INVITED PAPER *Special Section on Special Issue on Blind Signal Processing*

Blind Source Separation Algorithms with Matrix Constraints

Andrzej CICHOCKI[†,††], ***Member and*** **Pando GEORGIEV**[†,†††], ***Nonmember***

SUMMARY In many applications of Independent Component Analysis (ICA) and Blind Source Separation (BSS) estimated sources signals and the mixing or separating matrices have some special structure or some constraints are imposed for the matrices such as symmetries, orthogonality, non-negativity, sparseness and specified invariant norm of the separating matrix. In this paper we present several algorithms and overview some known transformations which allows us to preserve several important constraints. Computer simulation experiments confirmed validity and usefulness of the developed algorithms.

key words: *Blind sources separation, independent component analysis with constraints, non-negative blind source separation*

1. Introduction

The problem of blind source separation (BSS) and Independent Component Analysis (ICA) has received wide attention in various fields such as biomedical signal analysis and processing (EEG, MEG, fMRI), speech enhancement, geophysical data processing, data mining, wireless communications and image processing [1-33].

The mixing and filtering processes of the unknown input sources $s_j(k)$ $(j = 1, 2, ..., n)$ may have different mathematical or physical models, depending on the specific applications. In this paper, we will focus mainly on the simplest cases when m observed mixed signals $x_i(k)$ are linear combinations of n $(m \geq n)$ unknown, typically zero mean source signals $s_j(k)$ which are either statistically independent and/or they have different temporal structures [1], [21]. They are written as

$$x_i(k) = \sum_{j=1}^{n} h_{ij}\, s_j(k) + \nu_i(k), \quad (i = 1, 2, \ldots, m) \qquad (1)$$

or in the matrix notation

$$\mathbf{x}(k) = \mathbf{H}\,\mathbf{s}(k) + \boldsymbol{\nu}(k), \qquad (2)$$

where $\mathbf{x}(k) = [x_1(k), \ldots, x_m(k)]^T$ is a sensor vector, $\mathbf{s}(k) = [s_1(k), \ldots, s_n(k)]^T$ is a vector of source signals assumed to be statistically independent, $\boldsymbol{\nu}(k)$ is a vector of additive noise assumed to be independent with source signals, and $\mathbf{H}$ is an unknown full column rank $m \times n$ mixing matrix. It is assumed that only the sensor vector $\mathbf{x}(k)$ is available to use and it is desired to develop algorithms that enable estimation of primary sources and/or identification of the mixing matrix $\mathbf{H}$. The ICA of the sensor vector $\mathbf{x}(k) \in \mathbb{R}^m$ is obtained by finding an $n \times m$, (with $m \geq n$), a full rank separating matrix $\mathbf{W}$ such that the output signal vector $\mathbf{y}(k) = [y_1(k), y_2(k), \ldots, y_n(k)]^T$ defined by

$$\mathbf{y}(k) = \mathbf{W}\,\mathbf{x}(k), \qquad (3)$$

contains the estimated source components $\mathbf{s}(k) \in \mathbb{R}^n$ that are as independent as possible, evaluated by an information-theoretic cost function such as the minimum Mutual Information (MI).

The main objective of this contribution is to present several modifications and extensions of existing learning algorithms for ICA/BSS when some constraints are imposed for mixing matrix $\mathbf{H}$ and/or separating matrix $\mathbf{W}$ and estimated source signals. Typical constraints such as orthogonality or semi-orthogonality, symmetry and non-negativity constraints arise naturally in some physical models and they should be satisfied to obtain valid and reliable solutions [1], [6], [21], [25], [26], [28]–[30].

2. Blind Source Separation for Symmetric Matrices

Blind separation algorithms can be dramatically simplified if the mixing matrix $\mathbf{H}$ is symmetric and nonsingular. In such a case we can formulate the following.

Theorem 1: If in the model (2) the mixing matrix $\mathbf{H} \in \mathbb{R}^{n\times n}$ is nonsingular and symmetric, additive noise is zero or negligibly small and the zero-mean sources $\mathbf{s}(k)$ are spatially uncorrelated with unit variance, i.e. $\mathbf{R_{ss}} = E\{\mathbf{s}(k)\mathbf{s}^T(k)\} = \mathbf{I}_n$, then the demixing matrix can be uniquely estimated by eigenvalue decomposition of the covariance matrix $\mathbf{R_{xx}} = E\{\mathbf{x}(k)\mathbf{x}^T(k)\} = \mathbf{V_x}\boldsymbol{\Lambda}_\mathbf{x}\mathbf{V}_\mathbf{x}^T$ as

$$\mathbf{W}_* = \widehat{\mathbf{H}}^{-1} = \mathbf{R}_\mathbf{xx}^{-1/2} = \mathbf{V_x}\boldsymbol{\Lambda}_\mathbf{x}^{-1/2}\mathbf{V}_\mathbf{x}^T. \qquad (4)$$

Proof. The covariance matrix of sensor signals can be evaluated as

$$\mathbf{R_{xx}} = \mathbf{H}\mathbf{R_{ss}}\mathbf{H}^T. \qquad (5)$$

Taking into account that $\mathbf{R_{ss}} = \mathbf{I}_n$ and $\mathbf{H} = \mathbf{H}^T$, we

[†]The authors are with the Lab. for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Wako-shi, Saitama 351-0198, Japan.

[††]On leave from the Warsaw University of Technology, Poland

[†††]On leave from the Sofia University "St. Kl. Ohridski", Bulgaria

have $\mathbf{R_{xx}} = \mathbf{HH} = \mathbf{H}^2$. Hence, we obtain $\mathbf{W}_* = \widehat{\mathbf{H}}^{-1} = \mathbf{R_{xx}}^{-1/2} = \mathbf{V_x}\mathbf{\Lambda_x}^{-1/2}\mathbf{V_x}^T$, where $\mathbf{V_x}$ is the orthogonal matrix and $\mathbf{\Lambda_x}$ is the diagonal matrix obtained by the eigenvalue decomposition of the covariance matrix: $\mathbf{R_{xx}} = \mathbf{V_x}\mathbf{\Lambda_x}\mathbf{V_x}^T$. This means that the output signals $y_i(k)$ will be mutually orthogonal with unit variances. ■

In general, spatial decorrelation is not sufficient to perform instantaneous blind source separation from linear mixtures.

3. Independent Component Analysis with stable Frobenius Norm of the Demixing Matrix

In order to ensure the convergence of some learning algorithms and to provide their practical hardware implementations it is often necessary to restrict the values of entries of the separating matrices to a bounded subset. Such a bound can be, for example, imposed by a gradient descent learning system by keeping the Frobenius norm of the separating matrix bounded or fixed (invariant) during the learning process. In this section, we present some theorems which explain how to solve this problem. In the sequel we denote by $\mathrm{tr}(\mathbf{W})$ the trace of a matrix $\mathbf{W}$ and assume that a gradient $\dfrac{\partial J(\mathbf{W})}{\partial \mathbf{W}}$ of the cost function $J(\mathbf{W})$ satisfies reasonable conditions which guarantee uniqueness of the trajectories (for given initial condition) of the dynamical systems described below. For example, such a sufficient condition is $\dfrac{\partial J(\mathbf{W})}{\partial \mathbf{W}}$ to be a Lipschitz function on any bounded set of $\mathbb{R}^{n\times m}$.

