Piecewise Linear Source Separation

Rémi Gribonval

Piecewise linear source separation

Rémi Gribonval

IRISA, Rennes, France

ABSTRACT

We propose a new framework, called *piecewise linear separation*, for blind source separation of possibly degenerate mixtures, including the extreme case of a single mixture of several sources. Its basic principle is to : 1/ decompose the observations into "components" using some sparse decomposition/nonlinear approximation technique; 2/ perform separation on each component using a "local" separation matrix. It covers many recently proposed techniques for degenerate BSS, as well as several new algorithms that we propose. We discuss two particular methods of multichannel decompositions based on the Best Basis and Matching Pursuit algorithms, as well as several methods to compute the local separation matrices (assuming the mixing matrix is known). Numerical experiments are used to compare the performance of various combinations of the decomposition and local separation methods. On the dataset used for the experiments, it seems that BB with either cosine packets of wavelet packets (Beylkin, Vaidyanathan, Battle3 or Battle 5 filter) are the best choices in terms of overall performance because they introduce a relatively low level of artefacts in the estimation of the sources; MP introduces slightly more artefacts, but can improve the rejection of the unwanted sources.

Keywords: degenerate blind source separation, piecewise linear separation, sparse decomposition, nonlinear approximation, Best Basis, Matching Pursuit, denoising, Wiener filter, masking, clustering

1. INTRODUCTION

Source separation is a problem that arises when one or several sensor(s) record data to which can contribute several generating physical processes. Perhaps the most striking example of BSS problem consists in recovering the contributions of several musical instruments to a stereophonic audio recording. If we denote by $s_n(t)$ the signal emitted by the n-th instrument ($1 \leq n \leq N$) and $x_m(t)$ the data recorded on the m-th channel of the recording (here $1 \leq m \leq M = 2$), we can make the (simplistic) instantaneous linear mixture model

$$x_m(t) = \sum_n a_{m,n} s_n(t), \qquad 1 \leq m \leq M, \qquad \forall t$$

and try to recover the source signals $s_n(t)$ from the two mixtures $x_1(t), x_2(t)$. More generally, Blind Source Separation consists in recovering N unknown sources $\{s_n(t)\}_{n=1}^N$ from M instantaneous mixtures $\{x_m(t)\}_{m=1}^M$. The instantaneous linear mixture model is conveniently expressed using the matrix notation

$$\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ \dots \\ x_M(t) \end{pmatrix} = \mathbf{A} \begin{pmatrix} s_1(t) \\ \dots \\ s_N(t) \end{pmatrix} + \begin{pmatrix} w_1(t) \\ \dots \\ w_M(t) \end{pmatrix} = \mathbf{A}\mathbf{s}(t) + \mathbf{w}(t), \qquad \forall t \tag{1}$$

or $\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{w}$ where $\mathbf{w}$ is additive noise. Note that as a general notation in this paper, we will use bold letters to denote variables that are "multichannel", such as $\mathbf{x}$ or the mixing matrix $\mathbf{A}$, and plain letters to denote variables that correspond to only one channel, such as s_n. Note also that we consider real or complex data (signals and matrices).

Considering the case of discrete signals of T samples ($T \gg \max(M, N)$), with the assumption that there is no noise ($\mathbf{w} = 0$), BSS can be seen as a factorization problem : the $M \times T$ matrix $\mathbf{x}$ should be factored into the $M \times N$ matrix $\mathbf{A}$ and the $N \times T$ matrix $\mathbf{s}$. This is obviously an ill-posed problem, and its solution cannot be defined without additional assumptions, on the sources (such as independence [1] and positivity [2]) or on the mixing matrix.

