# SIMULTANEOUS MATRIX DIAGONALIZATION: THE OVERCOMPLETE CASE

*Lieven De Lathauwer*

ETIS (ENSEA, UCP, CNRS)
Cergy-Pontoise
France
delathau@ensea.fr

## ABSTRACT

Many algorithms for Independent Component Analysis rely on a simultaneous diagonalization of a set of matrices by means of a nonsingular matrix. In this paper we provide means to determine the matrix when it has more columns than rows.

## 1. INTRODUCTION

A lot of techniques for Independent Component Analysis (ICA) or Blind Source Separation (BSS) are based on a simultaneous congruence transformation. Given a set of matrices $\mathbf{A_1}, \ldots, \mathbf{A_K} \in \mathbb{C}^{N \times N}$, the aim is to find a nonsingular matrix $\mathbf{M} \in \mathbb{C}^{N \times N}$ such that

$$\begin{aligned} \mathbf{A_1} &= \mathbf{M} \cdot \mathbf{D_1} \cdot \mathbf{M}^H \\ &\vdots \\ \mathbf{A_K} &= \mathbf{M} \cdot \mathbf{D_K} \cdot \mathbf{M}^H, \end{aligned} \tag{1}$$

with $\mathbf{D_1}, \ldots, \mathbf{D_K} \in \mathbb{C}^{N \times N}$ as diagonal as possible according to some criterion. Examples are the JADE algorithm for the separation of non-Gaussian sources [3], the SOBI algorithm for the separation of sources that are mutually uncorrelated but individually exhibit some correlation in time [1], the algorithm proposed in [10] for the separation of non-stationary sources subject to a constant mixing, the algorithm described in [2] for source separation in the time-frequency domain (also for non-stationary sources), ... In JADE, one of the given matrices is the observed spatial covariance; the corresponding diagonal matrix is the covariance of the sources and the other given matrices are matrix slices of the spatial higher-order cumulant of the observations. In SOBI, the matrices $\{\mathbf{A_k}\}$ and the matrices $\{\mathbf{D_k}\}$ are the spatial covariance matrices for different time lags, of the observations and the sources, respectively. In [10] they correspond to spatial covariance matrices measured at different time instances, and in [2] they correspond to Spatial Time-Frequency Distributions, of the observations and the sources, respectively.

The original algorithms to solve (1) are prewhitening-based. In the prewhitening step, one picks a positive (semi-) definite matrix from $\{\mathbf{A_k}\}$, say $\mathbf{A_1}$, and computes its Eigenvalue Decomposition (EVD). (Actually, it is sufficient to compute any square root of $\mathbf{A_1}$, but the EVD allows for an optimal rank reduction when there are less sources than sensors.) Define $\mathbf{M}' = \mathbf{M} \cdot \mathbf{D_1}^{1/2}$, and let its Singular Value Decomposition (SVD) be given by

$$\mathbf{M}' = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^H, \tag{2}$$

then we have that

$$\mathbf{A_1} = \mathbf{M}' \cdot \mathbf{M}'^H = \mathbf{U} \cdot \mathbf{S}^2 \cdot \mathbf{U}^H. \tag{3}$$

By substituting (2), (3) in (1), we obtain a simultaneous *unitary* congruence transformation in the remaining unknown $\mathbf{V}$. As is well-known, the latter problem can be solved by means of a Jacobi-iteration.

In this prewhitening-based procedure, one computes more than half of the parameters in the SVD of $\mathbf{M}'$ from a single equation in (1), which is exactly satisfied. The other equations serve to estimate the factor $\mathbf{V}$; these equations are only approximately satisfied. It is intuitively clear that, instead, it is preferable to deal with the decompositions in (1) (or versions of them that are weighted / combined in an intelligent way) simultaneously, instead of sequentially. We call this principle "soft whitening". The idea was first proposed by Yeredor [15, 16]. The problem is now to find a non-unitary matrix $\mathbf{M}$ that approximately diagonalizes $\{\mathbf{A_k}\}$ in (1). Algorithms have been proposed in [5, 6, 8, 11, 13, 14, 16].