Wide class of algorithms for ICA can be expressed in general form as [1]

$$\frac{d\,\mathbf{W}(t)}{d\,t} = \mu(t)\,\mathbf{F}(\mathbf{y}(t))\,\mathbf{W}(t), \qquad (6)$$

where $\mathbf{y}(t) = \mathbf{W}(t)\mathbf{x}(t)$ and the matrix $\mathbf{F}(\mathbf{y})$ can take different forms, for example $\mathbf{F}(\mathbf{y}) = \mathbf{I}_n - \mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})$ with suitably chosen nonlinearities $\mathbf{f}(\mathbf{y}) = [f(y_1, ..., f(y_n)]$ and $\mathbf{g}(\mathbf{y}) = [g(y_1, ..., g(y_n)]$ [11], [12].

Alternatively, for signals corrupted by additive Gaussian noise, we can use higher order matrix cumulants as follows [1], [13], [14]. Let us use the following notation: $C_q(y_1)$ denotes the q-order cumulants of the signal y_i and $\mathbf{C}_{p,q}(\mathbf{y},\mathbf{y})$ denotes the cross-cumulant matrix whose elements are $[\mathbf{C}_{pq}(\mathbf{y},\mathbf{y})]_{ij} = Cum(\underbrace{y_i, y_i, \ldots, y_i}_{p}, \underbrace{y_j, y_j, \ldots, y_j}_{q})$.

Let us consider the following cost function:

$$J(\mathbf{y},\mathbf{W}) = -\frac{1}{2}\log|\det(\mathbf{W}\mathbf{W}^T)| - \frac{1}{1+q}\sum_{i=1}^{n}|C_{1+q}(y_i)|. \qquad (7)$$

The first term assures that the determinant of the global matrix will not approach zero. By including this term, we avoid the trivial solution $y_i = 0 \quad \forall i$. The second terms force the output signals to be as far as possible from Gaussianity, since the higher order cumulants are a natural measure of non-Gaussianity and they will vanish for Gaussian signals. It can be shown that for such a cost function, we can derive the following equivariant discrete time algorithm [14]

$$\begin{aligned}\Delta\mathbf{W}(k) &= \mathbf{W}(k+1) - \mathbf{W}(k)\\ &= \eta_k\,[\mathbf{I} - \mathbf{C}_{1,q}(\mathbf{y},\mathbf{y})\,\mathbf{S}_{q+1}(\mathbf{y})]\,\mathbf{W}(k),\end{aligned} \qquad (8)$$

where $\mathbf{S}_{q+1}(\mathbf{y}) = \mathrm{sign}(\mathrm{diag}(\mathbf{C}_{1,q}(\mathbf{y},\mathbf{y})))$ and $\mathbf{F}(\mathbf{y}) = \mathbf{I} - \mathbf{C}_{1,q}(\mathbf{y},\mathbf{y})\,\mathbf{S}_{q+1}(\mathbf{y})$. The above algorithm has no constraints imposed on entries of the separating matrix $\mathbf{W}$.

For some $\mathbf{F}$ the dynamical system doesn't correspond to minimization of any cost function, for example, this is the case of nonholonomic orthogonal learning algorithm [1], where, for specific $\mathbf{F}(\mathbf{y}) = \mathrm{diag}\{\mathbf{f}(\mathbf{y})\mathbf{y}^T\} - \mathbf{f}(\mathbf{y})\mathbf{y}^T$, the diagonal elements of $\mathbf{F}(\mathbf{y})$ are zero. The main observation for proving this fact is that for a given diagonal matrix $\mathbf{D}$ (different from the identity matrix) there is no such a cost function $\mathbf{J}(\mathbf{W})$ such that

$$\frac{\partial \mathbf{J}(\mathbf{W})}{\partial \mathbf{W}} = \mathbf{D}\mathbf{W}^{-T}.$$

This fact follows from the criterion for the existence of potential functions (see [23], Theorem 3.4).

For such a general case, the algorithm may diverge to infinity, or may converge to zero. We propose modifications which stabilize the Frobenius norm of $\mathbf{W}$ as follows.

Theorem 2: The learning rule

$$\frac{d\,\mathbf{W}(t)}{d\,t} = \mu(t)\;[\mathbf{F}(\mathbf{y}(t)) - \beta\gamma(t)\mathbf{I}_n]\;\mathbf{W}(t), \qquad (9)$$

where $\beta > 0$ is a scaling factor and $\gamma(t) = \mathrm{trace}\left(\mathbf{W}^T(t)\mathbf{F}(\mathbf{y}(t))\mathbf{W}(t)\right) > 0$, stabilizes the Frobenius norm of $\mathbf{W}(t)$ such $||\mathbf{W}(t)||_F^2 = \mathrm{tr}(\mathbf{W}^T(t)\mathbf{W}(t)) \approx \beta^{-1}$.

Proof. It is straightforward to show that

$$\frac{d\,\mathrm{tr}(\mathbf{W}^T\mathbf{W})}{dt} = -2\,\mu(t)\gamma(t)\beta\left[\mathrm{tr}(\mathbf{W}^T\mathbf{W}) - \beta^{-1})\right]. \qquad (10)$$

Denote $z(t) = ||\mathbf{W}||_F^2 = \mathrm{tr}\left(\mathbf{W}^T(t)\mathbf{W}(t)\right)$, and consider a differential equation of the form

$$\frac{dz}{dt} = -2\,\mu(t)\,\gamma(t)\beta\left(z(t) - \beta^{-1}\right).$$

The above differential equation has for an initial condition $z(0) = \beta^{-1} + \delta$ (where δ is a small coefficients representing perturbation) the following solution

$$z(t) = \beta^{-1} + \delta\exp(-2\int_0^t \mu(t)\,\gamma(t)\beta\,dt). \qquad (11)$$

Since $\mu(t)$, $\gamma(t)$ and β are assumed to be nonnegative, the exponential term in (11) decay to zero keeping the norm $||\mathbf{W}(t)||_F^2 = \mathrm{tr}(\mathbf{W}^T(t)\mathbf{W}(t))$ close to value β^{-1} and prevents the norm to explode. ∎

In the next corollaries from Theorem 2 we present stabilizing modifications of some known algorithms.

Corollary 1: The modified natural gradient descent learning algorithm with a forgetting factor described as

$$\frac{d\,\mathbf{W}(t)}{d\,t} = -\mu(t)\Big[\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}\,\mathbf{W}^T(t)\,\mathbf{W}(t) + \beta\gamma(t)\,\mathbf{W}(t)\Big], \tag{12}$$

where $J(\mathbf{W})$ is the cost function, $\nabla_{\mathbf{W}} J = \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}$ denotes its gradient with respect to the nonsingular matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$, $\mu > 0$ is the learning rate and

$$\gamma(t) = -\mathrm{tr}\left(\mathbf{W}^T(t)\,\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}\,\mathbf{W}^T(t)\,\mathbf{W}(t)\right) > 0 \tag{13}$$

is a forgetting factor, ensures that the Frobenius norm of the matrix $\mathbf{W}(t)$ is bounded.

Proof. Take $F(\mathbf{y}) = -\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}\,\mathbf{W}^T$ in Theorem 2. ∎

Remark 1 Theorem 2 does not discuss the stability problem of the corresponding learning algorithms, which depend on a cost function and a learning rate but only states that by introducing the forgetting factor we are able to bound entries of the matrix $\mathbf{W}$ in such a way that its Frobenius norm is bounded or even fixed.