Further author information: (Send correspondence to R. Gribonval)

R. Gribonval: E-mail: remi.gribonval@inria.fr, Telephone: (+33) 299 84 25 06, Address: IRISA, METISS Project, Campus de Beaulieu, F-35042 Rennes CEDEX, France

The most widely studied BSS situation is the (over)determined case where there is at least as many mixtures as there are sources, *i.e.* $M \geq N$. In this case, estimating the mixing matrix $\mathbf{A}$ is sufficient to get an estimate of the sources, and the standard methods (see [1] and the references therein) have essentially the following structure : an estimate $\widehat{\mathbf{A}}$ of the mixing matrix is obtained (by optimizing some contrast function which is generally highly nonlinear, or by joint diagonalization of higher order cumulants; the (pseudo)inverse $\widehat{\mathbf{B}} := \widehat{\mathbf{A}}^{\dagger}$ of the mixing matrix is applied to the mixtures to estimate the sources as $\widehat{\mathbf{s}} := \widehat{\mathbf{B}}\mathbf{x}$, *i.e.*

$$\widehat{s}_n(t) = \sum_{m=1}^{M} \widehat{b}_{n,m} x_n(t).$$

If there is no noise ($\mathbf{w} = 0$) and a perfect estimate of $\mathbf{A}$ is available, these methods provide perfect recovery of the sources. In general, there are however intrinsic limitations [3] to the accuracy of the estimation of $\mathbf{A}$.

In this paper we are particularly interested in the degenerate case $M < N$. In this case, even if it is still possible [4] to estimate the mixing matrix, the knowledge of $\mathbf{A}$ is not sufficient to estimate the sources, because (as noted by [5]) the equation $\mathbf{x} = \mathbf{A}\mathbf{s}$ has an affine set of solutions. To select a prefered solution $\widehat{\mathbf{s}}$ in this set, one can still choose to rely on a demixing matrix $\widehat{\mathbf{B}}$, *i.e.* $\widehat{s} = \widehat{\mathbf{B}}\mathbf{x}$. However the performance of such a linear separation in the degenerate case has intrinsic limitations [6] : "good" estimators of unknown sources from degenerate mixtures are necessarily nonlinear.

Recently, several algorithms for the separation of more sources than mixtures have been proposed [7–11]. They rely on some joint representation of at least $M = 2$ mixtures in some signal dictionary, followed by a clustering technique which is often used to build binary masks. Similarly, in the interesting (but somehow extreme) case where only one mixture is available, Roweis [12] proposed to build binary time-frequency masks based on Hidden Markov Models (HMM) of the sources. Benaroya [13–15] combined HMM with Gaussian Mixture Models (GMM) of the sources to replace the binary masks by "adaptive Wiener filters", and Jang *et al.* replaced binary masks with weights based on generalized Gaussian models of the sources [16].

We propose a general framework, called *piecewise linear separation*, for the separation of possibly more sources than sensors. It generalizes the techniques for degenerate BSS that we have mentioned above. In Section 2 we introduce our framework which is based on two elements : the choice of a decomposition of the observed mixtures into *components* (based, *e.g.*, on classical time-frequency or time-scale methods such as frame decompositions, Best Basis, Matching Pursuit or Basis Pursuit); the choice of *local separation matrices* to estimate the components of the sources from those of the mixtures. In Section 3 we discuss the possible multichannel decompositions that will serve as the first element of piecewise linear separation, and we detail two methods based on the Best Basis and Matching Pursuit algorithms. In Section 4 we discuss the computation of the local separation matrices, assuming the mixing matrix $\mathbf{A}$ is known, based on some Bayesian priors for the sources. In Section 5 we perform numerical experiments to compare the performance of some combinations of decomposition methods and local separation strategies. We conclude by discussing what seem to be the main challenges as well as the most promising research directions to design better algorithms within the piecewise linear separation framework.

2. PIECEWISE LINEAR SEPARATION

Several authors have proposed BSS algorithms for the possibly degenerate case that are based on some time-frequency/time-scale representation of the data followed by binary masking [7, 8, 11, 12]. The underlying model is that at most one source is "active" in each component of the representation. In the one microphone setting, some authors [13–16] proposed to replace binary masking with some smoother form of filtering, where each component is split adaptively to obtain the components of the sources. Here we want to gather the main ideas that have emerged and propose a global framework for degenerate BSS : *piecewise linear separation*. The key observation is that a good data representation often makes it possible to decompose a single degenerate BSS problem into several (over)determined problems. The basic principle is simply to :

1. decompose the observations $\mathbf{x}$ into "components" $\mathbf{x}^j$;
2. perform separation on each component using a "local" separation matrix.

