

Blind Source Separation: Biomedical applications

Alexander M. Bronstein¹

Michael M. Bronstein¹

Michael Zibulevsky²

¹Department of Computer Science,

²Department of Electrical Engineering,

Technion–Israel Institute of Technology,

Haifa 32000, Israel.

September 30, 2005

Abstract

Blind source separation (BSS) refers to a wide class of methods in signal and image processing, which extract the underlying sources from a set of mixtures without almost any prior knowledge about the sources nor about the mixing process. In biomedical applications, BSS is used for the analysis of electroencephalogram (EEG), magnetoencephalogram (MEG) and electrocardiogram (ECG) signals and functional magnetic resonance (fMRI) images.

Keywords: blind source separation, blind deconvolution, principal component analysis, independent component analysis, independent factor analysis, sparse component analysis, sparse representation, optimization, statistics, information theory, kurtosis, high-order statistics, cumulant tensor, entropy, negentropy, mutual information, maximum likelihood, infomax, short-time Fourier transform, time-frequency analysis, electroencephalography, magnetoencephalography, functional magnetic resonance imaging, electrocardiography

Cross References (mentioned in the article) electroencephalography, magnetoencephalography,

functional magnetic resonance imaging, signal processing, electrocardiography, fetal ECG, principal component analysis, probability function, independent component analysis, time-frequency analysis, deconvolution in biomedical signal processing, higher-order statistics, entropy

1 Introduction

The term *blind source separation* (BSS) refers to a wide class of problems in signal and image processing, where one needs to extract the underlying sources from a set of mixtures. Almost no prior knowledge about the sources, nor about the mixing is known, hence the name *blind*. In practice, the sources can be one-dimensional (e.g. acoustic signals), two-dimensional (images) or three-dimensional (volumetric data). The mixing can be *linear*, or *nonlinear* on one hand, and *instantaneous* or *convolutive*; in the latter case, the problem is referred to as *multichannel blind deconvolution* or *convolutive BSS*. In many medical applications, the instantaneous linear mixing model holds, hence the most common situation is when the mixtures are formed by superposition of sources with different scaling coefficients. These coefficients are usually referred to as *mixing* or *crosstalk coefficients*, and can be arranged into a *mixing (crosstalk) matrix*. The number of mixtures can be smaller, larger or equal to the number of sources.

Independent component analysis (ICA), a method for finding independent components in multi-dimensional statistical data, is usually exploited to solve the BSS problem. One of the fundamental assumptions of this approach is that the sources are *statistically independent*, i.e. the value of any one of the sources gives no information on the values of the other sources. Hence, finding the minimally dependent components by minimizing a certain measure of dependence (usually employing a *numerical optimization* procedure) produces an estimate of the original sources. An emerging class of alternative methods is the *sparse component analysis* (SCA or SPICA), which allows to relax the assumption of statistical independence of the sources and relies on the assumption that sources are *sparse* (i.e. have a small number of non-zero values) or *sparsely-representable* in an appropriate domain (e.g., in the domain of the Fourier transform, wavelet transform, Gabor transform, etc.) Another

important class of approaches to the BSS problem is based on *independent factor analysis* (IFA).

In the context of medical signal and image processing, the BSS problem arises e.g. in analysis of electroencephalogram (EEG), magnetoencephalogram (MEG) and electrocardiogram (ECG) signals and functional magnetic resonance images (fMRI). Recent research has also shown the feasibility of BSS techniques in hyperspectral analysis of tissues.

2 The linear BSS problem

In a typical scenario of an instantaneous linear BSS problem, real signals $s_i(t)$ from N sources are recorded by M sensors; t is a multi-index and refers to data of any dimension. The sensor signals $x_i(t)$ are instantaneous linear combinations of the source signals, and are possibly contaminated by additive *sensor noise* $\xi_i(t)$:

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + \dots + a_{1N}s_N(t) + \xi_1(t) \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \qquad \qquad \vdots \\ x_M(t) &= a_{M1}s_1(t) + \dots + a_{MN}s_N(t) + \xi_M(t) \end{aligned}$$

where a_{ij} are the crosstalk coefficients. In EEG, for example, source signals are resulting from electromagnetic activity of the brain cortex, and sensor signals are the electric potentials measured by the scalp electrodes. The crosstalk coefficients represent the attenuation the source signals undergo during their propagation and are related to the geometry of the head. The noise can result e.g. from electromagnetic disturbances.

In matrix notation, the linear BSS problem has the form

$$\mathbf{x}(t) = \mathbf{A}s(t) + \boldsymbol{\xi}(t),$$

where $s(t)$, $\mathbf{x}(t)$ and $\boldsymbol{\xi}(t)$ are column vectors with values $s_i(t)$, $x_i(t)$ and $\xi_i(t)$, respectively. When

the data is discrete, the problem can be written as

$$X = AS + \varepsilon, \quad (1)$$

where X and S are matrices containing the mixtures and the sources, respectively, as their rows, and ε is the corresponding matrix of noise. If the mixtures are two-dimensional (images), they are parsed into vectors and treated in an essentially similar way.

The BSS problem consists of finding an estimate $Y(X)$ of S , given only the observed (possibly noisy) data X . If no *a priori* information is available, the latter is possible up to an arbitrary permutation and scaling only. In a particular case when the matrix A is square (i.e. $N = M$) and invertible, and under the assumption of zero noise, the problem is equivalent to estimating the *unmixing matrix* $W = (\mathbf{w}_1, \dots, \mathbf{w}_N)^T = EA^{-1}$ such that

$$EY = WX,$$

where E is a permutation and scaling matrix.

3 Independent Component Analysis (ICA)

If the sources s_i are statistically independent, the ICA approach can be used to solve the BSS problem. Statistical independence implies that the joint probability density of the sources $f_{\mathbf{s}}(s_1, \dots, s_N)$ can be factorized into a product of marginal densities,

$$f_{\mathbf{s}}(s_1, \dots, s_N) = \prod_{i=1}^N f_{s_i}(s_i).$$

The assumption of statistical independence of sources is the guiding principle of ICA: the sources can be estimated as a linear transformation $Y = WX$ of the mixtures, yielding the "most independent" components. The sources estimated in this way are termed *independent components* (ICs).

The property of independence is stronger than *uncorrelatedness*, i.e. $\mathbf{E}s_i s_j = \mathbf{E}s_i \mathbf{E}s_j$ for $i \neq j$, where \mathbf{E} denotes expectation. Uncorrelated components can be extracted by means of linear *decorrelation* (sometimes termed as *sphering* or *whitening*) of the data $\mathbf{x}(t)$, i.e. by linearly transforming $\mathbf{x}(t)$ into $\mathbf{y}(t) = P\mathbf{x}(t)$ by a whitening matrix P , such that the *covariance matrix* of $\mathbf{y}(t)$ equals unity:

$$C_{\mathbf{y}\mathbf{y}} = \mathbf{E}(\mathbf{y}(t)\mathbf{y}^T(t)) = PC_{\mathbf{x}\mathbf{x}}P^T = I.$$

The covariance $C_{\mathbf{x}\mathbf{x}}$ is estimated from a finite realization of $\mathbf{x}(t)$. Substituting the eigendecomposition $C_{\mathbf{x}\mathbf{x}} = VDV^T$ (where V is a unitary diagonalizing matrix and D is diagonal), the covariance $C_{\mathbf{y}\mathbf{y}}$ can be expressed as

$$C_{\mathbf{y}\mathbf{y}} = PVC_{\mathbf{x}\mathbf{x}}V^T P^T.$$

The whitening matrix therefore equals $P = D^{-\frac{1}{2}}V^T$. This procedure is a version of *principal component analysis* (PCA) (also called *Hotteling-* or *Karhunen-Loeve transform*), in which the variances are normalized.