The well known Amari's natural gradient algorithm [1] based on minimization of the MI can be modified as follows:

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \eta_k\left[\mathbf{I}_n - \langle \mathbf{f}(\mathbf{y}_k)\mathbf{y}_k^T\rangle - \beta\gamma_k\mathbf{I}_n\right]\mathbf{W}_k, \tag{14}$$

where $\gamma_k = \mathrm{tr}\left(\mathbf{W}_k^T\boldsymbol{\Gamma}_k\right)$, $\boldsymbol{\Gamma}_k = \left(\mathbf{I}_n - \langle \mathbf{f}(\mathbf{y}_k)\mathbf{y}_k^T\rangle\right)\mathbf{W}_k$ and β satisfies the condition in Theorem 2.

Corollary 2: The stochastic gradient descent learning algorithm

$$\frac{d\,\mathbf{W}(t)}{d\,t} = -\mu(t)\left[\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} + \beta\gamma(t)\mathbf{W}(t)\right], \tag{15}$$

where $J(\mathbf{W})$ is the cost function, $\nabla_{\mathbf{W}} J(\mathbf{W}) = \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}$ denotes its gradient with respect to the nonsingular matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$, $\mu(t) > 0$ is the learning rate and

$$\gamma(t) = -\mathrm{tr}\left(\mathbf{W}^T(t)\,\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}\right) > 0 \tag{16}$$

is a forgetting factor, ensures that the Frobenius norm of the matrix $\mathbf{W}(t)$ is stable.

Corollary 3: The modified Atick-Redlich descent learning algorithm with forgetting factor [1]

$$\frac{d\,\mathbf{W}(t)}{d\,t} = -\mu(t)\Big[\mathbf{W}(t)\left[\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}\right]^T\mathbf{W}(t) + \beta\gamma(t)\mathbf{W}(t)\Big], \tag{17}$$

where $J(\mathbf{W})$ is a differentiable cost function, $\nabla_{\mathbf{W}} J = \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}$ denotes its gradient with respect to the nonsingular matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$, $\mu > 0$ is the learning rate and

$$\gamma(t) = -\mathrm{tr}\left(\mathbf{W}^T(t)\,\mathbf{W}(t)\left[\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}\right]^T\mathbf{W}(t)\right) > 0$$

is a forgetting factor, ensures that the Frobenius norm of the matrix $\mathbf{W}(t)$ is stable.

4. ICA Algorithms with Semi-Orthogonality Constraints

Recently, an interest has grown in gradient algorithms that attempt to impose semi-orthogonality constraints on the separating matrix [1], [4], [7], [13], [15], [16]. One advantage of this approach is possibility to extract arbitrary group of sources. Consider the problem of extraction of arbitrary group of sources simultaneously, say e, where $1 \le e \le n$, with $m \ge n$. Let us also assume that the sensor signals are prewhitened, for example, by using the PCA technique [1]. Then, the transformed sensor signals satisfy the condition

$$\mathbf{R}_{\overline{\mathbf{x}}\,\overline{\mathbf{x}}} = E\{\overline{\mathbf{x}}\,\overline{\mathbf{x}}^T\} = \mathbf{I}_n, \tag{18}$$

where $\overline{\mathbf{x}} = \mathbf{x}_1 = \mathbf{Q}\,\mathbf{x}$ and the new global $n\times n$ mixing matrix $\mathbf{A} = \mathbf{Q}\,\mathbf{H}$ is orthogonal, that is, $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}_n$. Hence, the ideal $n\times n$ separating matrix is $\mathbf{W}_* = \mathbf{A}^{-1} = \mathbf{A}^T$ for $e = n$.

In order to solve this problem, we can formulate the appropriate cost function expressed by the Kullback-Leibler divergence

$$K_{pq} = K(p(\mathbf{y},\mathbf{W}_e)||q(\mathbf{y})) = \int p(\mathbf{y},\mathbf{W}_e)\log\frac{p(\mathbf{y},\mathbf{W}_e)}{q(\mathbf{y})}d\mathbf{y}, \tag{19}$$

where $q(\mathbf{y}) = \prod_{i=1}^{e} q_i(y_i)$ represents an adequate independent probability distribution of the output signals $\mathbf{y} = \mathbf{W}_e\overline{\mathbf{x}}$, and $\mathbf{W}_e \in \mathbb{R}^{e\times n}$ is a demixing (separating) matrix, with $e \le n$.

Hence, the cost function takes the form:

$$J(\mathbf{y},\mathbf{W}_e) = -\sum_{i=1}^{e} E\{\log q_i(y_i)\} \tag{20}$$

subject to constraints $\mathbf{W}_e\mathbf{W}_e^T = \mathbf{I}_e$.

These constraints follow from the simple fact that the mixing matrix $\mathbf{A} = \mathbf{Q}\,\mathbf{H}$ is a square orthogonal matrix and the demixing matrix $\mathbf{W}_e$ should satisfy the following relationship after successful extraction of e sources (ignoring scaling and permutation for simplicity):

$$\mathbf{W}_e\mathbf{A} = [\mathbf{I}_e, \mathbf{0}_{n-e}]. \tag{21}$$

We say that the matrix $\mathbf{W}_e$ satisfying the above condition forms a Stiefel manifold since its rows are mutually orthogonal ($\mathbf{w}_i\mathbf{w}_j^T = \delta_{ij}$). In order to satisfy the constraints during the learning process, we employ the following continuous-time natural gradient formula [15], [17]

$$\frac{d\,\mathbf{W}_e(t)}{d\,t} = -\mu(t)\Big(\frac{\partial J(\mathbf{y},\mathbf{W}_e)}{\partial \mathbf{W}_e} - \mathbf{W}_e(t)\left[\frac{\partial J(\mathbf{y},\mathbf{W}_e)}{\partial \mathbf{W}_e}\right]^T \mathbf{W}_e(t)\Big). \tag{22}$$

It can be shown that the separating matrix $\mathbf{W}_e$ satisfies the relation $\mathbf{W}_e(t)\mathbf{W}_e^T(t) = \mathbf{I}_e$ on the condition that $\mathbf{W}_e(0)\mathbf{W}_e^T(0) = \mathbf{I}_e$. This property can be formulated in the form of the following theorem.

Theorem 3: The natural gradient dynamic systems on the Stiefel manifolds described by (22) with $\mu > 0, \mathbf{W} \in \mathbb{R}^{e\times n},\ \ e \le n$ satisfy during the learning process the semi-orthogonality constraints:

$$\mathbf{W}_e(t)\mathbf{W}_e^T(t) = \mathbf{I}_e,\ \ \forall t \ \text{ if } \ \ \mathbf{W}_e(0)\mathbf{W}_e^T(0) = \mathbf{I}_e \tag{23}$$

Proof. Consider the following system of differential equations

$$\frac{d\,\mathbf{W}_e(t)}{d\,t} = -\mu(t)\Big(\mathbf{W}_e(t)\mathbf{W}_e^T(t)\frac{\partial J(\mathbf{W}_e)}{\partial \mathbf{W}_e} - \mathbf{W}_e(t)\left[\frac{\partial J(\mathbf{W}_e)}{\partial \mathbf{W}_e}\right]^T \mathbf{W}_e(t)\Big) \tag{24}$$

with the initial condition $\mathbf{W}_e(0)\mathbf{W}_e^T(0) = \mathbf{I}_e$ the same as those of (22). It is straightforward to check that the solution $\mathbf{W}_e(t)$ of (24) satisfies $\frac{d(\mathbf{W}_e\mathbf{W}_e^T)}{dt} = \mathbf{0}$. Therefore, $\mathbf{W}_e(t)\mathbf{W}_e^T(t) = \mathbf{I}_e$ for every $t \ge 0$. Hence, the systems (22) and (24) coincide (considered with the same initial condition satisfying (23)). ■

