ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

Semi-Blind Approaches for Source Separation and Independent component Analysis

Massoud Babaie-Zadeh[1] and Christian Jutten[2] *

1- Sharif University of Technology - Electrical Engineering Department
Azadi avenue, Tehran - Iran

2- INPG - LIS (UMR CNRS 5083)
46, avenue Félix Viallet, 38031 Grenoble Cedex - France

Abstract. This paper is a survey of semi-blind source separation approaches. Since Gaussian iid signals are not separable, simplest priors suggest to assume non Gaussian iid signals, or Gaussian non iid signals. Other priors can also been used, for instance discrete or bounded sources, positivity, etc. Although providing a generic framework for semi-blind source separation, Sparse Component Analysis and Bayesian ICA will just sketched in this paper, since two other survey papers develop in depth these approaches.

1 Introduction

Source separation consists in retrieving unknown signals, $\mathbf{s} = (s_1(t), \ldots, s_n(t))^T$, which are observed through unknown mixture of them. Denoting the observations $\mathbf{x}(t) = (x_1(t), \ldots, x_p(t))^T$, one can write :

$$\mathbf{x}(t) = \mathcal{F}(\mathbf{s}(t)), \tag{1.1}$$

where $\mathcal{F}(.)$ denotes the unknown mixture, a function from $\mathbb{R}^n$ to $\mathbb{R}^p$. If the number of observations p is greater or equal to the number of sources, n, the main idea for separating the sources is to estimate a transform $\mathcal{G}(.)$ which inverses the mixture $\mathcal{F}(.)$, and, without extra effort, provides estimates of the unknown sources.

Of course, without other assumptions, this problem cannot be solved. Basically, it is necessary to have priors about

- the nature of the mixtures : it is very important to chose a separating transform $\mathcal{G}(.)$ suited to the mixture transform $\mathcal{F}(.)$,
- the sources : sources properties - even weak - are necessary for driving the $\mathcal{G}(.)$ estimation.

Because of the very weak assumptions, the problem is referred as blind source separation (BSS), and the method based on the property of source independence

*This work has been partially funded by Sharif University of Technology, and by Center for International Research and Collaboration (ISMO) through the France-Iran integrated action program (Gundi-Shapour).

has been called independent component analysis (ICA) [1, 2].

In fact, one often has priors on signals. A natural idea is to add these priors in the model, for simplifying or improving the separation methods. This paper, with its two companions, the Gribonval's paper on sparse component analysis (SCA) [3] and the Mohamad-Djafari's paper on Bayesian ICA [4], constitute a review of semi-blind methods in which various priors are used. For this reason, Sparse and Bayesian ICA, although related to generic approaches for, will not be developed in this paper.

This paper is organized as follows. In Section 2, we recall usual assumptions in the so-called blind source separation. Section 3 is devoted to Gaussian non iid sources. In Section 4, we show that prior like discrete-valued or bounded sources leads to simple geometrical algorithms. In Section 5, we suggest other priors, like video cue for enhancing speech separation, or positivity. Section 6 is a short conclusion, which briefly presents the papers of the special session on semi-blind source separation (SBSS) of the conference ESANN 2006.

2 Blind source separation

Source separation methods have been developed intensively for linear mixtures, instantaneous as well as convolutive, and more recently by a few researchers for nonlinear mixtures. In the most general case, the only assumption done on the sources is that they are statistically independent. From Darmois's result [5], one deduces that this problem has no solution for mutually independent sources which are Gaussian, with (temporally) independent and identically distributed (iid) samples. Then, since the Gaussian iid model has no solution, one must add priors, which are threefold [6]:

- Non Gaussian iid,
- Gaussian but non temporally independent (first *i* of *iid*), *i.e.* temporally correlated,
- Gaussian, but non identically distributed (*id* of *iid*), *i.e.* non stationary sources.

Initially, even if it was not clearly stated [7], the problem has been related to the non Gaussian iid model, and has been refered as blind source separation (BSS). The non Gaussian property appears clearly in the Comon's theorem [2] for linear mixtures.

Theorem 2.1 *Let* $\mathbf{x} = \mathbf{As}$ *be a p-dimension regular mixture of mutually independent random variables, with at most one Gaussian,* $\mathbf{y} = \mathbf{Bx}$ *has mutually independent components iff* $\mathbf{BA} = \mathbf{PD}$*, where* $\mathbf{P}$ *and* $\mathbf{D}$ *are permutation and diagonal matrices, respectively.*

This theorem is only based on the independence of random variables. The independence criterion involves (explicitly or implicitly) higher order (than 2) statistics, but does not take into account the order of samples. It means that the iid assumption is not required, it is just a default assumption: consequently, it works for iid as well as for not iid sources, but it cannot work for Gaussian sources.