In this paper we will solve the problem when $\mathbf{M}$ has more columns than rows, i.e., $\mathbf{M} \in \mathbb{C}^{N \times R}$ and $\mathbf{D_1}, \ldots, \mathbf{D_K} \in \mathbb{C}^{R \times R}$, with $R > N$. In the context of ICA, this is often called the "underdetermined" or "overcomplete" case. We will restrict ourselves to the identification of the mixing matrix and not consider the estimation of the source values.

As far as hard vs soft whitening is concerned, note also the following. When $R \leqslant N$, a hard whitening causes the remaining unknown matrix to be unitary, and hence optimally conditioned. This advantage is lost in the overcomplete case: the fact that $\mathbf{V}$ has mutually orthonormal columns

does not prevent rows from being arbitrarily close. Hence, the remaining simultaneous diagonalization may still be ill-conditioned in the sense that it may involve rank-1 terms that are close.

Our approach is an algebraic one. The tools we borrow from multilinear algebra will be explained in Section 2. Section 3 casts our problem in the framework of multilinear algebra. Section 4 deals with the case where $R > K$. More powerful results are obtained for the case where $R \leqslant K$ in Section 5. Section 6 is the conclusion.

## 2. ALGEBRAIC PRELIMINARIES

Multilinear algebra is the algebra of higher-order tensors. These may, for the purposes of this paper, be imagined as multi-indexed arrays (a vector is a first-order tensor and a matrix is a second-order tensor). The concept of rank is much more involved in multilinear algebra than in matrix algebra. We will use the following definitions.

*Definition 1.* [Rank] The rank $R$ of a matrix $\mathbf{A}$ is the maximal number of columns (or, equivalently, rows) of $\mathbf{A}$ that form a linearly independent set.

*Definition 2.* [Kruskal-rank [9]] The $k$-rank $k$ of a matrix $\mathbf{A}$ is the maximal number such that any set of $k$ columns of $\mathbf{A}$ is linearly independent.

By definition, we have that $k(\mathbf{A}) \leqslant R(\mathbf{A})$.

*Definition 3.* [Rank-1 tensor] An $N$th-order tensor $\mathcal{A}$ has rank 1 when it equals the outer product of $N$ vectors $U^{(1)}$, $U^{(2)}$, ..., $U^{(N)}$:

$$a_{i_1 i_2 \ldots i_N} = u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(N)}_{i_N}$$

for all values of the indices.

The outer product of $U^{(1)}$, $U^{(2)}$, ..., $U^{(N)}$ will be written as $U^{(1)} \circ U^{(2)} \circ \ldots \circ U^{(N)}$.

*Definition 4.* [Tensor rank] The rank $R$ of an $N$th-order tensor $\mathcal{A}$ is the minimal number of rank-1 tensors that yield $\mathcal{A}$ in a linear combination.

In the literature, such an expansion is called a "Canonical Decomposition" (CANDECOMP) [4] or a "Parallel Factors Model" (PARAFAC) [8]. In contrast to the matrix case, it can be unique even when the rank-1 terms are not mutually orthogonal. Moreover, the rank of an $(I_1 \times I_2 \times \ldots \times I_N)$-tensor is in general not bounded by $\min\{I_1, I_2, \ldots, I_N\}$, as for matrices. These two properties form the basis of our contribution.

We have the following uniqueness theorem [9, 12]:

| $N$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| $R_{max}$ | 2 | 5 | 8 | 12 | 18 | 24 | 31 |

**Table 1**. Real case: the maximum number of columns of $\mathbf{M}$ that can be allowed in Alg. 1, provided $R_{max} \leqslant K$.

*Theorem 1.* Consider $\mathbf{U} = [U_1 U_2 \ldots U_R] \in \mathbb{C}^{I_1 \times R}$, $\mathbf{V} = [V_1 V_2 \ldots V_R] \in \mathbb{C}^{I_2 \times R}$, $\mathbf{W} = [W_1 W_2 \ldots W_R] \in \mathbb{C}^{I_3 \times R}$ and a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ given by

$$\mathcal{A} = \sum_{r=1}^{R} U_r \circ V_r \circ W_r. \tag{4}$$

If

$$k(\mathbf{U}) + k(\mathbf{V}) + k(\mathbf{W}) \geqslant 2(R+1), \tag{5}$$

then the rank-1 terms in decomposition (4) are unique. In terms of **U**, **V** and **W**, this means that any other triple **U'**, **V'**, **W'** that defines a CANDECOMP of $\mathcal{A}$, is related to **U**, **V**, **W** via