Geometrically, decorrelation of the data undoes the mixing up to a rotation matrix, i.e. the unmixing matrix can be found as a product $W = UP$ of the whitening matrix P and some rotation matrix U . When the sources are jointly Gaussian, independence and uncorrelatedness are equivalent. Since such distribution is invariant under rotation, in case of Gaussian sources unmixing is only possibly up to arbitrary rotation (see Figure 1). For this reason, the Gaussian case is usually excluded in the classical ICA model.

More precisely, the classical setting of the (zero-noise) ICA problem is identical to (1) with the addition of the following *identifiability conditions*: (i) at most one of the sources s_i is Gaussian; (ii) the number of mixtures is at least as large as the number of sources, ($M \geq N$); and (iii) the mixing matrix A is of full column rank; (iv) $s_i(t)$ are stationary and ergodic. Typically, a stronger version of these conditions, assuming non-Gaussianity of all the sources and independent identical distribution (i.i.d.) of the values of $s_i(t)$ over t is used.

ICA is usually performed by formulating a criterion of statistical dependence $\varphi(\mathbf{y})$, referred to as *contrast function* or *contrast* (the terms *objective-*, *cost-* and *loss function* are synonymous) and minimizing it – sometimes referred to as *minimum contrast estimation*. The selection of specific contrast and numerical algorithm for its optimization gives rise to different ICA methods.

Typically, decorrelation of the data using PCA is performed as a pre-processing stage. Some methods using *orthogonal contrasts* demand explicitly uncorrelatedness of the data. If the number of sources N equals the number of mixtures M , it is usually convenient to pose the problem as estimating the unmixing matrix W :

$$W = \underset{W}{\operatorname{argmin}} \varphi(W\mathbf{x}).$$

If $M > N$, the first N principal components of X are used as the data.

Depending on whether all the ICs are estimated simultaneously, or one-by-one, *multi-unit* or *one-unit* contrast functions are used, respectively. Another distinction is made between *online* (or *adaptive*) methods that use instantaneous data at each iteration, and *batch* methods that use all the data simultaneously; this terminology is common in the neural networks community, where some fundamental works on ICA came from.

4 Contrast functions

Minimum contrast estimation is a general statistical estimation approach commonly used in ICA. Different multi-unit contrast functions can be derived from different principles such as maximum likelihood, information maximization, etc. It can be shown that in many cases these contrasts are equivalent.

Mutual information: independence can be measured as the *Kullback-Leibler divergence* between

the true density $f_{\mathbf{y}}(\mathbf{y})$ of \mathbf{y} and the product of marginal densities $f_0(\mathbf{y}) = \prod_{i=1}^N f_i(y_i)$

$$\varphi_{MI}(\mathbf{y}) = \int f_{\mathbf{y}}(\mathbf{y}) \log \left(\frac{f_{\mathbf{y}}(\mathbf{y})}{f_0(\mathbf{y})} \right) d\mathbf{y},$$

also known as the *mutual information* of \mathbf{y} . Mutual information reflects the quantity of information shared between the elements of \mathbf{y} ; it is always non-negative and vanishes only if y_i are statistically independent. Mutual information can be also expressed via the *differential entropy* as,

$$\varphi_{MI}(\mathbf{y}) = \sum_{i=1}^N H(y_i) - H(\mathbf{y}) = \sum_{i=1}^N H(w_i^T \mathbf{x}) - H(\mathbf{x}) - \log |\det W|,$$

with w_i denoting the i -th row of W and

$$H(\mathbf{x}) = - \int f_{\mathbf{x}}(\mathbf{x}) \log f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}. \quad (2)$$

If the data is decorrelated, φ_{MI} becomes the orthogonal mutual information contrast:

$$\varphi_{MI}^{\perp}(\mathbf{y}) = \sum_{i=1}^N H(y_i). \quad (3)$$

Likelihood: assuming that $s_i(t), t = 1, \dots, T$ are i.i.d., the samples of the mixtures $x_i(t)$ can be considered as independent measurements of the data. The distribution of $\mathbf{s}(t)$ in this case is given by

$$f(\mathbf{x}|W) = |\det W| \cdot f_{\mathbf{s}}(W\mathbf{x}) = |\det W| \cdot \prod_{i=1}^N f_{s_i}(\mathbf{w}_i^T x_i), \quad (4)$$

where W are the parameters of the model which need to be estimated. The *likelihood* is the probability of the observed data as a function of the parameters of the model:

$$L(X; W) = |\det W| \cdot \prod_{t=1}^T \prod_{i=1}^N f_{s_i}(\mathbf{w}_i^T x_i(t)).$$

Maximizing the likelihood function $L(X, W)$ with respect to W produces an estimate of the unmixing matrix. Usually, it is more convenient to minimize the normalized *minus log-likelihood function*

$$\ell(X; W) = -\frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N \log f_{s_i}(\mathbf{w}_i^T x_i(t)) - \log |\det W|.$$

The normalized minus-log-likelihood $\ell(X; W)$ can be seen as an approximation (up to a constant) of the Kullback-Leibler divergence

$$d_{KL}(f_{W\mathbf{x}}, f_{\mathbf{s}}) = \int f_{W\mathbf{x}}(\mathbf{y}) \log \left(\frac{f_{W\mathbf{x}}(\mathbf{y})}{f_{\mathbf{s}}(\mathbf{y})} \right) d\mathbf{y},$$

on a finite sample of size T , i.e.

$$\ell(X; W) \xrightarrow{T \rightarrow \infty} d_{KL}(f_{W\mathbf{x}}, f_{\mathbf{s}}) + \text{const.}$$

Hence, the *maximum likelihood (ML)* estimation is associated with the contrast function

$$\varphi_{ML}(\mathbf{y}) = d_{KL}(f_{W\mathbf{x}}, f_{\mathbf{s}}). \quad (5)$$

The probability densities f_{s_i} do not need to be necessarily known exactly; when approximate f_{s_i} are used, the estimation is termed as *quasi maximum likelihood (QML)* [1]. Often, if the sources are known to be sub-Gaussian or super-Gaussian, approximate f_{s_i} can be selected accordingly.

Infomax: estimated ICs can be considered as outputs of a single-layer multi-input multi-output neural network, with $x(t)$ serving as inputs and outputs given by $y_i(t) = g_i(\mathbf{w}_i^T \mathbf{x}(t))$, where g_i are non-linear scalar functions. When g_i are selected as the cumulative distribution functions F_{s_i} of the sources (or in other words, $g_i' = f_{s_i}$), maximizing the entropy of the outputs

$$\varphi_{IM}(\mathbf{y}) = H(g_1(\mathbf{w}_1^T \mathbf{x}(t)), \dots, g_N(\mathbf{w}_N^T \mathbf{x}(t))), \quad (6)$$

is equivalent to maximum likelihood estimation. This approach is known as *network entropy- or information maximization (infomax)* and was introduced by T. Bell and L. Sejnowski [2]. The intuition of infomax is the following: $\mathbf{g}(\mathbf{s}) = (g_1(\mathbf{s}), \dots, g_N(\mathbf{s}))$ is uniformly distributed on $[0, 1]^N$ and possesses the largest entropy among all the distributions on $[0, 1]^N$. Therefore, $\mathbf{g}(\mathbf{y}) = \mathbf{g}(W\mathbf{A}\mathbf{s})$ has the largest entropy when $W\mathbf{A} = I$.