Before we discuss how to perform these two steps, let us see how piecewise linear separation works in an ideal case.

2.1. Ideal model; some notations

Let us assume the sources can be expressed as a sum $\mathbf{s} = \sum_j \mathbf{s}^j$, which may be written componentwise $s_n(t) = \sum_j s_n^j(t)$, where we will call s_n^j the j-th "component" of the source s_n. For example, if $\mathcal{B} = \{g_j(t),\ 1 \leq j \leq T\}$ is an orthonormal basis we have such a decomposition with $s_n^j(t) = \langle s_n, g_j \rangle g_j(t)$. More general decompositions will be considered in Section 3. Assume in addition that for each component, the set $I_j = \{n,\ s_n^j \neq 0\}$ of "active" sources is known and contains at most M entries. From the noiseless mixing model $\mathbf{x} = \mathbf{A}\mathbf{s}$, we easily derive *a* decomposition $\mathbf{x} = \sum_j \mathbf{x}^j$ of the observations, with $\mathbf{x}^j = \mathbf{A}\mathbf{s}^j = \mathbf{A}_{I_j}\mathbf{s}^j_{I_j}$, where $\mathbf{A}_{I_j}$ is the $M \times \text{card}(I_j)$ matrix with columns $\mathbf{A}_n, n \in I_j$ ($\mathbf{A}_n$ is the n-th column of the mixing matrix $\mathbf{A}$), and $\mathbf{s}^j_{I_j}$ is the $\text{card}(I_j) \times T$ matrix with rows $s_n^j, n \in I_j$. Thus, there exist *a* split of the original problem into an equivalent collection of subproblems :

$$\left\{ \begin{array}{rcll} \mathbf{x}^j & = & \mathbf{A}_{I_j}\mathbf{s}^j_{I_j} & \\ s_n^j & = & 0, \qquad n \notin I_j & , \end{array} \right. \qquad \forall j$$

Because $\text{card}(I_j) \leq M$, each subproblem is (over)determined, so we get

$$\mathbf{s}^j_{I_j} \ := \ \mathbf{A}^\dagger_{I_j}\mathbf{x}^j; \tag{2}$$

$$s_n^j \ := \ 0, \qquad n \notin I_j. \tag{3}$$

The ideal case requires a noiseless problem with several assumptions that are generally unrealistic. We are indeed supposed to know :

- the mixing matrix $\mathbf{A}$;
- how to decompose $\mathbf{x}$ so as to "match" the decomposition of $\mathbf{s}$;
- the set I_j of "active" sources on each component, which should satisfy $\text{card}(I_j) \leq M$;

Under these ideal assumptions, we have seen that it is possible to perfectly recover the sources.

2.2. Principle of piecewise linear separation algorithms

Of course, real BSS problems may include additive noise and, most of all, they are blind :

- the mixing matrix is unknown;
- it is not known in advance which (and how many) sources are "active" in which component;
- it may not be possible to claim that only one source is active in each component;
- in the extreme –but interesting– *one microphone* setting, one may even have to relax the assumption that at most M sources are active on one component.

Consider for example the separation of N instruments in a commercial musical recording. Not only are the sources dependent (at least at a high level point of view : they follow the same musical score), they are also temporally synchronized and have common harmonics, so there almost certainly exist time-frequency components that contain contributions of several sources. If one ever wants to address audio BSS problems of this nature (and it seems to be the implicit dream of almost every researcher in the field of BSS!), the assumption that at most one source contributes to each component is not quite realistic.