Remark 2 It should be noted that for discrete-time version of the algorithm

$$\Delta\mathbf{W}_e(k) = -\eta_k(\boldsymbol{\nabla}_k - \mathbf{W}_e(k)\boldsymbol{\nabla}_k^T\mathbf{W}_e(k)), \tag{25}$$

where $\boldsymbol{\nabla}_k = \frac{\partial J(\mathbf{W}_e(k))}{\partial \mathbf{W}_e}$ the semi-orthogonality constraints can be satisfied only approximately. Direct calculations show that

$$\mathbf{W}_e(k+1)\mathbf{W}_e^T(k+1) = \mathbf{W}_e(k)\mathbf{W}_e^T(k) + \eta_k^2\Big(\boldsymbol{\nabla}_k\boldsymbol{\nabla}_k^T - (\boldsymbol{\nabla}_k\mathbf{W}_e^T(k))^2 - (\mathbf{W}_e(k)\boldsymbol{\nabla}_k^T)^2 + \mathbf{W}_e(k)\boldsymbol{\nabla}_k^T\boldsymbol{\nabla}_k\mathbf{W}_e(k)^T\Big), \tag{26}$$

where the second term can be neglected if the learning step is sufficiently small. In order to satisfy the semi-orthogonality constraints we can employ modified algorithm

$$\Delta\mathbf{W}_e(k) = -\eta_k[\mathbf{W}_e(k)\mathbf{W}_e^T(k)\boldsymbol{\nabla}_k - \mathbf{W}_e(k)\boldsymbol{\nabla}_k^T\mathbf{W}_e(k). \tag{27}$$

Applying the natural gradient formula (25), we obtain a learning rule:

$$\mathbf{W}_e(k+1) = \mathbf{W}_e(k) - \eta(k)[\langle \mathbf{f}[\mathbf{y}(k)]\,\overline{\mathbf{x}}^T(k)\rangle - \langle \mathbf{y}(k)\,\mathbf{f}^T[\mathbf{y}(k)]\rangle\ \mathbf{W}_e(k)], \tag{28}$$

with the initial conditions satisfying $\mathbf{W}_e(0)\mathbf{W}_e^T(0) = \mathbf{I}_e$.

It is worth to note, that for $e = n$, the separating matrix $\mathbf{W}_e = \mathbf{W}$ is orthogonal and the learning rule simplifies to the well-known algorithm proposed by Cardoso and Laheld [7]:

$$\mathbf{W}(k+1) = \mathbf{W}(k) - \eta(k)\Big[\langle \mathbf{f}[\mathbf{y}(k)]\mathbf{y}^T(k)\rangle - \langle \mathbf{y}(k)\mathbf{f}^T[\mathbf{y}(k)]\rangle\Big]\mathbf{W}(k), \tag{29}$$

since $\overline{\mathbf{x}} = \mathbf{Q}\,\mathbf{x} = \mathbf{Q}\,\mathbf{H}\,\mathbf{s} = \mathbf{W}^{-1}\,\mathbf{y} = \mathbf{W}^T\,\mathbf{y}$.

The above algorithm (28) can be implemented as on-line (moving average) algorithm [1]:

$$\mathbf{W}_e(k+1) = \mathbf{W}_e(k) - \eta(k)\,[\mathbf{R}_{\mathbf{f}\,\overline{\mathbf{x}}}^{(k)} - \mathbf{R}_{\mathbf{y}\,\mathbf{f}}^{(k)}\,\mathbf{W}_e(k)], \tag{30}$$

where

$$\mathbf{R}_{\mathbf{f}\,\overline{\mathbf{x}}}^{(k)} = (1-\eta_0)\,\mathbf{R}_{\mathbf{f}\,\overline{\mathbf{x}}}^{(k-1)} + \eta_0\,\mathbf{f}(\mathbf{y}(k))\,\overline{\mathbf{x}}^T(k), \tag{31}$$

$$\mathbf{R}_{\mathbf{y}\,\mathbf{f}}^{(k)} = (1-\eta_0)\,\mathbf{R}_{\mathbf{y}\,\mathbf{f}}^{(k-1)} + \eta_0\,\mathbf{y}(k)\,\mathbf{f}^T(\mathbf{y}(k)), \tag{32}$$

and $\mathbf{f}(\mathbf{y}) = [f(y_1), f(y_2), \ldots, f(y_e)]^T$, with suitably chosen score functions ($f(y_i) = -d\,\log(q_i(y_i))/dy_i = q_i'(y_i)/q_i(y_i)$).

An interesting modification of the natural gradient learning algorithm with the orthogonality constraint has been proposed by Nishimori and Fiori [18], [27]. The algorithm can be represented as

$$\mathbf{W}(k+1) = \exp[-\eta_k\,\tilde{\boldsymbol{\nabla}}_{\mathbf{W}}J]\ \mathbf{W}(k), \tag{33}$$

where the gradient $\tilde{\boldsymbol{\nabla}}_{\mathbf{W}}J$ can take a special form:

$$\tilde{\boldsymbol{\nabla}}_{\mathbf{W}}J = \boldsymbol{\nabla}_k\mathbf{W}^T(k) - \mathbf{W}(k)\boldsymbol{\nabla}_k^T = \mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k) - \mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k)), \tag{34}$$

with the standard gradient denoted as $\boldsymbol{\nabla}_k = \frac{\partial J(\mathbf{W}(k))}{\partial \mathbf{W}}$. It can be easily shown that in the special case of Grassmann-Stiefel manifolds the learning rule (33) exactly satisfies semi-orthogonality constraints independent of the value of the learning rate η_k under the condition that $\mathbf{W}(0)\mathbf{W}^T(0) = \mathbf{I}$.

Proof. Let us assume that $\mathbf{W}(k)\mathbf{W}^T(k) = \mathbf{I}$. Then we have $\mathbf{W}(k+1)\mathbf{W}^T(k+1) = \exp[-\eta_k \tilde{\boldsymbol{\nabla}}_{\mathbf{W}} J]$ $\mathbf{W}(k)\mathbf{W}^T(k)\exp[\eta_k \tilde{\boldsymbol{\nabla}}_{\mathbf{W}} J] = \exp(\mathbf{0}) = \mathbf{I}$.

The above learning formula can be extended or generalized to the following forms (that not necessarily satisfy orthogonality constraints):

$$\mathbf{W}(k+1) = \exp[\eta_k \mathbf{F}(\mathbf{y}(k))]\ \mathbf{W}(k), \tag{35}$$

and

$$\widehat{\mathbf{H}}(k+1) = \widehat{\mathbf{H}}(k)\ \exp[-\eta_k \mathbf{F}(\mathbf{y}(k))], \tag{36}$$

where the matrix $\mathbf{F}(\mathbf{y}(k))$ can take various forms as discussed in previous sections of this paper (see also [1]). The matrix $\exp[\eta_k \mathbf{F}(\mathbf{y}(k))]$ can be efficiently computed in MATLAB using a Pade approximation.