Negentropy: Another non-Gaussianity maximization approach stems from a fundamental result of information theory, according to which a Gaussian variable possesses the largest entropy among all random variables with equal variance. A normalized version of entropy $J(\mathbf{y}) = H_0 - H(\mathbf{y})$, where H_0 is the entropy of a Gaussian vector possessing the same covariance as \mathbf{y} , is called *negentropy*. Negentropy is non-negative and vanishes only if \mathbf{y} has a Gaussian distribution. Minimization of the contrast function

$$\varphi_{NE}(\mathbf{y}) = - \sum_i J(y_i) \quad (7)$$

leads to ICs with maximum non-Gaussianity. Minimization of $\varphi_{NE}(\mathbf{y})$ can be approximated by maximization the sum of squared kurtoses $\sum_i \kappa_i^2$. A. Hyvärinen *et al.* proposed to use the more general approximation of the form

$$J(\mathbf{y}) \approx (\mathbf{E}h(\mathbf{y}) - \mathbf{E}h(\mathbf{v}))^2, \quad (8)$$

where \mathbf{v} is a normal random variable and h can be virtually any non-quadratic function.

High-order cumulants: another way to derive contrast functions is by using high-order statistics. Such contrasts can be considered approximations of the information theoretic contrasts. High-order statistical information is expressed via *cumulant tensors*; for zero-mean variables x_i, x_j, x_k, x_l the

second- and the fourth-order cumulants are:

$$\begin{aligned}\mathbf{C}_{ij}(\mathbf{x}) &= \mathbf{E}(x_i x_j), \\ \mathbf{C}_{ijkl}(\mathbf{x}) &= \mathbf{E}(x_i x_j x_k x_l) - \mathbf{E}(x_i x_j) \mathbf{E}(x_k x_l) - \mathbf{E}(x_i x_k) \mathbf{E}(x_j x_l) - \mathbf{E}(x_i x_l) \mathbf{E}(x_j x_k).\end{aligned}$$

The auto-cumulants $\mathbf{C}_{ii}(\mathbf{x}) = \mathbf{E}x_i^2 = \sigma_i^2(\mathbf{x})$ and $\mathbf{C}_{iiii}(\mathbf{x}) = \mathbf{E}x_i^4 - 3\mathbf{E}^2x_i^2 = \kappa_i(\mathbf{x})$ are the *variance* and the *kurtosis* of the i -th source, respectively. For statistically independent sources, the cross-cumulants vanish, i.e. $\mathbf{C}_{ij}(\mathbf{s}) = \sigma_i^2 \delta_{ij}$, $\mathbf{C}_{ijkl}(\mathbf{s}) = k_i \delta_{ijkl}$; where δ denotes the Kronecker symbol.

If the data is decorrelated, the mutual information contrast φ_{MI} can be approximated as:

$$\varphi_{KUR}^\perp(\mathbf{y}) = \sum_{ijkl \neq iiii} \mathbf{C}_{ijkl}^2(\mathbf{y}) = - \sum_i \mathbf{C}_{iiii}^2(\mathbf{y}) + \text{const.} \quad (9)$$

This orthogonal contrast was used already in the early works on ICA [3]. The intuition behind this contrast is based on the *non-Gaussianity maximization* principle. According to the central limit theorem, the distribution of a sum of non-Gaussian sources is closer to Gaussian distribution than the distributions of the sources. Gaussian distribution has zero kurtosis; distributions with positive kurtosis are termed as *super-Gaussian* and distributions with negative kurtosis are termed as *sub-Gaussian*. The first class includes, e.g. sparse distributions. Non-Gaussianity of y_i can be measured as the absolute or the squared value of the kurtosis $\kappa_i(\mathbf{y})$. Hence, the minimizer of $\varphi_{KUR}^\perp(\mathbf{y})$ corresponds to the most non-Gaussian components.

Tensorial methods rely on eigendecomposition of the fourth-order cumulant. A fundamental result from statistics states that every matrix $\mathbf{w}_i \mathbf{w}_i^T$, where \mathbf{w}_i^T is the i -th row of the mixing matrix W , is an eigenmatrix of the cumulant tensor of the whitened data \mathbf{y} , with the corresponding eigenvalue κ_i , i.e.,

$$\mathbf{C}(\mathbf{w} \mathbf{w}^T) = \kappa_i \cdot \mathbf{w} \mathbf{w}^T,$$

where

$$\mathbf{C}(\mathbf{w}\mathbf{w}^T)_{ij} = \sum_{k,l} w_k w_l \mathbf{C}_{ijkl}$$

is the product of the tensor \mathbf{C} with the matrix $\mathbf{w}\mathbf{w}^T$. This gives a direct method for estimating the mixing matrix: the observation \mathbf{y} is whitened, then the fourth-order cumulant tensor \mathbf{C} is estimated and its eigendecomposition is computed. Estimate of the mixing matrix is constructed from the eigenmatrices of \mathbf{C} .

Joint approximate diagonalization of eigenmatrices (JADE): is a particular case of tensorial methods introduced by J.-F. Cardoso and A. Souloumiac [4]. The tensor \mathbf{C} has N^2 orthonormal eigenmatrices D_m and the corresponding eigenvalues λ_m , of which only N are non-zero. If the data is decorrelated, the mixing matrix is unitary (a rotation matrix), and diagonalizes the eigenmatrices of \mathbf{C} , i.e.: $B_m = W D_m W^T$, with D_m a diagonal matrix. Thus, the unmixing matrix can be found by joint diagonalization of the N eigenmatrices B_m , which can be carried out efficiently using Jacobi iterations. The orthogonal contrast associated with the JADE algorithm is the joint diagonalization criterion:

$$\varphi_{JADE}^\perp(\mathbf{y}) = \sum_{ijkl \neq ijkk} \mathbf{C}_{ijkl}^2(\mathbf{y}). \quad (10)$$

5 Numerical ICA algorithms

Independent component analysis is often carried out by means of minimization of the contrast function φ with respect to the argument W . This task is performed by an iterative numerical procedure, called *optimization algorithm*. An generic optimization algorithm produces a sequence $W^{(0)}, W^{(1)}, \dots, W^{(K)}$ of estimates of the function minimizer, on which the function approaches the optimal value. The iterations are stopped when e.g. the norm of the gradient $\|\nabla\varphi(W^{(k)}\mathbf{y})\|$ is sufficiently small. Classical optimization algorithms produce $W^{(k+1)}$ by additive update, i.e.

by making a step in direction $V^{(k)}$ on every iteration: $W^{(k+1)} = W^{(k)} + \alpha^{(k)}V^{(k)}$. Typically, the choice of the direction depend on the gradient $\nabla\varphi$ or the Hessian $\nabla^2\varphi$ of φ at each iteration. For example, choosing $V^{(k)} = -\nabla\varphi(W^{(k)}\mathbf{y})$ is known as *steepest* or *gradient descent*, and $V^{(k)} = -(\nabla^2\varphi(W^{(k)}\mathbf{y}))^{-1}\nabla\varphi(W^{(k)}\mathbf{y})$ is known as *Newton step*.