$$\mathbf{U} = \mathbf{U}' \cdot \mathbf{P} \cdot \mathbf{D_1} \qquad \mathbf{V} = \mathbf{V}' \cdot \mathbf{P} \cdot \mathbf{D_2} \qquad \mathbf{W} = \mathbf{W}' \cdot \mathbf{P} \cdot \mathbf{D_3}, \tag{6}$$

in which $\mathbf{D_1}$, $\mathbf{D_2}$, $\mathbf{D_3}$ are diagonal matrices, satisfying $\mathbf{D_1} \cdot \mathbf{D_2} \cdot \mathbf{D_3} = \mathbf{I}$, and $\mathbf{P}$ is a permutation matrix.

Along the lines of [7], we also have:

*Theorem 2.* Consider $\mathbf{U} = [U_1 U_2 \ldots U_R] \in \mathbb{C}^{N \times R}$, $\mathbf{V} = [V_1 V_2 \ldots V_R] \in \mathbb{C}^{K \times R}$ and a tensor $\mathcal{A} \in \mathbb{C}^{N \times N \times K}$ given by

$$\mathcal{A} = \sum_{r=1}^{R} U_r \circ U_r^* \circ V_r. \tag{7}$$

If $R \leqslant K$, then we generically have that the rank-1 terms in decomposition (7) are unique if $2R(R-1) \leqslant N^3(N-1)$. For real-valued tensors, we have uniqueness if $R \leqslant R_{max}$, with the values of $R_{max}$ for $N = 2, \ldots, 8$ listed in Table 1.

We call a property "generic" when it holds for all matrices, except for a set of Lebesgue measure 0.

## 3. PROBLEM REFORMULATION

Let us stack the matrices $\mathbf{A_1}, \ldots, \mathbf{A_K}$ in Eq. (1) in a tensor $\mathcal{A} \in \mathbb{C}^{N \times N \times K}$. Define a matrix $\mathbf{D} \in \mathbb{C}^{K \times R}$ as follows:

$$\mathbf{D} = \begin{pmatrix} \mathrm{diag}(\mathbf{D_1}) \\ \vdots \\ \mathrm{diag}(\mathbf{D_K}) \end{pmatrix}, \tag{8}$$

in which $\mathrm{diag}(\cdot)$ is the operator that extracts the diagonal from its argument and puts it in a row. Then we have that

$$\mathcal{A} = \sum_{r=1}^{R} M_r \circ M_r^* \circ D_r, \tag{9}$$

in which $\{M_r\}$ and $\{D_r\}$ are the columns of $\mathbf{M}$ and $\mathbf{D}$, respectively. Eq. (9) is a CANDECOMP of our data tensor. Since its rank is not bounded by $N$, and since the decomposition may still be unique when $R > N$, we may be able to determine $\mathbf{M}$ (up to a scaling and permutation of its columns) for overcomplete mixtures.

We will now discuss some means to calculate $\mathbf{M}$.

## 4. CASE 1: $R > K$

Generically, a matrix is full rank and full k-rank. Hence, in practice, $k(\mathbf{M}) = \min(N, R) = N$ and $k(\mathbf{D}) = \min(K, R) = K$. Theorem 1 then guarantees identifiability if $2N + K \geqslant 2R + 2$, i.e., when $R \leqslant N - 1 + K/2$.

Two methods for the calculation of the canonical components have been described in the literature. They are both of the "alternating least squares" (ALS) type. The aim is to minimize the (squared) Frobenius norm of the difference between $\mathcal{A}$ and its estimated decomposition in rank-1 terms by means of an iteration in which each step consists of fixing a subset of unknown parameters to their current estimates, and optimizing w.r.t. the remaining unknowns, followed by fixing an other subset of parameters, and optimizing w.r.t. the complimentary set, etc. (Like for matrices, the squared Frobenius norm of a tensor is the sum of the squared moduli of its entries.)