5. Blind Source Separation with Non-negativity Constraints

In many applications such as computer tomography and biomedical image processing non-negative constraints are imposed for entries ($h_{ij} \geq 0$) of the mixing matrix $\mathbf{H}$ and/or estimated source signals ($s_j(k) \geq 0$) [1], [6], [20], [26], [28], [30]. Moreover, recently several authors suggested that a decomposition of a observation $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \ldots \mathbf{x}(N)] = \mathbf{HS}$ into non-negative factors or Non-Negative Matrix Factorization (NMF), is able to produce useful and meaningful representation of real-world data, especially in image analysis, hyperspectral data processing, biological modeling and sparse coding [20], [25], [28], [29].

In this section, we present very simple and practical technique for estimation of nonnegative sources and entries of the mixing matrix using standard ICA approach and suitable postprocessing. In other words, we will show that by simple modifications of existing ICA or BSS algorithms we are able to satisfy non-negativity constraints of sources and simultaneously impose they are sparse or independent as possible. Without loss of generality, we assume that matrix and all sources are non-negative, i.e., $s_j(k) = \widetilde{s}_j(k) + c_j \geq 0 \quad \forall j, k$. Moreover, we assume that the zero mean subcomponents $\widetilde{s}_j(k)$ are mutually statistically independent†.

Furthermore, we may assume if necessary that the entries of nonsingular mixing matrix $\mathbf{H}$ are also non-negative i.e, $h_{ij} \geq 0 \ \forall i, j$ and optionally that columns of the mixing matrix are normalized vectors with 1-norm equal to unity [6], [26], [28].

We propose a two stage procedure. In the first stage, we can apply any standard ICA or BSS algorithm for zero-mean (pre-processed) sensor signals without any constraints in order to estimate the separating matrix $\mathbf{W}$ up to an arbitrary scaling and permutation and estimate the waveforms of the original sources by projecting the (non-zero mean) sensor signals $s_j(k)$ via the estimated separating matrix ($\widehat{\mathbf{s}}(k) = \mathbf{W}_* \mathbf{x}(k)$).

It should be noted that since the global mixing-unmixing matrix defined as $\mathbf{G} = \mathbf{W}_* \mathbf{H}$ after successful extraction of the sources is a generalized permutation matrix containing only one nonzero (negative or positive) element in each row and each column, thus each estimated source in the first stage will be either non-negative or non-positive for every time instant.

In the second stage in order to recover the original waveforms of the sources with correct sign all the estimated non positive sources should be inverted, i.e. multiplied by -1. It should be noted that this procedure is valid for an arbitrary nonsingular mixing matrix with both positive and negative elements.

If the original mixing matrix $\mathbf{H}$ has non-negative entries, then in order to identify it the corresponding vectors of the estimating matrix $\widehat{\mathbf{H}} = \mathbf{W}^{-1}$ should be multiplied by the factor -1. In this way, we can estimate the original sources and blindly identify the mixing matrix satisfying non-negativity constraints. Furthermore, if necessary, we can redefine $\widehat{\mathbf{H}}$ and $\hat{\mathbf{s}}$ as follows: $\widehat{\hat{h}}_{kj} = \hat{h}_{kj} / \sum_{i=1}^n \hat{h}_{ij}$ and $\widehat{\hat{s}}_j = \hat{s}_j (\sum_{i=1}^n \hat{h}_{ij})$. After such transformation, the new estimated mixing matrix $\widehat{\widehat{\mathbf{H}}}$ has column sums equal to one and the vector $\mathbf{x} = \widehat{\widehat{\mathbf{H}}}\widehat{\widehat{\mathbf{s}}}$ is unchanged.

Summarizing, from this simple explanation it follows that it is not necessary to develop any special kind of algorithms for BSS with non-negativity constraints (see for example [25], [28], [30]). Any standard ICA algorithm (batch or on-line) can be applied first for zero-mean signals, and the waveforms of original sources and optionally the desired mixing matrix with non-negativity constraints can be estimated exploiting basic properties of the assumed model (see Example 2).

6. Multiresolution Subband Decomposition – Independent Component Analysis (MSD-ICA)

Despite the success of using standard ICA in many applications, the basic assumptions of ICA may not hold for some kind of signals hence some caution should be taken when using standard ICA to analyze real world problems, especially in biomedical signal processing. In fact, by definition, the standard ICA algorithms are not able to estimate statistically dependent original sources, that is, when the independence assumption is violated. In this section we will present some natural extensions and generalizations of ICA called Multiresolution Subband Decomposition ICA (MSD-ICA) which relaxes considerably the assumption regarding mutual

†It should be noted that the non negative sources $s_j(k) = \tilde{s}_j(k) + c_j$ are non independent, even zero mean sub-components $\widetilde{s}_j(k)$ are independent, since dc (constant) sub-components c_j are dependent. Due to this reason we refer to the problem as non-negative blind source separation rather than non-negative ICA [29], [30].

independence of primarily sources. The key idea in this approach is the assumption that the wide-band source signals are generally dependent, however some narrow band subcomponents of the sources are independent. In other words, we assume that each unknown source can be modeled or represented as a sum of narrow-band sub-signals (sub-components):

$$s_i(k) = s_{i1}(k) + s_{i2}(k) + \cdots + s_{iK}(k). \quad (37)$$

The basic concept in MSD-ICA is to divide the sensor signal spectra into their subspectra or subbands, and then to treat those subspectra individually for the purpose at hand. The subband signals can be ranked and processed independently. Let us assume that only a certain set of sub-components is independent. Provided that for some of the frequency subbands (at least one) all sub-components, say $\{s_{ij}(k)\}_{i=1}^{n}$, are mutually independent or temporally decorrelated, then we can easily estimate the mixing or separating system (under condition that these subbands can be identified by some *a priori* knowledge or detected by some self-adaptive process) by simply applying any standard ICA algorithm, however not for all available raw sensor data but only for suitably pre-processed (e.g., band pass filtered) sensor signals.

In one of the most simplest case, source signals can be modeled or decomposed into their low- and high-frequency sub-components:

$$s_i(k) = s_{iL}(k) + s_{iH}(k) \qquad (i = 1, 2, \ldots, n). \quad (38)$$

In practice, the high-frequency sub-components $s_{iH}(k)$ are often found to be mutually independent. In such a case in order to separate the original sources $s_i(k)$, we can use a High Pass Filter (HPF) to extract the mixture of independent high frequency sub-components and then apply any standard ICA algorithm to such preprocessed sensor (observed) signals. We have implemented these concepts in our ICALAB software and extensively tested these concepts for some experimental data [1], [8]. Such explanation can be summarized as follows.

Definition 1 (MSD-ICA): The MSD-ICA (Multiresolution Subband Decomposition ICA) can be formulated as a task of estimation of the separating matrix $\mathbf{W}$ and/or the mixing matrix $\mathbf{H}$ on the basis of suitable multiresolution subband decomposition of sensor signals and by applying a classical ICA (instead for raw sensor data) for one or several preselected subbands for which source sub-components are independent.

In the preprocessing stage, more sophisticated methods, such as block transforms, multirate subband filter bank or wavelet transforms, can be applied. We can extend and generalize further this concept by performing the decomposition of sensor signals in a composite time-frequency domain rather than in frequency subbands. This naturally leads to the concept of wavelets packets (subband hierarchical trees) and to block transform packets [9], [35].

7. Computer Simulation Experiments

In this section, we present a few examples illustrating the validity and performance of some of the proposed algorithms.

Example 1. Five independent zero-mean sources mixed by a randomly generated full column rank mixing matrix $\mathbf{H} \in \mathbb{R}^{9\times 5}$ and additive uniform noise was added with SNR 15dB. The plots of original sources and their noisy mixtures are shown respectively in Fig 1 (a) and (b).