Relative optimization algorithms is a class of optimization algorithm that rely on the fact that the mixing matrices form a multiplicative group. The latter implies that starting from the observed mixtures, one can iteratively improve the source estimate (in terms of some contrast function) by finding an estimate of the unmixing matrix leading to a decrease of the contrast function, and using the obtained source estimate as the observation at the next iteration. A general relative optimization algorithm has the following structure [5]:

1. Start with an initial source estimate $Y^{(0)}$.
2. For $k = 0, 1, 2, \dots$, until convergence
 - (a) Starting from $V^{(k)} = I$ and using one or more steps of some unconstrained minimization algorithm, find such a matrix $V^{(k)}$ that decreases the contrast function $\varphi(V^{(k)}Y^{(k)})$.
 - (b) Update the source estimate: $Y^{(k+1)} = V^{(k)}Y^{(k)}$.
3. End of loop

The output of the algorithm is the source estimate $Y = Y^{(K)}$ minimizing the contrast φ . The unmixing matrix is obtained as the product $W = V^{(0)} \cdot V^{(1)} \cdot \dots \cdot V^{(K)}$.

Different algorithms can be obtained by different choices of the unconstrained minimization algorithm used in Step (a). The *relative gradient* (also referred to as the *natural gradient*) algorithm, introduced by A. Cichocki *et al.* [6], J.-F. Cardoso and B. Laheld [7] and S.-I. Amari *et al.* [8], although derived from different considerations, can be obtained by using a steepest descent iteration at Step (a). The *relative Newton* algorithm [1, 5]) is obtained by using a single Newton iteration at Step (a). An efficient block-coordinate version of this algorithm was proposed by Bronstein *et al.* [9].

Relative optimization techniques can be used in both batch and online modes. Extensions for blind deconvolution are also available [10, 11].

Fixed point algorithms were introduced by A. Hyvärinen and E. Oja [12] and are commonly known under the name of *Fast ICA*. This class of algorithms is based on fixed point iterations, a method known to have very fast convergence. The one-unit Fast ICA algorithm has the following form:

- Center and whiten the data \mathbf{y} .
- Start with some estimate $\mathbf{w}^{(0)}$ of a row vector of W .
- For $k = 1, 2, \dots$, until convergence
 - Update $\mathbf{w}^{(k)} \leftarrow \mathbf{E} \left\{ \mathbf{y} \varphi' \left((\mathbf{w}^{(k-1)})^T \mathbf{y} \right) \right\} - \mathbf{E} \left\{ \varphi'' \left((\mathbf{w}^{(k-1)})^T \mathbf{y} \right) \right\} \cdot \mathbf{w}^{(k-1)}$.
 - Normalize: $\mathbf{w}^{(k)} \leftarrow \mathbf{w}^{(k)} / \|\mathbf{w}^{(k)}\|$.
- End of loop

In practice, the expectations in Step 4 are replaced by empirical averages and the separation vectors \mathbf{w} are orthogonalized. φ can be virtually any twice-differentiable function and does not have to correspond to any statistically or information theoretically based contrast. Different choices of φ result in different algorithms suitable for different classes of sources. Fast ICA is a batch algorithm and generally used for estimating the ICs one-by-one. Extensions for the multi-unit case are available.

6 Sparse Component Analysis (SCA)

The fundamental assumption of statistical independence of source used in ICA can be relaxed and replaced by the assumption of their *sparsity*. This gives rise to the *sparse component analysis (SCA)*, an alternative class of methods for BSS [13, 14]. Sparsity implies that most of the values of the source are zero or near-zero. For simplicity, consider the problem of two sources and two mixtures

($M = N = 2$). The observed mixture $\mathbf{x}(t)$ is a set of points in the plane (*scatter plot*), given by

$$\mathbf{x}(t) = \mathbf{a}_1 \cdot \mathbf{s}_1(t) + \mathbf{a}_2 \cdot \mathbf{s}_2(t),$$

where $\mathbf{a}_1, \mathbf{a}_2$ are the column vectors of the mixing matrix A . If the sources are sparse, most of the instantaneous mixture values are contributed by a single "active" source only (either $\mathbf{s}_1(t)$ or $\mathbf{s}_2(t)$). Statistical independence is not necessary for this to hold; it is enough that in the realization of the sources the non-zero values do not overlap in most of their occurrences. Consequently, $\mathbf{x}(t)$ consists mainly of points lying along directions \mathbf{a}_1 and \mathbf{a}_2 (Figure 3). Recovering these directions allows to estimate the mixing matrix. This BSS approach is termed as *geometric separation*. The applicability of geometric separation is practically limited to problems with number of mixtures $M = 2$ or 3. However, the advantage of this method compared to ICA is that it allows to handle BSS problems with number of sources $N > M$.

If the sources are not originally sparse, in many cases they can be *sparsified*, i.e. brought to a sparse representation by application of a linear *sparsifying transformation* given by a $T_1 \times T$ matrix Φ :

$$S' = S\Phi^T,$$

where T_1 is the dimension of the resulting representation, and superscript $()^T$ denotes matrix transposition. The transformation Φ is not necessarily invertible. Since the mixing is linear, applying Φ on the mixtures is equivalent to mixing the sparsified sources:

$$X' = X\Phi^T = (AS)\Phi^T = A(S\Phi^T) = AS'.$$

Unmixing is performed by first estimating the unmixing matrix W given the sparsified mixtures X' , and then applying it to the original mixtures X . The sparsifying transformation is generally dependent on the source types. A possible selection of Φ is e.g. the *short-time Fourier transform* (STFT) for

acoustic signals [15] and a discrete derivative for images [16]. A richer family of transformations is obtained by using multi-resolution representations such as the *wavelet* or the *wavelet-packet* (WP) transforms.

Sparsification can be employed together with QML estimation in cases when $M = N$. QML methods require that the distribution of sources is modelled at least approximately. In some cases, it is very difficult or practically impossible to write the source distribution in an analytical form. In addition, many distributions result in non-convex objective functions which are problematic for optimization. Instead, the sources are assumed sparse (with the absolute value or its smoothed version used as a model of $-\log f_{s_i}(s)$) and an appropriate sparsification transformation is applied on the mixtures, making the problem equivalent to separation of sparse sources. This usually outperforms QML estimation without sparsification.

7 Independent Factor Analysis (IFA)

One of the weaknesses of ICA algorithms is the fact they do not incorporate noise into the model. For this reason, the applicability of ICA is limited to low- to medium-noise cases. SCA is more robust to noise in some cases, however, it also ignores the explicit noise model. As a possible remedy, H. Attias [17] proposed the *independent factor analysis* (IFA), which can be viewed as a generalization of ICA, PCA and ordinary factor analysis.

Similarly to ICA, the IFA framework assumes instantaneous linear mixing model. Observations $\mathbf{x}(t)$ are assumed to be contaminated by zero-mean Gaussian noise $\boldsymbol{\xi}(t)$ with some covariance matrix Λ . Both Λ and the mixing matrix A are unknown. The sources $s(t)$ are considered mutually statistically independent and their probability density functions are usually modelled as a *mixture of Gaussians* (MOG)

$$f_{s_i}(s_i|\boldsymbol{\theta}_i) = \sum_{k=1}^{n_i} z_{ik} f_G(s_i - \mu_{ik}, \sigma_{ik}^2),$$

where $f_G(s - \mu, \sigma^2)$ is the Gaussian probability density function with mean μ and variance σ^2 ,

where $a_{ij}(t)$ are some unknown filters. This problem is called *multichannel blind deconvolution* (BD). As in the BSS case, the sources and the filters can be one-, two- or three-dimensional. It is common to distinguish between the following cases: $M, N > 1$ - *multiple-input multiple-output* (MIMO), which is the most general setting; $M > 1, N = 1$ - *single-input multiple-output* (SIMO); and $M = N = 1$ - *single-input single-output* (SISO). In image processing, the latter case is sometimes simply referred to as *blind deconvolution*.