We propose the following general algorithmic structure for piecewise linear BSS, where some of the steps may be performed jointly :

1. Compute a decomposition $\mathbf{x} = \sum_j f_j[\mathbf{x}]$;
2. Compute local separation matrices $\mathbf{B}_j$ for each j;

3. Recover the sources

$$\widehat{\mathbf{s}} := \sum_j \mathbf{B}_j f_j[\mathbf{x}]; \tag{4}$$

In Section 3 we discuss the decomposition step (step 1), which may rely on nonlinear approximation and sparse decomposition techniques. The local separation step (step 2) is probably the most critical one. In Section 4 we discuss how it can be performed based on Bayesian priors when one knows an estimate of the mixing matrix $\mathbf{A}$. Estimating $\mathbf{A}$ is a problem by itself which can be adressed by various techniques [4,7,17–20] but its not dealt with in this paper.

Before we discuss both steps and show that that this general framework covers many recently proposed algorithms, let us emphasize that it also makes it possible to design new algorithms that :

- deal with more than one "active" source per component (and sometimes even more than M sources, such as in the *one microphone* setting);
- decide *globally* which sources are active in which component, by taking into account the dependencies between components (harmonic relations, persistence of instantaneous frequency across time, persistence of transients across scale, ...) rather than making independent decisions for each component.

3. MULTICHANNEL SIGNAL DECOMPOSITIONS

Besides the raw (time-domain) representation of a signal, there are many ways to decompose a (monochannel) signal : linear time-frequency/time-scale transforms such as wavelets bases, wavelet frames, Gabor frames, Wilson bases [21]; adaptive methods using local cosine bases, wavelet packets, with the Best Basis algorithm [22]; dictionary decompositions using Matching Pursuit [11,23] or Basis Pursuit [24]. In this section we discuss how these *monochannel* signal decompositions can be adapted to get *multichannel* ones.

3.1. Notations

First, let us introduce a few notations. For multichannel signals $\mathbf{y}(t) = (y_1(t), \ldots, y_L(t))^T$ ($(\cdot)^T$ denotes the real transpose while $(\cdot)^H$ denotes the Hermitian transpose) and $\mathbf{z}(t) = (z_1(t), \ldots, z_L(t))^T$ with an arbitrary number L of channels we define the inner product and its associated norm

$$\langle \mathbf{y}, \mathbf{z} \rangle := \sum_{l=1}^{L} \langle y_l, z_l \rangle = \sum_{l,t} y_l(t) z_l^\star(t) \tag{5}$$

$$||\mathbf{y}||^2 := \sum_{l=1}^{L} ||y_l||^2 = \sum_{l,t} |y_l(t)|^2 \tag{6}$$

where $(\cdot)^\star$ denotes complex conjugation.

3.2. Frame decompositions

Frames are families of "atoms" $\{g_j(t)\}$ such that for all (monochannel) signals y $A||y||^2 \leq \sum_j |\langle y, g_j \rangle|^2 \leq B||y||^2$, where $A > 0$ and $B < \infty$. For any frame, there exists a dual frame $\{\widetilde{g}_j\}$ such that every signal y has the frame decomposition $y(t) = \sum_j \langle y, \widetilde{g}_j \rangle g_j(t)$. For multichannel signals (of any dimension L) $\mathbf{y}(t) = (y_1(t), \ldots, y_L(t))^T$ we define, (by abuse of notation we use the same notation f_j for all dimensions L)

$$f_j[\mathbf{y}] := \begin{pmatrix} \langle y_1, \widetilde{g}_j \rangle g_j \\ \cdots \\ \langle y_L, \widetilde{g}_j \rangle g_j \end{pmatrix}$$

and get the desired multichannel decomposition $\mathbf{y} = \sum_j f_j[\mathbf{y}]$. With such a decomposition, solving the initial BSS problem (1) is equivalent to solving the collection of "sub-problems" $f_j[\mathbf{x}] = \mathbf{A} f_j[\mathbf{s}] + f_j[\mathbf{w}]$ (notice that $f_j[\mathbf{x}]$ and $f_j[\mathbf{w}]$ have M rows while $f_j[\mathbf{s}]$ has N rows).