In order to reconstruct original sources the learning rule (14) has been applied for square separating matrix $\mathbf{W} \in \mathbb{R}^{9\times 9}$ with $\mathbf{F}(y) = \mathbf{I}_5 - \langle \mathbf{f}(y)\mathbf{y}^T \rangle$, $\beta = 0.01$ and $\mathbf{f}(y) = \tanh(2\mathbf{y})$. After 100 iterations the algorithm was able to estimate the sources (see Fig. 1 (c)) by keeping the Frobenius norm of the separating matrix in the range $||\mathbf{W}(k)||_F^2 = \mathrm{tr}(\mathbf{W}^T(k)\mathbf{W}(k)) \approx 100$. Without a stabilizing factor the norm of $\mathbf{W}$ grows to very large value as the number of iteration increases.

Example 2. In this example sparse non-negative source signals are mixed by a normalized sparse matrix with non-negative entries (assumed to be unknown).

$$\mathbf{H} = \begin{bmatrix} 0 & 0.7 & 0.2 & 0 \\ 0.4 & 0 & 0 & 0.3 \\ 0.6 & 0.3 & 0 & 0.5 \\ 0 & 0 & 0.8 & 0.2 \end{bmatrix} \quad (39)$$

The mixed (observed sensor) signals are shown in Fig. 2 (a).

After standard prewhitening algorithm (8) with fourth order cumulants has been applied. The estimated sources are shown in Fig 2 (c) and the estimating mixing matrix has the form

$$\widehat{\mathbf{H}} = \mathbf{W}^{-1} = \begin{bmatrix} 0.000 & 3.549 & 0.000 & 0.888 \\ -1.861 & 0.000 & -1.245 & 0.000 \\ -2.783 & 1.521 & -2.095 & 0.001 \\ -0.005 & 0.003 & -0.909 & 3.563 \end{bmatrix}$$

It is seen that source $\hat{s}_1(k)$ and $\hat{s}_3(k)$ are non positive so they should be inverted and the corresponding columns (first and third one) of the matrix $\widehat{\mathbf{H}}$ should be multiplied by -1. After normalizing the mixing matrix we obtained

$$\widehat{\widehat{\mathbf{H}}} = \begin{bmatrix} 0.000 & 0.700 & 0.000 & 0.199 \\ 0.395 & 0.000 & 0.293 & 0.000 \\ 0.605 & 0.300 & 0.493 & 0.000 \\ 0.000 & 0.000 & 0.214 & 0.801 \end{bmatrix} \quad (40)$$

which is a very close approximation of the original mixing matrix $\mathbf{H}$ neglecting permutation ambiguity. In this way, we were able to reconstruct the original sources

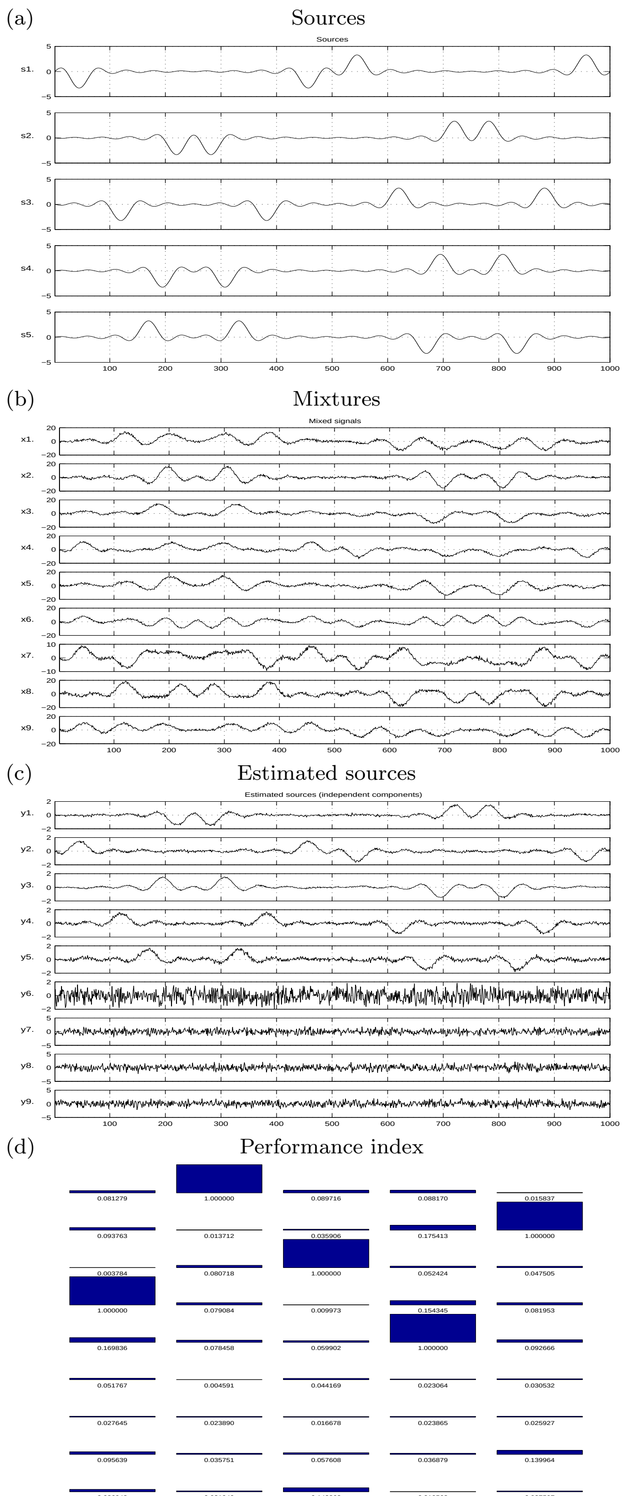

Fig. 1 Plots illustrating Example 1: (a) Original 5 sources. (b) 9 sensor (observed) noisy signals (the number of sources is assumed to be unknown). (c) Estimated source signals + noise, (d) performance matrix $\mathbf{G} = \mathbf{W}\mathbf{H}$.

and estimate the mixing matrix with only permutation ambiguity. We reduced the scaling ambiguity with the assumption that the mixing matrix $\mathbf{H}$ is normalized and