A particular setting of the MIMO BD problem is a generalization of the instantaneous (delay-less) linear BSS problem to the cases when the signal propagation velocity is low. The filters model the delays in signal propagation. Such a situation is common e.g. when the sources are acoustic signals. Often, in order to account for acoustic effects like reverberation, the filters must be more complicated than just delays.

The SISO BD problem arises in optical imaging applications, where the source image is acquired through scattering medium. The action of the medium can be modelled as a *linear shift-invariant* (LSI) system and described by convolution with some two-dimensional filter (usually referred to as the *point spread function* or PSF). The source reconstruction requires to undo the effect of the convolution with the PSF. If more than one observation of the same source is available degraded by different PSFs, the problem is formulated as SIMO BD.

Some theoretical methods and numerical algorithms used for BSS can be generalized to BD problems. For example, ML and QML estimators can be formulated in a way similar to the BSS case. An emerging field is the extension of SCA approaches to BD [11, 10]. In some cases, the MIMO BD problem can be posed as an instantaneous BSS problem. For example, in case of acoustic signals, P. Smaragdis proposed the *time-frequency domain ICA* approach [18]. The sensor signals are transformed into the STFT domain

$$x_i(t) \xrightarrow{STFT} X_i(t, \omega),$$

where the convolutive mixtures are translated into instantaneous ones, with a frequency-dependent

mixing matrix:

$$X_i(t, \omega) = A_{i1}(\omega)S_1(t, \omega) + \dots + A_{iN}(\omega)S_N(t, \omega).$$

Assuming that in narrow frequency subbands the filters are approximately constant (i.e. that A is constant), the BD problem can be formulated as a set of instantaneous BSS problems in each subband. The difficulty arising in this method is the permutation and scale ambiguity, which must be resolved before the separated signals in each subband are merged. Usually, some additional information such as the directivity pattern is necessary for this purpose [19].

Nonlinear mixture is a generalization of the linear mixing model. The mixing matrix A is replaced by an invertible nonlinear mixing function \mathcal{A} from \mathbf{R}^N to \mathbf{R}^M . The sensor signals in this case are given by

$$\mathbf{x}(t) = \mathcal{A}(\mathbf{s}(t)).$$

A particular case when the nonlinear operator is applied in a component-wise manner after linear mixing

$$\mathbf{x}(t) = \mathcal{A}(A \cdot \mathbf{s}(t)),$$

is referred to as *post-nonlinear* BSS. Typically, the nonlinear BSS problem is significantly harder than the linear BSS. Several approaches, e.g. based on correlation maximization and kernel learning [20] were proposed, yet, the nonlinear BSS problem is still an open research field.

9 Applications

BSS techniques have been successfully employed in biomedicine, e.g. for the analysis of EEG, MEG, ECG and fMRI data. In these applications, the linear mixture assumption is usually justified by the physical principles of signal formation, and high signal propagation velocity allows to use the

instantaneous mixture model [21]. Otherwise, nonlinear BSS or BD methods are used.

Electroencephalography (EEG): The brain cortex can be thought of as a field of K tiny sources, which in turn are modelled as current dipoles. The j -th dipole is characterized by the location vector \mathbf{r}_j and the dipole moment vector \mathbf{q}_j . The electromagnetic field produced by the neural activity determines the potential on the scalp surface, sampled at a set of M sensors. Denote the i -th sensor location by \mathbf{r}'_i , and the potential it measures by v_i . The mapping from the neural current sources to the measured scalp potentials $\{\mathbf{q}_j, \mathbf{r}_j\} \rightarrow \{v_i, \mathbf{r}'_i\}$ is usually referred to as the *forward model*, and the measured potentials as the *forward field*. The forward model is linear in the dipole moment \mathbf{q} . Since the propagation velocity of electromagnetic waves is very large compared to head dimensions, zero propagation time can be assumed. Hence, one can express the forward field at sensor j due to dipole i as the inner product $v_j^i = \mathbf{g}_{ij}^T \mathbf{q}_i$, where $\mathbf{g}_{ij} = \mathbf{g}(r_j, r'_i)$ is the *vector kernel* or the *lead field* depending on the geometry and the electromagnetic properties of the head. By electromagnetic superposition, $v_j = \sum_i v_j^i$. Assuming T time samples of $\mathbf{v}(t) = (v_1(t), \dots, v_M(t))^T$ are observed, the entire spatio-temporal forward model can be expressed in matrix form as

$$\begin{pmatrix} v_1(t_1) & \dots & v_1(t_T) \\ \vdots & \ddots & \vdots \\ v_M(t_1) & \dots & v_M(t_T) \end{pmatrix} = \begin{pmatrix} \mathbf{g}_{11}^T & \dots & \mathbf{g}_{1K}^T \\ \vdots & \ddots & \vdots \\ \mathbf{g}_{M1}^T & \dots & \mathbf{g}_{MK}^T \end{pmatrix} \cdot \begin{pmatrix} \mathbf{q}_1(t_1) & \dots & \mathbf{q}_1(t_T) \\ \vdots & \ddots & \vdots \\ \mathbf{q}_K(t_1) & \dots & \mathbf{q}_K(t_T) \end{pmatrix},$$

or $V = GQ$, where V is an $M \times T$ matrix of the measured forward field, G is the $M \times 3K$ *gain matrix*, which is assumed to be fixed in time, and Q is a $3K \times T$ matrix of current sources.

Typically, the number of sensors ranges from tens to hundreds, and the number of dipoles K in a faithful forward model is larger by several orders of magnitude. As consequence, the *inverse problem*, i.e. the problem of determining the entire set of K current sources from the measured forward field is underdetermined; however, many types of brain activity can be modelled as the forward field from sets of spatially fixed (either localized or widely-spread around the cortical surface) concurrently acting dipoles activated by temporally independent sources. Makeig *et al.* were the pioneers in using ICA for separation of such sources [22]. ICA may determine what temporally independent activations

compose the collected scalp recordings without specifying directly where in the brain these activations arise. ICA has been shown particularly efficient for analysis of *event-related potential* (ERP) data [23] and for removing encephalographic artifacts [24].

Magnetoencephalography (MEG): Similarly to EEG, the forward model in MEG is also essentially linear. The sensors measure the vector of the magnetic field \mathbf{b}_j around the scalp. The forward field at sensor j due to dipole i can be expressed as $\mathbf{b}_j^i = G_{ij} \mathbf{q}_j$, where $G_{ij} = G(r_j, r_i')$ is the *matrix kernel* depending on the geometry and the electromagnetical properties of the head. BSS can be used for separation of independent temporal components in the same way it is used in EEG [21].

Electrocardiography (ECG): The mechanical action of the heart is initiated by a quasi-periodic electrical stimulus, which causes an electrical current to propagate through the body tissues and results in potential differences. The potential differences measured by electrodes on the skin (*cutaneous recording*) as a function of time is termed *electrocardiogram* (ECG). The measured ECG signal can be considered as a superposition of several independent processes, resulting, for example, from *electromyographic* activity (electrical potentials generated by muscles), 50Hz or 60Hz net interferences, or the electrical activity of the fetal heart (FECG). The latter contains important indications about the fetus health.

