A Fully-Multiplicative Orthogonal-Group ICA Neural Algorithm

Simone Fiori*

August 29, 2003

Abstract

The aim of this Letter is to present a fully-multiplicative, batch learning algorithm for neural independent component analysis on the orthogonal group $O(N)$. It exploits the known principle of diagonalization of a tensor of warped network's outputs, which is achieved through fast iteration.

Indexing terms. Neural networks, independent component analysis, orthogonal group.

1 Introduction

The independent component analysis (ICA) technique aims at recovering statistically independent source signals from their mixtures [1, 6]. In the present Letter we consider linear, instantaneous, noiseless mixtures of the form $\mathbf{x} = \mathbf{M}\mathbf{s}$, where $\mathbf{s} \in \mathcal{R}^N$ denotes the vector of zero-mean, stationary, statistically indepen-

*The author is with the Faculty of Engineering of the Perugia University, Polo Didattico e Scientifico del Ternano, Loc. Pentima bassa, 21, I-05100 Terni (Italy). Email: sfr@unipg.it (S. Fiori).

dent signals, $\mathbf{x} \in \mathcal{R}^N$ denotes the vector of observable signals and $\mathbf{M} \in \mathcal{R}^{N\times N}$ is a full-rank constant mixing matrix.

The separating neural network has the input-output description $\mathbf{y} = \mathbf{C}^T\mathbf{x}$, where $\mathbf{y} \in \mathcal{R}^N$ denotes the network response vector, $\mathbf{C} \in \mathcal{R}^{N\times N}$ denotes the neural network connection matrix.

It is well-known that every full-rank mixture may be reduced to an orthogonal mixture through *whitening* [1, 6], which aim at removing second-order statistics from the signal $\mathbf{x}$ by making its covariance matrix equal to the identity. In this case, the separating network may be described by an orthogonal connection pattern as well, thus we shall also assume $\mathbf{C}$ belonging to the orthogonal group $O(N) \stackrel{\text{def}}{=} \{\mathbf{A} \in \mathcal{R}^{N\times N} | \mathbf{A}^T\mathbf{A} = \mathbf{I}_N\}$. This fact may be advantageously exploited to design learning algorithms that make the network connection pattern belong to the orthogonal group at any time [3, 4].

A class of ICA algorithms stems from the following well-known principle: Given two odd non-linear functions $\phi(\cdot)$ and $\psi(\cdot)$, the separating network ought to diagonalize the matrix $\mathbb{E}_{\mathbf{y}}[\phi(\mathbf{y})\psi(\mathbf{y}^T)]$, where the symbol $\mathbb{E}_{\mathbf{x}}[f(\mathbf{x})]$ denotes statistical expectation of the function $f(\cdot)$ over the distribution of the random-vector $\mathbf{x}$. Note that, in the present Letter, a vector always denotes a *column array*.

The above principle was initially exploited in the neural-network field by Jutten and Hérault (see [6] and references therein) and later on by Cichocki, Umbehauen and Rummert [2]).

In the next section we propose a fully-multiplicative orthogonal-group ICA neural algorithm that enables us to analyze simultaneously the components forming a mixture, on the basis of the above principle.

2 Fully-multiplicative orthogonal-group learning algorithm

Let us define the matrix $\mathbf{G} \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{y}}[\phi(\mathbf{y})\psi(\mathbf{y}^T)] \in \mathcal{R}^{N\times N}$ and let us also define a $N \times N$ real, invertible, positive-definite diagonal matrix $\mathbf{\Gamma}$. The above recalled diagonalization principle may be expressed in the following way which proves useful in the following development:

$$\mathbf{G}^{-1} = \mathbf{\Gamma}^{-1} \ , \tag{1}$$

By pre-multiplying by $\mathbf{C}$ and post-multiplying by $\mathbf{\Gamma}$ both members of the above equation by $\mathbf{C}$ we obtain:

$$\mathbf{C}\mathbf{G}^{-1}\mathbf{\Gamma} = \mathbf{C} \ . \tag{2}$$

This expression readily suggests the following batch iterative learning algorithm:

$$\begin{cases} \tilde{\mathbf{C}}_{n+1} &= \mathbf{C}_n\mathbf{G}_n^{-1}\mathbf{\Gamma} \ , \ n = 0, 1, 2, \ldots \ , \\ \mathbf{C}_{n+1} &= (\tilde{\mathbf{C}}_{n+1}\tilde{\mathbf{C}}_{n+1}^T)^{-1/2}\tilde{\mathbf{C}}_{n+1} \ , \end{cases} \tag{3}$$

where $\mathbf{G}_n \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{y}}[\phi(\mathbf{y}_n)\psi(\mathbf{y}_n^T)]$, with $\mathbf{y}_n \stackrel{\text{def}}{=} \mathbf{C}_n^T\mathbf{x}$, $\mathbf{C}_0 \in O(N)$ is an initial guess and n denotes the iteration index. The last expression denotes the symmetric orthogonalization step that it necessary to project the partial update matrix $\tilde{\mathbf{C}}_{n+1}$ into the orthogonal group. It is convenient to plug the first of equations (3) into the second equation to obtain a single update rule:

$$\mathbf{C}_{n+1} = \sqrt{\mathbf{C}_n\mathbf{G}_n^T\mathbf{\Gamma}^2\mathbf{G}_n\mathbf{C}_n^T}\mathbf{C}_n\mathbf{G}_n^{-1}\mathbf{\Gamma} \ , \ n = 0, 1, 2, \ldots \ . \tag{4}$$

The matrix square-root may be computed via the Parlett's algorithm [5] that is based on the Schur decomposition of the argument. Note that, in the expression (4), the argument of the square-root operator is symmetric, thus its Schur form is diagonal and the square-root may be computed accurately.

As anticipated, the algorithm is expressed in a fully-multiplicative fashion. Also, it is interesting to observe that, in contrast to other well-known on-line or batch-type algorithms, it does not stem from first-order nor second-order (Newton-like) cost-function optimization.

3 Experimental results

In order to assess the numerical behavior of the presented algorithm, the result of the same experiment considered in [2] are illustrated and commented. It concerns the analysis of five synthetic signals ($N = 5$) having very different scales, mixed by a 5×5 Hilbert matrix, which possess the characteristic of being really badly conditioned.

As mentioned, signal pre-whitening is performed by computing the covariance matrix $\mathbf{\Sigma}_x \stackrel{\text{def}}{=} \mathbb{E}[\mathbf{x}\mathbf{x}^T]$, its eigen-pair $(\mathbf{E}, \mathbf{\Lambda})$ such that $\mathbf{\Sigma}_x = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T$ and then the projection $\sqrt{\mathbf{\Lambda}}\mathbf{E}^T\mathbf{x}$. As non-linear functions, according to [2] we selected $\phi(u) = \tanh(2u)$ and $\psi(u) = u^2\text{sign}(u)$. The actual expectation value $\mathbb{E}_{\mathbf{y}}[\phi(\mathbf{y})\psi(\mathbf{y}^T)]$ is replaced by sample mean. The simplest choice for the constant $\mathbf{\Gamma} = \mathbf{I}_N$ is assumed, that is, the learning algorithm writes:

$$\begin{cases} \mathbf{C}_{n+1} & = & \sqrt{\mathbf{C}_n\mathbf{G}_n^T\mathbf{G}_n\mathbf{C}_n^T}\mathbf{C}_n\mathbf{G}_n^{-1} \ , \ n = 0, 1, 2, \ldots \ , \\ \mathbf{C}_0 & = & \mathbf{I}_N \ . \end{cases}$$

As performance indices, the inter-channel interference (ICI) was selected to measure the separation ability:

$$ICI \stackrel{\text{def}}{=} \frac{\sum_{ij} D_{ij}^2}{\sum_i \max_k\{D_{ik}^2\}} - 1 \ , \tag{5}$$

where the *separation product* $\mathbf{D}$ is defined by $\mathbf{y} = \mathbf{D}\mathbf{s}$ and compactly represents the deviation of the estimated sources with respect to the true source signals. Note that, as long as the network connection pattern belongs to $O(N)$ then $\det(\mathbf{D}) \neq 0$.