The different (blind) ICA algorithms then use different ideas for achieving the required higher order independences. Some of different ideas used for ICA are:

- Non-linear decorrelation [8, 9, 10, 11].
- Methods based on Higher (than 2) Order Statistics (HOS)[12, 13]).
- Cancellation the mutual information (MI) of the outputs [2, 14, 15, 16, 17]. This approach may be shown to provide asymptotically a Maximum Likelihood (ML) estimation of the source signals [18].
- Algorithms based on non-Gaussianity [19, 20]. These algorithms may be shown to have a close correspondence to the algorithms based on MI minimization (refer to section 10.2 of [21]).

More complicated mixing systems have also been studied in the literature. For example, in (linear) convolutive mixtures, the mixing model is $\mathbf{x}(n) = \mathbf{B}_0\mathbf{x}(n)+\mathbf{B}_1\mathbf{x}(n-1)+\cdots+\mathbf{B}_p\mathbf{x}(n-p) = [\mathbf{B}(z)]\mathbf{x}(n)$, which has been shown [22] to be separable. Non-linear mixtures are not in general separable (refer to chapter 3 of [23]). A practically important case of non-linear mixtures is Post Non-Linear (PNL) mixtures [24, 23, 25], in which a linear mixture is followed by non-linear sensors. It has been shown that PNL mixtures are separable, too [24, 23].

However, if some weak prior information about the source signals is available, then the performance of the source separation algorithms may be significantly improved. Thus, these methods are not 'Blind' but 'Semi-Blind'. In the rest of this paper, some of most frequently used priors have been considered. It should be noted, however, that the 'sparsity prior' and 'Bayesian approaches' are mostly considered in the two companions of this paper [4, 3].

3 Separation of non iid sources

Suppose that we know a priori that the source samples are not iid, *i.e.* if sources are temporally correlated, or non stationary.

3.1 Separation of correlated sources

Several approaches had been proposed for separating correlated sources [26, 27, 28]. Pham and Garat [29] showed that time-correlated Gaussian sources can be separated provided than their spectra are different. In that case, the separation can be achieve by estimating a separation matrix $\mathbf{B}$ which minimizes the criterion

$$C(\mathbf{B}) = \sum_{l=1}^{L} w_l \text{off}(\mathbf{B}\hat{\mathbf{R}}(\tau_l)\mathbf{B}^T), \tag{3.1}$$

where w_l are weighting coefficients, off(.) is a measure of deviation from diagonality, which is positive and vanishes iff (.) is diagonal and which satisfies:

$$\text{off}(\mathbf{R}) = \text{D}(\mathbf{R} \mid \text{diag}\mathbf{R}), \tag{3.2}$$

where $\text{D}(\mathbf{R}_i \mid \mathbf{R}_j)$ denotes the Kullback-Leibler divergence of two zero mean multivariate normal densities, with variance-covariance matrices $\mathbf{R}_i$ and $\mathbf{R}_j$, and diag$\mathbf{R}$ is the diagonal matrice composed by diagonal entries of $\mathbf{R}$ and zeros elsewhere.

The criterion (3.1) involves a set of variance-covariance matrices with various delays τ_l : $\hat{\mathbf{R}}(\tau_l) = \hat{E}[\mathbf{y}(t-\tau_l)\mathbf{y}(t)^T]$, where $\hat{E}[.]$ is estimated using an empirical mean. Basically, minimizing this criterion is equivalent to estimate the separation matrix $\mathbf{B}$ which diagonalizes jointly the set of the variance-covariance matrices. The advantages of this approache are:

- it only requires second-order statistics,
- it can then separate Gaussian sources,
- there exist many very fast and efficient algorithms for jointly diagonalizing matrices [30, 31].

3.2 Separation of nonstationary sources

Source nonstationarity have been first used by Matsuoka *et al.* [32]. More recently, Pham et Cardoso developed a rigourous formalization, and proved that nonstationary Gaussian sources can be separated provided than the variance ratios $\sigma_i^2(t)/\sigma_j^2(t)$ are not constant. In that case, the separation can be achieve by estimating a separation matrix $\mathbf{B}$ which minimizes the criterion

$$C(\mathbf{B}) = \sum_{l=1}^{L} w_l \text{off}(\mathbf{B}\hat{\mathbf{R}}_l\mathbf{B}^T), \tag{3.3}$$

where we use the same notations than in the previous subsection. In Eq. (3.3), matrices $\hat{\mathbf{R}}_l$ are variance-covariance matrices estimated by empirical mean on successive sample blocks T_l. Among a few algorithms, the separation matrix $\mathbf{B}$ can be computed as the matrix which jointly diagonalizes the set of the variance-covariance matrices $\mathbf{R}_l$.

The method has the same advantages than the method exploiting the temporal correlation.