The transfer from bioelectrical sources to the electrodes can be considered linear and delay-less, due to the high propagation velocity of electromagnetic waves. Cutaneous ECG voltage recordings, $v_1(t), \dots, v_M(t)$, measured at M electrodes on the skin are given as a superposition of N electrical sources $s_1(t), \dots, s_N(t)$:

$$\begin{pmatrix} v_1(t) \\ \vdots \\ v_M(t) \end{pmatrix} = \begin{pmatrix} g_{11} & \dots & g_{1K} \\ \vdots & \ddots & \vdots \\ g_{M1} & \dots & g_{MK} \end{pmatrix} \cdot \begin{pmatrix} s_1(t) \\ \vdots \\ s_N(t) \end{pmatrix} + \begin{pmatrix} \xi_1(t) \\ \vdots \\ \xi_N(t) \end{pmatrix},$$

where $\xi(t)$ stands for additive noise, and g_{ij} are the transfer coefficients (conductivity) from source j to electrode i . BSS method have been successfully used for separation of interferences in ECG data. A particular success was demonstrated in separation of fetal ECG from mother's ECG [25].

Functional Magnetic Resonance Imaging (fMRI): The principle of fMRI is based on different magnetic properties of oxygenated and deoxygenated hemoglobin, which allows to obtain a *blood oxygenation level dependent* (BOLD) signal. The observed spatio-temporal signal $q(\mathbf{r}, t)$ of magnetic induction can be considered as a superposition of N *spatially independent* components, each associated with a unique time course $\beta_k(t)$ and a spatial map $s_k(\mathbf{r})$. Each source represent the *loci* of concurrent neural activity and can be either *task-related* or *non task-related* (e.g., physiological pulsations, head movements, background brain activity, etc). The spatial map corresponding to each source determines its influence in each volume element (*voxel*), and is assumed to be fixed in time. Spatial maps can be overlapping.

ICA has been successfully used to separate either the independent spatial sources [26], or the independent time courses. These techniques are usually known as *spatial* and *temporal* ICA, respectively. In the spatial approach, fMRI data $q(\mathbf{r}, t_1), \dots, q(\mathbf{r}, t_T)$ acquired at T different times can be considered as a superposition of N independent source images $s_1(\mathbf{r}), \dots, s_N(\mathbf{r})$, mixed with different contributions:

$$\begin{pmatrix} q(\mathbf{r}, t_1) \\ \vdots \\ q(\mathbf{r}, t_T) \end{pmatrix} = \begin{pmatrix} \beta_1(t_1) & \dots & \beta_N(t_1) \\ \vdots & \ddots & \vdots \\ \beta_1(t_T) & \dots & \beta_N(t_T) \end{pmatrix} \cdot \begin{pmatrix} s_1(\mathbf{r}) \\ \vdots \\ s_N(\mathbf{r}) \end{pmatrix} + \begin{pmatrix} \xi(\mathbf{r}, t_1) \\ \vdots \\ \xi(\mathbf{r}, t_T) \end{pmatrix},$$

where $\xi(\mathbf{r}, t)$ stands for additive noise. In the temporal approach the observed data is considered as K time signals $q(\mathbf{r}_1, t), \dots, q(\mathbf{r}_K, t)$, consisting of linear mixtures of N independent time courses $\beta_1(t), \dots, \beta_N(t)$:

$$\begin{pmatrix} q(\mathbf{r}_1, t) \\ \vdots \\ q(\mathbf{r}_K, t) \end{pmatrix} = \begin{pmatrix} s_1(\mathbf{r}_1) & \dots & s_N(\mathbf{r}_1) \\ \vdots & \ddots & \vdots \\ s_1(\mathbf{r}_K) & \dots & s_N(\mathbf{r}_K) \end{pmatrix} \cdot \begin{pmatrix} \beta_1(t) \\ \vdots \\ \beta_N(t) \end{pmatrix} + \begin{pmatrix} \xi(\mathbf{r}_1, t) \\ \vdots \\ \xi(\mathbf{r}_K, t) \end{pmatrix},$$

where K stands for the number of voxels. Combined *spatio-temporal* techniques maximize some measure of independence over space and time simultaneously, without necessarily achieving inde-

pendence over either of them separately [27]. The main advantage of BSS techniques over other fMRI analysis tools is that there is no need to assume any *a priori* information about the time course of processes contributing to the measured signals.

Suggested Reading

- J. Karhunen A. Hyvärinen and E. Oja. *Independent Component Analysis*. John Wiley and Sons, 2001. – *A comprehensive introduction to ICA. The book includes the fundamental mathematical background needed to understand and utilize it. It offers a general overview of the basics of ICA, important solutions and algorithms, and in-depth coverage of different applications.*
- S. J. Roberts and R. M. Everson, editors. *Independent Components Analysis: Principles and Practice*. Cambridge University Press, 2001. – *A self-contained book built as a structured series of edited papers by leading researchers in the field, including an extensive introduction to ICA. The major theoretical bases are reviewed from a modern perspective, current developments such as the SCA paradigm are surveyed and many case studies of applications, including biomedical ones, are described in detail.*
- J.-F. Cardoso. Blind signal separation: statistical principles. *Proc. IEEE*, 9(10):2009–2025, October 1998. [Available Online]. <http://www.tsi.enst.fr/~cardoso/jfbib.html> – *One of the best review articles on ICA. The author, one of the leading researchers in the field, gives a comprehensive survey on different approaches to ICA and algorithms used therein.*
- A. Hyvärinen. Survey on Independent Component Analysis. *Neural Computing Surveys*, 2:94–128, 1999. [Available Online] <http://www.cis.hut.fi/aapo/papers/NCS99web> – *A comprehensive survey paper on ICA by a leading researcher in the field.*
- T.-P. Jung *et al.* Imaging brain dynamics using Independent Component Analysis. In

Proceedings of the IEEE, vol. 89(7), pp. 1107–1122, 2001. [Available Online]
<http://www.sccn.ucsd.edu/~scott> – *A review article on biomedical applications of ICA written by the pioneers of the field. The paper is based mainly on research of S. Makeig et al. and highlights several important applications of ICA in EEG, MEG, ECG and fMRI.*

Appendix: implementations of BSS algorithms

- A. Hyvärinen - fixed-point ICA (FastICA) algorithm (language: MATLAB)
<http://www.cis.hut.fi/projects/ica/fastica>
- M. Zibulevsky - Relative Newton algorithm (language: MATLAB)
<http://iew3.technion.ac.il/~mcib>
- EEGLab - toolbox for processing and visualization of electrophysiological data. Includes an implementation of the InfoMax BSS algorithm (language: MATLAB, C)
<http://www.sccn.ucsd.edu/eeglab>

Cross-references (incorporated in or identical to the article)

Blind deconvolution, See: **Blind source separation: biomedical applications**

Sparse component analysis, See: **Blind source separation: biomedical applications**

Index

blind source separation

 mixing: linear, nonlinear, post-nonlinear, instantaneous, convolutive

 blind deconvolution: MIMO, SIMO, SISO

 short-time Fourier transform

permutation
sparse representation
information maximization
contrast function: orthogonal, non-orthogonal, multi-unit, one-unit
identifiability conditions
independent component analysis: online, batch
independent factor analysis
Kullback-Leibler divergence
mutual information
entropy: differential, neg-
likelihood: maximum, quasi maximum, log-
sub-Gaussian
super-Gaussian
kurtosis
relative optimization
relative Newton: see relative optimization
relative gradient: see relative optimization
natural gradient: see relative optimization
sparse component analysis