The most basic frame decomposition is the trivial representation where the atoms are simply Diracs $g_j(t) = \delta(t-j)$, it is used in Van Hulle's clustering approach to degenerate BSS [7]. Another classical transform is the Short Time Fourier Transform (STFT) used in the Degenerate Unmixing and Estimation Technique (DUET algorithm) of Jourjine *et al*. [8]. To the STFT correspond time-frequency atoms $g_{t,f}(t') = g(t'-t)e^{2\pi i f(t-t')}$ localized at time t and frequency f (see, *e.g*., the book of Mallat [21]). All sorts of Gabor and wavelet frames as well as local cosine bases or wavelet packets bases can be used similarly.

3.3. Other linear decompositions

Instead of decomposing the signals along "atoms", it is also possible to cut them linearly into larger pieces (or "molecules") that correspond to higher dimensional subspaces of the signal space. For example, in overlap-add methods (OLA), it is classical to decompose a signal into windowed pieces $g(t-jL)y(t)$ where the shifted windows $\{g(t-jL)\}$ sum to one. Similarly, it is possible to decompose a signal into subbands using a family of filters. For example, based on an orthonormal basis of wavelets [21] $\{\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)\}$ one can build :

$$f_j[\mathbf{y}] := \begin{pmatrix} \sum_k \langle y_1, \psi_{j,k}\rangle \psi_{j,k} \\ \cdots \\ \sum_k \langle y_L, \psi_{j,k}\rangle \psi_{j,k} \end{pmatrix}.$$

Which linear decomposition is better for piecewise linear separation certainly depends on the nature of the sources, and it requires numerical experiments on large databases to choose a linear representation that is well adapted to a target application. For instance, DUET [8] is based on the STFT on the ground that independent audio sources seem to be quasi "W-disjoint orthogonal" [25,26], while Jang *et al*. [16] rely on sparse coding [27–31] to compute the representation used for their algorithm. As an alternative to selecting *a priori* the linear transform based on knowledge of the class of sources we want to separate, it is possible to choose it adaptively depending on the observed mixtures $\mathbf{x}$, using nonlinear optimization techniques such as Best Basis [22] (BB) or Matching Pursuit [11,23] (MP).

3.4. Best orthogonal basis

Assume we have at hand a "library" of orthonormal bases $\{\mathcal{B}_\lambda\}$ and some cost function that predicts the performance of a basis $\mathcal{B}$ for a target application. The principle of Best Basis (BB) is simply to pick up the best basis $\mathcal{B}_{\lambda_0} = \{g_j^{\lambda_0}(t)\}_{j=1}^T$ in the library according to the cost function. When the target application is the nonlinear approximation of a monochannel signal y, a typical cost function will be, for $\mathcal{B} = \{g_j\}$:

$$C(y, \mathcal{B}) := \sum_{j=1}^{T} \Phi\left(\frac{|\langle y, g_j\rangle|^2}{||y||^2}\right)$$

where Φ is some concave function [21]. Such cost function will be small when the energy of y is well concentrated on a few coefficients, and it will take the largest value $T\Phi(1/T)$ when the energy is spread equally on all coefficients. For nice tree-structured libraries such as wavelet packets or local cosines, the best basis can be found with a fast search, provided that the cost function is additive [22]. Once the basis $\lambda(y) := \arg\min_\lambda C(y, \mathcal{B}_\lambda)$ is selected, we can perform the decomposition using the linear transform corresponding to $\mathcal{B}_{\lambda(y)}$, *i.e*.