In [8], one optimizes the cost function

$$f(\mathbf{U}, \mathbf{V}, \mathbf{W}) = \|\mathcal{A} - \sum_{r=1}^{R} U_r \circ V_r^* \circ W_r\|^2. \qquad (10)$$

Due to the multi-linearity of the model, estimation of one of the arguments, given the other two, is a classical linear least squares problem. One alternates between updates of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$. After updating $\mathbf{U}$ and $\mathbf{V}$, their columns are rescaled to unit length, to avoid under- and overflow. Although during the iteration the symmetry of the problem is broken, one supposes that eventually $\mathbf{U}$ and $\mathbf{V}$ both converge to $\mathbf{M}$.

The algorithm described in [15, 16] has been developed for the situation in which $N \geqslant R$, but it can actually be applied to the overdetermined case as well. Here the symmetry is kept ($\mathbf{U} = \mathbf{V}$), and one alternates between updates of a single column of $\mathbf{U}$ and updates of $\mathbf{W}$. Updating a column of $\mathbf{U}$ involves computing the dominant positive eigenvalue of an $(N \times N)$ Hermitean matrix. Like in the previous method, updating $\mathbf{W}$ is an ordinary linear least squares problem.

In both techniques, the rank of A has to be estimated by trial-and-error.

## 5. CASE 2: $R \leqslant K$

Rearrange $\mathcal{A}$ in an $(N^2 \times K)$ matrix $\mathbf{B}$. In this matrix notation, the basic model (9) is written as

$$\mathbf{B} = (\mathbf{M} \odot \mathbf{M}^*) \cdot \mathbf{D}^T, \qquad (11)$$

in which $\odot$ symbolizes the Katri-Rao or column-wise Kronecker product.

Hence, $R$ can generically be estimated as the rank of $\mathbf{B}$. Let the columns of $\mathbf{E} \in \mathbb{C}^{N^2 \times R}$ span the dominant left singular subspace of $\mathbf{B}$, then the problem is to find a nonsingular matrix $\mathbf{F} \in \mathbb{C}^{R \times R}$ such that the columns of $\mathbf{E} \cdot \mathbf{F}$ are vectorized rank-1 matrices. This is the same problem as in [7], except for the fact that we are now looking for a matrix that is merely complex nonsingular, instead of real orthogonal. Again, we have that the solution is essentially unique if $2R(R-1) \leqslant N^3(N-1)$ (complex case) or $R \leqslant R_{\max}$, with $R_{\max}$ given in Table 1 (real case). This means that, under the condition on $R$, any $\mathbf{F}$ that induces the rank-1 structure causes $\mathbf{M} \odot \mathbf{M}^*$ and $\mathbf{E} \cdot \mathbf{F}$ to be equal up to a scaling and a permutation of their columns, and vice-versa. $\mathbf{F}$ can be estimated from

$$\begin{aligned} \mathbf{C_1} &= \mathbf{F} \cdot \mathbf{\Lambda_1} \cdot \mathbf{F}^H \\ &\vdots \\ \mathbf{C_R} &= \mathbf{F} \cdot \mathbf{\Lambda_R} \cdot \mathbf{F}^H, \end{aligned} \qquad (12)$$

in which $\mathbf{\Lambda_1}, \mathbf{\Lambda_2}, \ldots, \mathbf{\Lambda_R} \in \mathbb{R}^{R \times R}$ are as diagonal as possible and in which $\mathbf{C_1}, \mathbf{C_2}, \ldots, \mathbf{C_R}$ are $(R \times R)$ matrices, constructed from $\mathbf{E}$ in the same way as explained in [7], Section 3 (see algorithmic description below). In other words, we have reformulated the overcomplete case as a regular simultaneous congruence transformation, involving a square matrix $\mathbf{F}$. References that address this problem have been listed in the Introduction. Note that the noise-free solution is given by the EVD

$$\mathbf{C_{r_1}} \mathbf{C_{r_2}}^{-1} = \mathbf{F} \cdot (\mathbf{\Lambda_{r_1}} \mathbf{\Lambda_{r_2}}^{-1}) \cdot \mathbf{F}^{-1}. \qquad (13)$$

The overall algorithm takes the following form:

*Algorithm 1*

*In:* matrices $\mathbf{A_1}, \ldots, \mathbf{A_K} \in \mathbb{C}^{N \times N}$
*Out:* $\mathbf{M} \in \mathbb{C}^{N \times R}$ that (approximately) diagonalizes $\{\mathbf{A_k}\}$ in the sense of Eq. (1).