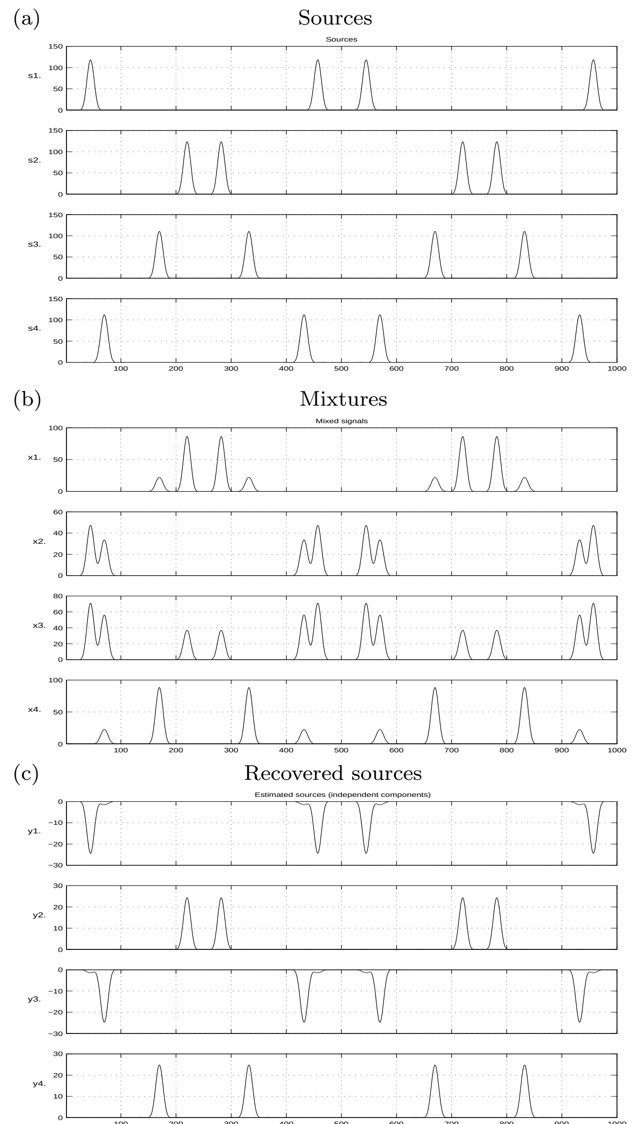

Fig. 2 Plots illustrating Example 2: (a) Original non-negative sources. (b) Mixed (observed) signals. (c) Estimated sources before post-processing.

non-negative.

Example 3 In this experiment 4 human faces are mixed with the Hilbert (ill-conditioned) mixing matrix $\mathbf{H}$. The mixing images shown in Fig. 3 (a) are strongly correlated, thus any classical ICA algorithm would fail to separate them. In order to reconstruct the original images we applied the pre-processing stage first high pass filtering the observed images in order to enhance edges (which appear to be independent for the original images). For such preprocessed mixed images we have applied the learning rule (8) and after convergence of the algorithm the post processing technique described in section 5 has been employed.

The estimated original images are shown in Fig. 3 (b). It should be noted that original images are reconstructed almost perfectly without sign scaling am-

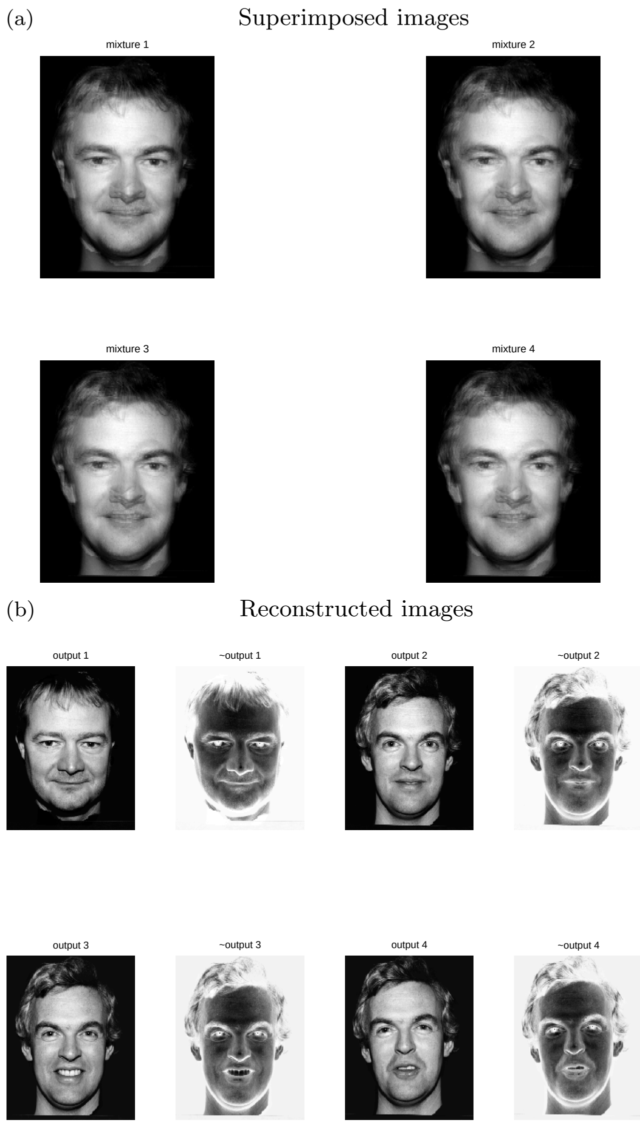

Fig. 3 Example 3: (a) Observed overlapped images. (b) Reconstructed original images using MSD-ICA with high pass filter preprocessing.

biguity. Similar results have been obtained for other randomly generated matrices.

8. Conclusions

In this paper we have presented several extensions and modifications of algorithms for blind source separation and independent component analysis where various constraints are imposed such as symmetries, semi-orthogonality, non-negativity and constant or bounded Frobenius norm. Mathematical analysis and/or computer simulations confirmed validity and satisfactory performance of the proposed algorithms. Moreover, we proposed generalization and extension of ICA to MSD-ICA which relaxes considerably the condition on independence of original sources. Using these concepts in many cases, we are able to reconstruct (recover) the original sources and to estimate mixing and separating matrices, even if the original sources are not independent and in fact they are strongly correlated. In the future work, we are going to establish a self-adaptive system which detect automatically optimal subbands for which ICA should be performed.

References

[1] A. Cichocki and S. Amari, "Adaptive Blind Signal and Image Processing", John Wiley, Chichester, April 2002.
[2] S. Amari, "ICA of temporally correlated signals - learning algorithm", Proc. ICA'99, Aussois, France, pp. 13-18, 1999.
[3] S. Amari, T.-P. Chen and A. Cichocki, "Nonholonomic Orthogonal Learning Algorithms for Blind Source Separation", Neural Computation, Vol.12, pp. 1463-1484, 2000.
[4] S. Amari and A. Cichocki, "Adaptive blind signal processing - neural network approaches", Proc. IEEE (invited paper), Vol.86, No.10, pp. 2026-2048, Oct. 1998.
[5] A. Belouchrani and A. Cichocki, "Robust whitening procedure in blind source separation context", Electronics Letters, vol.36, no.24, pp.2050–2053, 2000.
[6] C. Byrne, "Block-iterative interior point optimization methods for image reconstruction from limited data, Inverse Problems, Vol.15, pp.1405–1419, 2000.
[7] J.-F. Cardoso and B.H. Laheld, "Equivariant adaptive source separation", IEEE Trans. Signal Processing, Vol.44, No.12, pp.3017–3030, December 1996.
[8] A. Cichocki, A. Amari, K. Siwek et al. ICALAB Toolboxes for Signal and Image Processing *www.bsp.brain.riken.go.jp*
[9] A. Cichocki and T. Tanaka, "Non indpendent blind source separtion", NIPS-2002 Postconfrence Workshop "Independent Component Analysis and Beyond", Vancouver, Canada, 2002 (to appear).
[10] A. Cichocki, R. Thawonmas, and S. Amari, "Sequential blind signal extraction in order specified by stochastic properties", Electronics Letters, Vol.33, No.1, pp.64–65, 1997.
[11] A. Cichocki and R. Unbehauen, "Robust neural networks with on-line learning for blind identification and blind separation of sources", "IEEE Trans. Circuits and Systems I : Fundamentals Theory and Applications", Vol.43, No.11, pp.894–906, Nov. 1996.
[12] A. Cichocki, R. Unbehauen, and E. Rummert, "Robust learning algorithm for blind separation of signals", Electronics Letters, Vol.30, No.17, pp.1386–1387, August 1994.
[13] S. Cruces, A. Cichocki, and S. Amari, "The minimum entropy and cumulants based contrast functions for blind source extraction", Proc. 6th Int. Work-Conference on Artificial and Natural Networks: Bio-Inspired Applications of Connectionism, Granada, Spain, Lect. Notes Comp. Sci., Vol.2085, pp.786–793, 2001.
[14] S. Cruces, L. Castedo, and A. Cichocki, "Novel blind source separation algorithms using cumulants", Proc. Int. Conf. Acoustic and Signal Processing, ICASSP'2000, Istanbul, Turkey, Vol.V, pp. 3152–3155, June 2000.
[15] S.C. Douglas, S. Amari, and S.-Y. Kung, "On gradient adaptation with unit-norm constraints", IEEE Trans. Signal Processing, Vol.48, No.6, pp.1843-1847.
[16] S.C. Douglas, "Prewhitened blind source Separation with orthogonality constraints", presented at 33rd Annual Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA, October 24-27, 1999.
[17] A. Edelman, T. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints", SIAM Journal on Matrix Analysis and Applications, Vol.20, no.2, pp.303–353, 1998.