$$y = \sum_{j=1}^{T} \langle y, g_j^{\lambda(y)}\rangle g_j^{\lambda(y)}.$$

In the case of multichannel signals, assume we have at hand some cost function $C(\mathbf{x}, \mathcal{B})$. We can adopt the BB strategy by selecting $\lambda(\mathbf{x}) := \arg\min_\lambda C(\mathbf{x}, \mathcal{B}_\lambda)$ and decomposing $\mathbf{x} = \sum_j f_j[\mathbf{x}]$ with

$$f_j[\mathbf{x}] = \begin{pmatrix} \langle x_1, g_j^{\lambda(\mathbf{x})}\rangle g_j^{\lambda(\mathbf{x})} \\ \cdots, \\ \langle x_M, g_j^{\lambda(\mathbf{x})}\rangle g_j^{\lambda(\mathbf{x})} \end{pmatrix} = \mathbf{A}\begin{pmatrix} \langle s_1, g_j^{\lambda(\mathbf{x})}\rangle g_j^{\lambda(\mathbf{x})} \\ \cdots, \\ \langle s_M, g_j^{\lambda(\mathbf{x})}\rangle g_j^{\lambda(\mathbf{x})} \end{pmatrix} + \begin{pmatrix} \langle w_1, g_j^{\lambda(\mathbf{x})}\rangle g_j^{\lambda(\mathbf{x})} \\ \cdots, \\ \langle w_M, g_j^{\lambda(\mathbf{x})}\rangle g_j^{\lambda(\mathbf{x})} \end{pmatrix}.$$

The most straightforward choice of a cost function is certainly

$$C(\mathbf{x}, \mathcal{B}) := \sum_{j=1}^{T} \Phi\left(\frac{\sum_{n=1}^{N} |\langle x_n, g_j \rangle|^2}{||\mathbf{x}||^2} \right),$$

where $||\mathbf{x}||^2 = \sum_{n=1}^{N} ||x_n||^2$ measures the joint energy of the channels, and each term $\sum_{n=1}^{N} |\langle x_n, g_j \rangle|^2 / ||\mathbf{x}||^2$ gives the relative amount of joint energy that is carried by the j-th component.

3.5. Basis Pursuit

Zibulevsky and Pearlmutter [9] proposed BSS algorithms based on sparse decompositions of the sources in some redundant signal dictionary of atoms $\mathcal{D} = \{g_k(t), 1 \leq k \leq RT\}$. The dictionary typically consists in a family of time-frequency atoms $g_{t,f}$, or wavelets $g_{s,t}(t') = \psi((t'-t)/s)$, etc, but other less standard dictionaries can be considered thay may be estimated from training data [27–32] or may combine several classical dictionaries into a larger one [23, 33–35]. Because of the assumed redundancy of the dictionary, any monochannel signal has infinitely many representations $y(t) = \sum_k c_k g_k(t)$ with coefficients $\mathbf{c} = \{c_k\}_k$, so one has some freedom in choosing the most convenient set of coefficients. Zibulevsky and Pearlmutter choose a "sparse" decomposition $\mathbf{c}$ among all the possibilities according to some "sparseness" criterion, which is derived in [9] from a probabilistic model of the unknown coefficients. When the sparseness measure is the ℓ^1 norm $||\mathbf{c}||_1$ (it corresponds to a Laplacian model on the unknown coefficients), the coefficients are recovered using Basis Pursuit (BP) [24]. They are generally computed with Linear Programming algorithms which are computationally intensive. In the multichannel case, Zibulevsky and Pearlmutter [9] propose a joint estimation of the sources and the matrix via a MAP optimization under the Laplacian model.

3.6. Matching Pursuit

BP is computationally intensive and tricky to implement with arbitrary dictionaries, even in the monochannel case. To the opposite, the Matching Pursuit algorithm is quite generic and easy to implement, as well as easily generalized to multichannel decompositions. Moreover, it shares some of the good properties of Basis Pursuit for the perfect recovery of very sparse expansions in well-behaved dictionaries [36–39]. Matching Pursuit (MP) [23] is a strategy that selects a sequence $\lambda = \{k_j\}$ of indexes of atoms, and iteratively computes some residuals $R_j^{\lambda}[y]$ (starting from $R_1^{\lambda}[y] := y$) as

$$R_{j+1}^{\lambda}[y] := R_j^{\lambda}[y] - \langle R_j^{\lambda}[y], g_{k_j} \rangle g_{k_j}. \tag{7}$$