*Step 1:* $B_k = \text{vec}(\mathbf{A_k})$. $[B_1 \ldots B_K] = \mathbf{B} \in \mathbb{C}^{N^2 \times K}$.
*Step 2:* $R$ is the dimension of the signal subspace of $\mathbf{B}$. Compute the $R$ dominant left singular vectors $U_r$ and values $\sigma_r$. $\mathbf{E_r} = \text{unvec}(\sigma_r U_r)$, with $\mathbf{E_r} \in \mathbb{C}^{N \times N}$.
*Step 3:* Construct $\{\mathbf{\Psi_{rs}} \in \mathbb{C}^{N \times N \times N \times N}\}_{1 \leqslant r,s \leqslant R}$:

$$(\mathbf{\Psi_{rs}})_{ijkl} = (\mathbf{E_r})_{ij} (\mathbf{E_s})^*_{kl} - (\mathbf{E_r})_{ik} (\mathbf{E_s})^*_{jl}.$$

*Step 4:* $G_{(r-1)R+s} = \text{vec}(\mathbf{\Psi_{rs}})$. $[G_1 \dots G_{R^2}] = \mathbf{G} \in \mathbb{C}^{N^4 \times R^2}$. Compute the right singular vectors $V_r$ corresponding to the smallest singular values $\lambda_r$. $\mathbf{C_r} = \text{unvec}(V_r)$ (possibly weighted inversely proportional to $\lambda_r$), with $\mathbf{C_r} \in \mathbb{C}^{R \times R}$.

*Step 5:* Obtain $\mathbf{F} \in \mathbb{C}^{R \times R}$ from the simultaneous diagonalization [13]

$$\begin{aligned} \mathbf{C_1} &= \mathbf{F} \cdot \mathbf{\Lambda_1} \cdot \mathbf{F}^H \\ &\vdots \\ \mathbf{C_R} &= \mathbf{F} \cdot \mathbf{\Lambda_R} \cdot \mathbf{F}^H. \end{aligned}$$

*Step 6:* $E_r = \text{vec}(\mathbf{E_r})$. $[E_1 \dots E_R] = \mathbf{E} \in \mathbb{C}^{N^2 \times R}$. $\mathbf{E} \cdot \mathbf{F} = [H_1 \dots H_R] \in \mathbb{C}^{N^2 \times R}$. $\mathbf{H_r} = \text{unvec}(H_r)$, with $\mathbf{H_r} \in \mathbb{C}^{N \times N}$.

*Step 7:* Compute $M_r$ as the dominant left singular vector of $\mathbf{H_r}$. $\mathbf{M} = [M_1 M_2 \dots M_R]$.

[ *Step 8:* Optimize Eq. (1) in least-squares sense [8, 16]. ]

In this algorithmic description, vec($\mathbf{X}$) is a vector representation of a matrix $\mathbf{X}$, in which all the columns are stacked one after the other. More in general, we assume that, in vec($[x_{ijkl}]$), the index $l$ varies faster than index $k$, which in turn varies faster than index $j$, etc. unvec($\bullet$) is the inverse operation of vec($\bullet$).

## 6. CONCLUSION

Many signal processing problems, including several variants of ICA / BSS, can be formulated as a search for a set of $R$ symmetric rank-1 matrices of which linear combinations yield a set of $K$ given $(N \times N)$ matrices. In this paper we have considered the case where $N < R$. We have rephrased this problem as a CANDECOMP of a third-order tensor and proposed multilinear algebraic means to compute this decomposition.

**Acknowledgements**

Lieven De Lathauwer holds a permanent research position with the French Centre National de la Recherche Scientifique (C.N.R.S.) (http://www-etis.ensea.fr); he also holds a honorary research position with the K.U.Leuven (http://www.esat.kuleuven.ac.be/sista-cosic-docarch).

