0.0003$ for $i > m$. Clearly, the BSS algorithm extracts both positive- and negative-kurtosis sources without knowing or estimating any of their statistical characteristics.

6. CONCLUSIONS

In this paper, we have presented and analyzed novel algorithms for separating mixtures of arbitrary sources. The algorithms are simple and only require knowledge of the number of positive- and negative-kurtosis sources within the mixtures to function. Simulations verify their robust extraction abilities. Extensions to multichannel blind deconvolution are currently being developed.

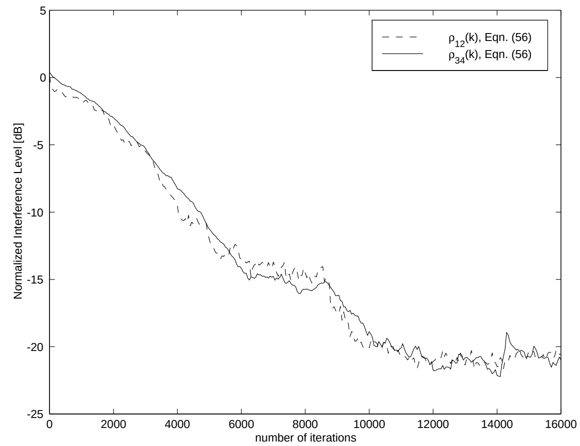

Fig. 3: Evolutions of $\rho_{ij}(k)$ for the equivariant prewhitening/BSS method in the second simulation example.

7. REFERENCES

[1] A.J. Bell and T.J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Comput.,* vol. 7, no. 6, pp. 1129-1159, Nov. 1995.

[2] S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation," *Adv. Neural Inform. Proc. Sys. 8* (Cambridge, MA: MIT Press, 1996), pp. 757-763.

[3] D.-T. Pham, "Blind separation of instantaneous mixture of sources via an independent component analysis," *IEEE Trans. Signal Processing,* vol. 44, pp. 2768-2779, Nov. 1996.

[4] S.C. Douglas, A. Cichocki, and S. Amari, "Multichannel blind separation and deconvolution of sources with arbitrary distributions," *Proc. IEEE Workshop Neural Networks Signal Processing,* Amelia Island, FL, pp. 436-445, Sept. 1997.

[5] S.-Y. Kung, "Independent component analysis in hybrid mixture: KuicNet learning algorithm and numerical analysis," *Proc. Int. Symp. Multimedia Inform. Processing,* Taipei, Taiwan, pp. 368-381, Dec. 1997.

[6] S.C. Douglas and S.-Y. Kung, "An ordered-rotation KuicNet algorithm for separating arbitrarily-distributed sources," *Proc. IEEE Int. Workshop Indep. Compon. Anal. Signal Separation,* Aussois, France, pp. 81-86, Jan. 1999

[7] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing,* vol. 44, pp. 3017-3030, Dec. 1996.

[8] K.I. Diamantaras and S.-Y. Kung, *Principal Component Neural Networks: Theory and Applications* (New York: Wiley, 1996).

[9] S.C. Douglas and S. Haykin, "Relationships between blind deconvolution and blind source separation," in *Unsupervised Adaptive Filtering, Vol. II: Blind Deconvolution,* S. Haykin, ed. (New York: Wiley, in press), Chap. 3.

[10] P. Comon, "Independent component analysis: A new concept?" *Signal Processing,* vol. 36, no. 3, pp. 287-314, Apr. 1994.

[11] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: a deflation approach," *Signal Processing,* vol. 45, no. 1, pp. 59-83, July 1995.

[12] A. Hyvarinen and E. Oja, "Independent component analysis by general nonlinear Hebbian-like learning rules," *Signal Processing,* vol. 64, no. 3, pp. 301-313, Feb. 1998.

[13] S.-Y. Kung, "Independent component analysis: extrema-correspondence properties for higher-order moment functions," *Proc. IEEE Workshop Neural Networks Signal Processing,* Cambridge, UK, pp. 53-62, Aug. 1998.

[14] S.C. Douglas, "Prewhitened blind source separation with orthogonality constraints," *Proc. 33rd Asilomar Conf. Signals, Syst. Comput.,* Pacific Grove, CA, vol. 2, pp. 1027-1031, Oct. 1999.

[15] S.C. Douglas, "Self-stabilized gradient algorithms for blind source separation with orthogonality constraints," *IEEE Trans. Neural Networks* (in press).

[16] S.C. Douglas, S.-Y. Kung, and S. Amari, "A self-stabilized minor subspace rule," *IEEE Signal Processing Lett.,* vol. 5, pp. 328-330, Dec. 1998.

[17] S.C. Douglas and A. Cichocki, "Neural networks for blind decorrelation of signals," *IEEE Trans. Signal Processing,* vol. 45, pp. 2829-2842, Nov. 1997.