References

- [1] D. Pham and P. Garrat. Blind separation of a mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Trans. Sig. Proc.*, 45:1712–1725, 1997.
- [2] A.J. Bell and T.J. Sejnowski. An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

- [3] P. Comon. Independent component analysis – a new concept. *Signal Processing*, 36(3):287–314, 1994.
- [4] J.-F. Cardoso and A. Souloumiac. An efficient technique for blind separation of complex sources. In *Proc. IEEE SP Workshop on Higher-Order Stat.*, pages 275–279, 1993.
- [5] M. Zibulevsky. Blind source separation with relative Newton method. In *Proc. International Symposium on Independent Component Analysis and Blind Signal Separation*, pages 897–902, 2003.
- [6] A. Cichocki, R. Unbehauen, and E. Rummert. Robust learning algorithm for blind separation of signals. *Electronics Letters*, 30(17):1386–1387, 1994.
- [7] J.-F. Cardoso and B. Laheld. Equivariant adaptive source separation. *IEEE Trans. Sig. Proc.*, 44(12):3017–3030, 1996.
- [8] S.-I. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems*, 8:757–763, 1996.
- [9] A. M. Bronstein, M. M. Bronstein, and M. Zibulevsky. Blind source separation using block-coordinate relative Newton method. *Signal Processing*, 84(8):1447–1459, August 2004.
- [10] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi. Blind deconvolution of images using optimal sparse representations. *IEEE Trans. Image Processing*, 14(6):726–736, June 2005.
- [11] A. M. Bronstein, M. M. Bronstein, and M. Zibulevsky. Relative optimization for blind deconvolution. *IEEE Trans. Signal Processing*, 53(6):2018–2026, June 2005.
- [12] A. Hyvarinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.
- [13] M. S. Lewicki and T. J. Sejnowski. Coding time-varying signals using sparse, shift-invariant representations. *Advances in Neural Information Processing Systems*, 11:730–736, 1999.

- [14] M. Zibulevsky and B. A. Pearlmutter. Blind source separation by sparse decomposition. *Neural Computation*, 13(4), 2001.
- [15] M. Zibulevsky, P. Kisilev, Y. Y. Zeevi, and B. A. Pearlmutter. Blind source separation via multinode sparse representation. *Advances in Neural Information Processing Systems*, 12, 2002.
- [16] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi. Sparse ICA for blind separation of transmitted and reflected images. *International Journal of Imaging Science and Technology*, 15(1):84–91, 2005.
- [17] H. Attias. Independent factor analysis. *Neural Computation*, 11(4):803–851, 2001.
- [18] P. Smaragdis. Blind separation of convolved mixtures in the frequency domain. *Neurocomputing*, 22:21–34, 1998.
- [19] N. Mitianoudis and M. Davies. Permutation alignment for frequency domain ICA using subspace beamforming methods. In *Proc. International Symposium on Independent Component Analysis and Blind Signal Separation*, 2004.
- [20] S. Harmeling, A. Ziehe, M. Kawanabe, B. Blankertz, and K.-R. Muller. Nonlinear blind source separation using kernel feature spaces. In *Proc. International Symposium on Independent Component Analysis and Blind Signal Separation*, 2001.
- [21] T.-P. Jung, S. Makeig, M. J. McKeown, A. J. Bell, T.-W. Lee, and T. J. Sejnowski. Independent component analysis of biomedical signals. *Imaging brain dynamics using Independent Component Analysis*, 89(7):1107–1122, 2001.
- [22] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski. Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, 8:145–151, 1996.

- [23] S. Makeig, T. P. Jung, A. J. Bell, and T. J. Sejnowski. Blind separation of auditory event-related brain responses into independent components. *Proc. of National Academy of Sciences*, 94:10979–10984, 1997.
- [24] T.-P. Jung, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, S. Makeig, and T. J. Sejnowski. Removing electroencephalographic artifacts from by blind source separation. *Psychophysiology*, 37:163–178, 2000.
- [25] V. Zarzoso, A. K. Nandi, and E. Bacharakis. Maternal and foetal ecg separation using blind source separation methods. *IMA Journal of Mathematics Applied in Medicine and Biology*, 14(3):207–225, 1997.
- [26] M. J. McKeown, S. Makeig, C. G. Brown, T.-P. Jung, S. S. Kindermann, and T. J. Sejnowski. Analysis of fMRI by blind separation into independent spatial components. *Human Brain Mapping*, 6(3):160–188, 1998.
- [27] J. V. Stone, J. Porrill C. Buchel, and K. Friston. Spatial, temporal, and spatiotemporal independent component analysis of fMRI data. In *Proc. 18th Leeds Statistical Research Workshop on Spatial-temporal modelling and its applications*, 1999. Online: <http://www.shef.ac.uk/pc1jvs/papers/published.html>.

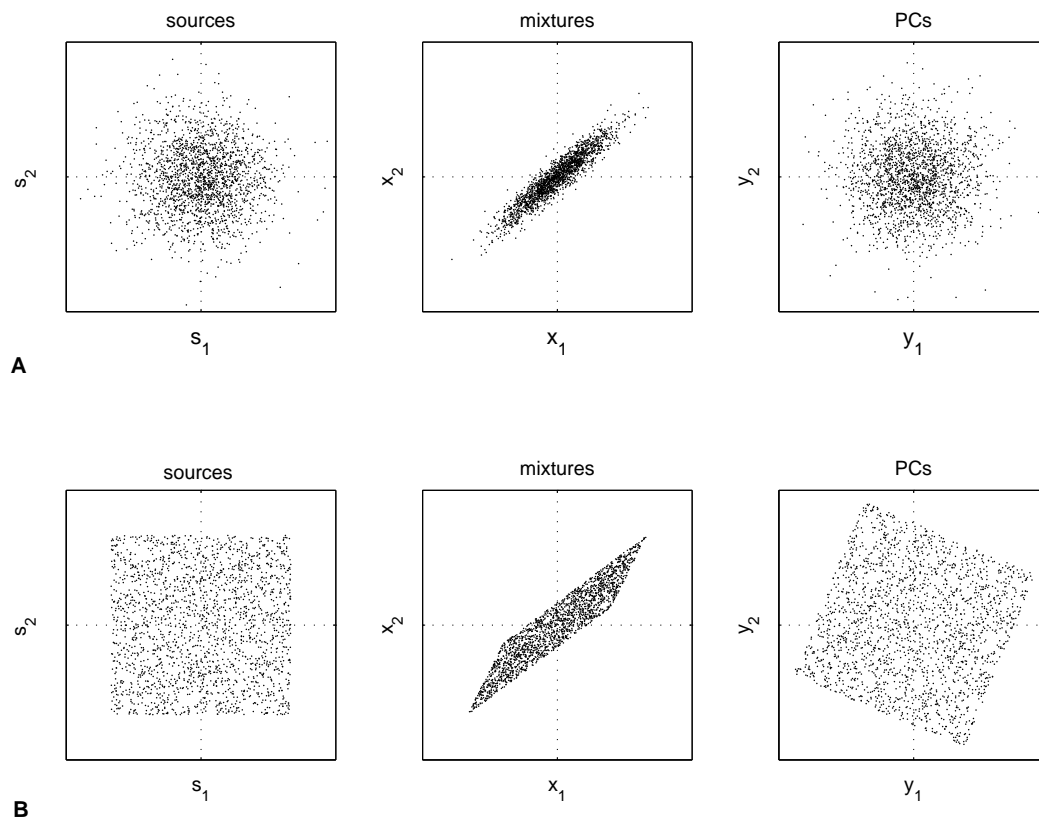


Figure 1: PCA applied to mixtures of two different kinds of sources: Gaussian (A) and uniform (B).