This work is supported by several institutions:

1. Research Council K.U.Leuven: Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology),
2. Flemish Government: the F.W.O. projects G.0240.99 (Multilinear Generalisations of the Singular Value Decomposition and Applications in Signal Processing and System Identification) and G.0256.97 (Numerical Algorithms for Subspace System Identification, Extension to Special Cases), the F.W.O. Research Communities ICCoS (Identification and Control of Complex Systems) and ANMMM (Advanced Numerical Methods for Mathematical Modelling),
3. Belgian Federal Government: Interuniversity Poles of Attraction Programme IUAP IV-02 (1997-2001): Modeling, Identification, Simulation and Control of Complex Systems, and IUAP V-22 (2002-2006): Dynamical Systems and Control: Computation, Identification & Modelling.

The scientific responsibility is assumed by the author.

## 7. REFERENCES

[1] A. Belouchrani, et al., "A Blind Source Separation Technique using Second Order Statistics", *IEEE Trans. Signal Processing*, Vol. 45, No. 2, Feb. 1997, pp. 434–444.

[2] A. Belouchrani, M.G. Amin, "Blind Source Separation Based On Time-Frequency Signal Representations", *IEEE Trans. Signal Processing*, Vol. 46, No. 11, Nov. 1998, pp. 2888–2897.

[3] J.-F. Cardoso, A. Souloumiac, "Blind Beamforming for Non-Gaussian Signals", *IEE Proceedings-F*, Vol. 140, No. 6, 1994, pp. 362–370.

[4] J.D. Carroll, J. Chang, "Analysis of Individual Differences in Multidimensional Scaling via an $N$-Way Generalization of "Eckart-Young" Decompositions", *Psychometrika*, Vol. 35, 1970, pp. 283-319.

[5] L. De Lathauwer, *Signal Processing Based on Multilinear Algebra*, Ph.D. thesis, K.U.Leuven, E.E. Dept., Belgium, 1997.

[6] L. De Lathauwer, J. Vandewalle, B. De Moor, "Jacobi-Algorithm for Simultaneous Generalized Schur Decomposition in Higher-Order-Only ICA", *Proc. IEEE Benelux Signal Processing Chapter Signal Processing Symposium (SPS'98)*, Leuven, Belgium, March 26–27, 1998, pp. 67–70.

[7] L. De Lathauwer, B. De Moor, J. Vandewalle, J.-F. Cardoso, "Independent Component Analysis of Largely Underdetermined Mixtures", *Proc. ICA 2003*.

[8] R.A. Harshman, "Foundations of the PARAFAC Procedure: Model and Conditions for an "Explanatory" Multi-Mode Factor Analysis", *UCLA Working Papers in Phonetics*, Vol. 16, No. 1, 1970.

[9] J.B. Kruskal, "Three-Way Arrays: Rank and Uniqueness of Trilinear Decompositions, with Application to Arithmetic Complexity and Statistics", *Linear Algebra and its Applications*, No. 18, 1977, pp. 95–138.

[10] D.-T. Pham, J.-F. Cardoso, "Blind Separation of Instantaneous Mixtures of Non-stationary Sources", *IEEE Trans. Signal Processing*, Vol. 49, No. 9, pp. 1837–1848.

[11] D.-T. Pham, “Joint Approximate Diagonalization of Positive Definite Hermitian Matrices”, *SIAM J. Matrix Anal. Appl.*, Vol. 22, No. 4, 2001, pp. 1136–1152.

[12] N. Sidiropoulos, G. Giannakis, R. Bro, “Blind PARAFAC Receivers for DS-CDMA Systems”, *IEEE Trans. on Signal Processing*, Vol. 48, 2000, pp. 810–823.

[13] A.-J. van der Veen, A. Paulraj, “An Analytical Constant Modulus Algorithm”, *IEEE Trans. Signal Processing*, Vol. 44, 1996, pp. 1136–1155.

[14] A.-J. van der Veen, “Joint Diagonalization via Subspace Fitting Techniques”, *Proc. IEEE ICASSP*, Salt Lake City (UT), May 2001.

[15] A. Yeredor, “Approximate Joint Diagonalization Using Non-Orthogonal Matrices”, *Proc. ICA 2000*, June 19–22, Helsinki, Finland, 2000, pp. 33–38.

[16] A. Yeredor, “Non-Orthogonal Joint Diagonalization in the Least-Squares Sense With Application in Blind Source Separation”, *IEEE Trans. Signal Processing*, Vol. 50, No. 7, July 2002, pp. 1545–1553.