The Figure 1 illustrates the estimated source signals as well as the course of the ICI index during iteration. As it can be observed, in this experiment the algorithm behaved satisfactorily and converged in a few (< 10) iterations. Also, the result keep stable over time, showing that the algorithm does not exhibit post-separation oscillations.

4 Conclusion

We presented a novel batch-type learning algorithm over the orthogonal group. Its numerical performance in independent component analysis was illustrated. The achieved results suggest further theoretical endeavors on such class of learning algorithms.

References

[1] A. Chicocki and S.-i. Amari, *Adaptive Blind Signal and Image Processing*, J. Wiley & Sons, 2002

[2] A. Cichocki, R. Umbehauen and E. Rummert, *Robust learning for blind separation of signals*, Electronics Letters, Vol. 30, No. 17, pp. 1386 – 1387, Aug. 1994

[3] S. Fiori, *A Theory for Learning by Weight Flow on Stiefel-Grassman Manifold*, Neural Computation, Vol. 13, No. 7, pp. 1625 – 1647, July 2001

[4] S. Fiori, *A Theory for Learning Based on Rigid Bodies Dynamics*, IEEE Trans. on Neural Networks, Vol. 13, No. 3, pp. 521 – 531, May 2002

[5] G.H. Goulb and C.F. van Loan, *Matrix Computations*, The John Hopkins University Press, Third Edition, 1996

[6] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001

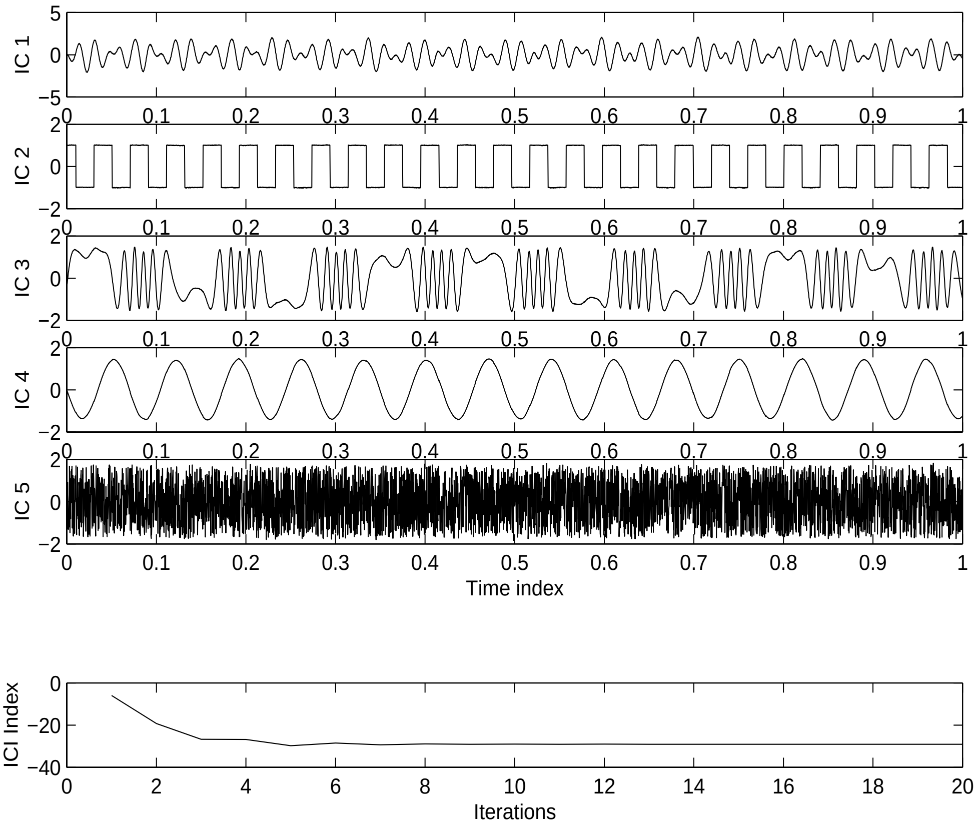

Figure 1: Estimated source signals (first five rows) and the course of the ICI index during iteration (last row).