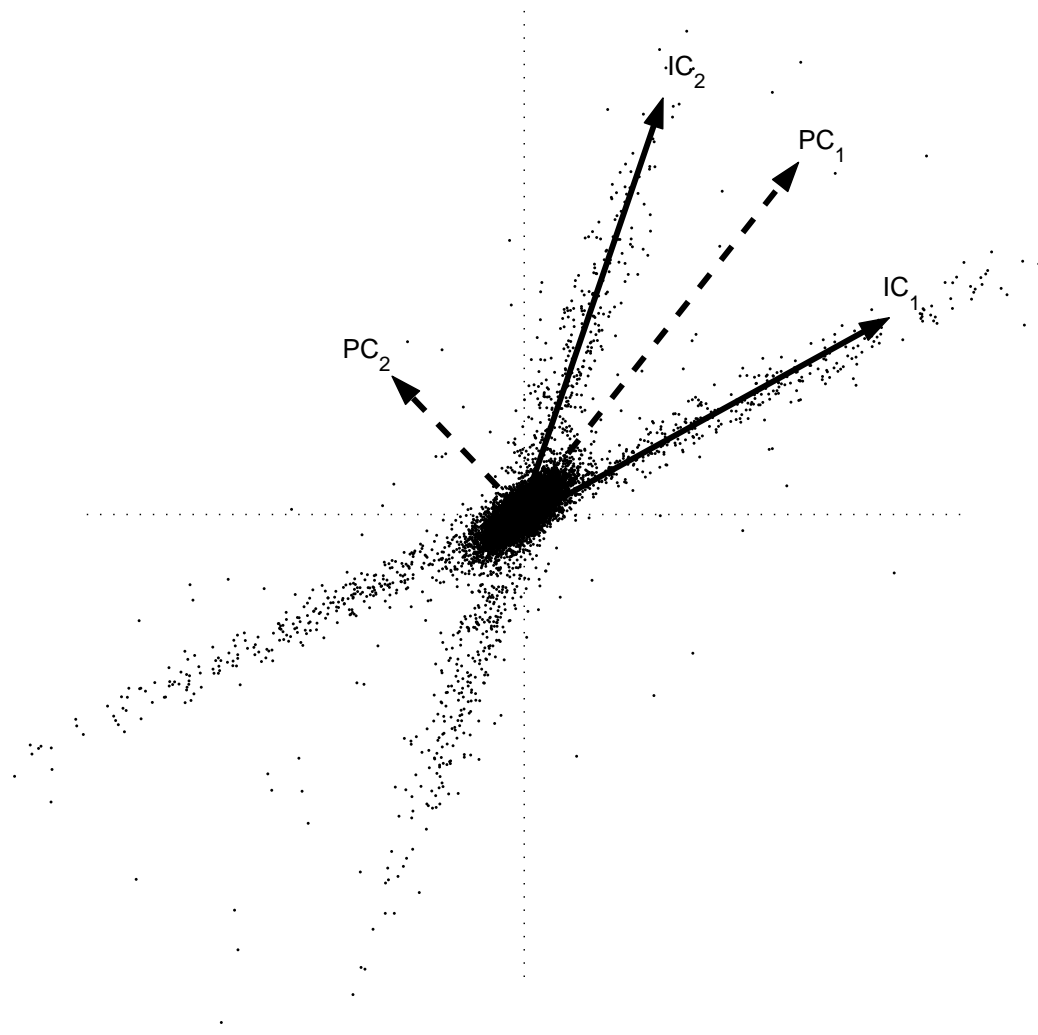


Figure 2: The difference between PCA and ICA on non-Gaussian sources.

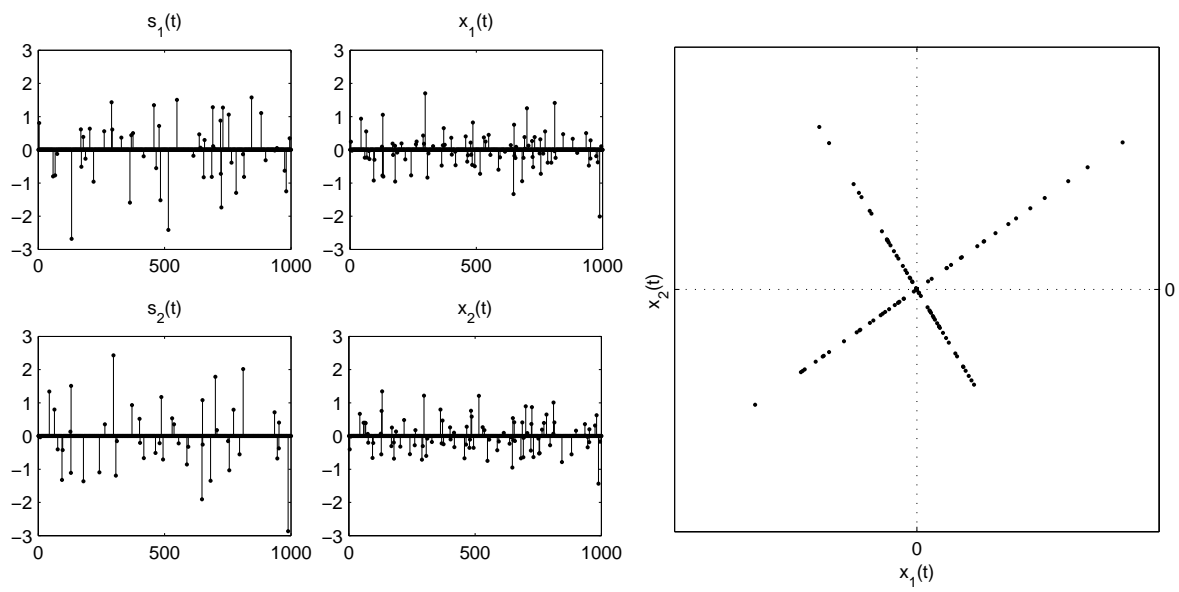


Figure 3: Sparse sources (left), their mixtures (center) and the mixtures scatter plot (right).

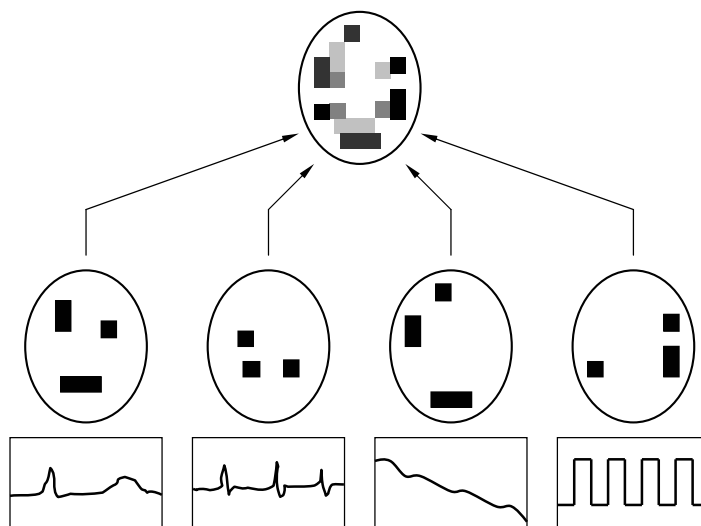


Figure 4: Linear mixing model in fMRI.