[18] S. Fiori. A theory for learning by weight flow on Stiefel-Grassmann manifold. *Neural Computation*, vol.13, 2001, pp.1625-1637.

[19] P. Georgiev, A. Cichocki, and S. Amari, "On some extensions of the natural gradient algorithm", Proc. 3rd Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA 2001), San Diego, USA pp.581–585, 2001.

[20] P.O. Hoyer, "Non-negative sparse coding", Proc. Workshop on Neural Network for Signal Processing, (NNSP-2002), Martigny, Switzerland, pp.557-565, Sept. 2002

[21] A. Hyvärinen, J. Karhunen, and E. Oja, "Independent Component Analysis", John Wiley, New York, 2001.

[22] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee and T. J. Sejnowski, "Dictionary Learning Algorithms for Sparse Representation", Neural Computation, 2002, Vol.15, No.2, 2003.

[23] S. Lang, "Undergraduate Analysis", Second Edition, Springer, New York, 1997.

[24] J. S. Lee, D. D. Lee, S. Choi, K. S. Park, and D. S. Lee. "Non-negative matrix factorization of dynamic images in nuclear medicine", Proc. IEEE Medical Imaging Conference, San Diego, California, November 4-10, 2001.

[25] D. D. Lee and H. S. Seung, "Learning of the parts of objects by non-negative matrix factorization", Nature, Vol. 401, p.788, Oct. 1999

[26] J. Nagy and Z. Strakos, "Enforcing non-negativity in image reconstruction algorithms", Mathematical Modeling, Estimation, and Imaging, Eds. David C. et.al., pp.182–190, 2000.

[27] Y. Nishimori. "Learning algorithm for ICA by geodesic flows and on orthogonal group". *In Proc. International Joint Conference on Neural Networks (IJCNN'99)*, Washington DC, 10-16 July 1999, Vol.2, pp.933-938.

[28] L. Parra, K.R. Mueller, C. Spence, A. Ziehe, P. Sajda. "Unmixing Hyperspectral Data", Advances in Neural Information Processing Systems 12, MIT Press, pp. 942-948, 2000.

[29] M. D. Plumbley, "Algorithms for non-negative independent component analysis", to appear in IEEE Trans on Neural Networks, 2003, web page
$http://www.eee.kcl.ac.uk/member/staff/m_plumbley/$.

[30] M. D. Plumbley. "Adaptive lateral inhibition for non-negative ICA", Proc. 3rd Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA 2001), San Diego, USA pp.581–585, 2001. web page
$http://www.eee.kcl.ac.uk/member/staff/m_plumbley$.

[31] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection", IEEE Trans. Sig. Proc., vol.47, no.1, pp.187–200, January 1999.

[32] L. Tong, V.C. Soon, Y. Inouye, Y. Huang, and R. Liu, "Waveform-preserving blind estimation of multiple sources", Proc. IEEE Conference on Decision and Control, Brighton, UK, Vol.3, pp.2388–2393, 1991.

[33] L. Tong, Y. Inouye, and R. Liu, "A finite-step global algorithm for the parameter estimation of multichannel MA processes", IEEE Trans. Signal Proc., vol.40, no.10, pp.2547–2558, 1992.

[34] L. Tong, S. Yu, Y. Inouye, and R. Liu, "A necessary and sufficient condition of blind identification", J.L. Lacoume, ed., Higher Order Statistics, Elsevier, Proc. Int. Signal Processing Workshop, Chamrousse, France, pp.263–266, 1992.

[35] M. Zibulevsky, P. Kisilev, Y.Y. Zeevi, B.A. Pearlmutter, "Blind source separation via multinode sparse representation", Morgan Kaufmann, ed., Advances in Neural Information Processing Systems, (NIPS2001) vol.14. , pp. 185-191, 2002.

Andrzej Cichocki received the M.Sc.(with honors), Ph.D., and Habilitate Doctorate (Dr.Sc.) degrees, all in electrical engineering, from Warsaw University of Technology (Poland) in 1972, 1975, and 1982, respectively. Since 1972, he has been with the Institute of Theory of Electrical Engineering and Electrical Measurements at the Warsaw University of Technology, where he became a full Professor in 1991. He is the co-author of three books: Adaptive Blind Signal and Image Processing - Learning Algorithms and Applications, Wiley, 2002, MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems (Springer-Verlag, 1989) and Neural Networks for Optimization and Signal Processing (Teubner-Wiley,1993/94) and more than 150 research papers. He spent at University Erlangen-Nuernberg (Germany) a few years as Alexander Humboldt Research Fellow and Guest Professor.
Since 1995 he has been working in the Brain Science Institute RIKEN (Japan), as a team leader for laboratory for Open Information Systems and currently as a head of laboratory for Advanced Brain Signal Processing. His current research interests include optimization, bio-informatics, neurocomputing and signal and image processing, especially analysis and processing of multi-sensory biomedical data.
He is currently associate editor of IEEE Transaction on Neural Networks and Fluctuation and Noise Letters. He was 3 years the member of the Technical Committee for Neural Network for Signal Processing and recently the member of core group who established a new IEEE Circuits and Systems Technical Committee for Blind Signal Processing.
More information about his research activities can be found at
http://www.bsp.brain.riken.go.jp/

Pando G. Georgiev received his MS, PhD and "Doctor of Mathematical Sciences" degrees in Mathematics (Operations Research) from Sofia University "St. Kl. Ohridski", Bulgaria, in 1982, 1987 and 2001 respectively. He has been with the Department of Probability, Operations Research and Statistics at Faculty of Mathematics and Informatics, Sofia University "St. Kl. Ohridski", Bulgaria, as Assistant Professor (1989-1994) and since 1994, as Associate Professor. He was a Visiting Professor at the University of Rome II, Italy (CNR grants), the International Center for Theoretical Physics, Trieste, Italy (ICTP grant, six months), the University of Pau, France (NATO grant, three months), Hirosaki University, Japan (JSPS grant, nine months). Since October 2000 he has been working as a research scientist at the Laboratory of Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Wako-shi. His interests include Optimization and Operations Research, Signal and Image Processing, Independent Component Analysis and Blind Signal Separation, Neural Networks, Probability and Statistics, Numerical Methods and Algorithms, Nonlinear and Nonsmooth Analysis. He is a member of AMS, IEEE, UBM.