In MP the sequence $\lambda(y)$ is iteratively selected depending on the signal y :

$$k_j(y) := \arg\max_k |\langle R_j^{\lambda(y)}[y], g_k \rangle|^2. \tag{8}$$

Hence, after J iterations, y is decomposed as

$$y(t) = \sum_{j=1}^{J} \langle R_j^{\lambda(y)}[y], g_{k_j(y)} \rangle g_{k_j(y)}(t) + R_{J+1}^{\lambda(y)}[y](t).$$

The author proposed a version of the Matching Pursuit for stereophonic audio signals [11] which we readily extend to an arbitrary number of channels. After J iterations, it provides a decomposition of $\mathbf{x} = \sum_{j=1}^{J+1} f_j[\mathbf{x}]$ with

$$f_j[\mathbf{x}] = \begin{pmatrix} \langle R_j^{\lambda(\mathbf{x})}[x_1], g_{k_j(\mathbf{x})} \rangle g_{k_j(\mathbf{x})} \\ \cdots \\ \langle R_j^{\lambda(\mathbf{x})}[x_M], g_{k_j(\mathbf{x})} \rangle g_{k_j(\mathbf{x})} \end{pmatrix}, \qquad 1 \leq j \leq J \tag{9}$$

$$f_{J+1}[\mathbf{x}] = \begin{pmatrix} R_{J+1}^{\lambda(\mathbf{x})}[x_1] \\ \cdots \\ R_{J+1}^{\lambda(\mathbf{x})}[x_M] \end{pmatrix}. \tag{10}$$

As a natural extension of the monochannel case, we propose to select the indexes as

$$k_j(\mathbf{x}) := \arg\max_k \sum_{n=1}^{N} |\langle R_j^{\lambda(\mathbf{x})}[x_n], g_k\rangle|^2 \tag{11}$$

so as to pick up at each iteration a component that carries as much as possible of the joint energy of the multichannel residual. Such a multichannel MP shares the same convergence properties as the monochannel one [11,40], that is to say $\lim_{J\to\infty} ||f_{J+1}[\mathbf{x}]||^2 = 0$.

4. FROM BINARY MASKING TO LOCAL LINEAR SEPARATION

The first step in piecewise linear separation is a decomposition step, and we have seen several possible choices to perform it. In this section we discuss the second step, that is to say the computation of local separation matrices $\mathbf{B}_j$. As mentioned in the introduction, there are methods to estimate the mixing matrix $\mathbf{A}$ even in the degenerate case [4,7,17–20]. However, in the degenerate case, the knowledge of the mixing matrix is not sufficient to recover the sources [5]. In this section, we want to concentrate on source recovery, so we assume that an estimate of the mixing matrix $\mathbf{A}$ is available. Choosing a particular solution in the affine set of solutions to the equation $\mathbf{x} = \mathbf{A}\mathbf{s}$ requires *prior models* $p(\mathbf{s})$ of the sources, and the selection can rely on Bayesian estimators. In this section, we model the noise $\mathbf{w}$ as Gaussian, spatially and temporally white, with sample variance σ_w^2. Then, the Maximum A Posteriori (MAP) estimator is

$$\widehat{\mathbf{s}}_{\text{MAP}} := \arg\min_{\mathbf{s}} \Big\{ -\log p(\mathbf{s}|\mathbf{x}, \mathbf{A}) \Big\} = \arg\min_{\mathbf{s}} \Big[||\mathbf{x} - \mathbf{A}\mathbf{s}||^2 + h(\mathbf{s}) \Big] \tag{12}$$

where $h(\mathbf{s}) := -2\sigma_w^2 \log p(\mathbf{s})$, while the Conditional Mean (CM), which minimizes the mean square error, is

$$\widehat{\mathbf{s}}_{CM} := E(\mathbf{s}|\mathbf{x}, \mathbf{A}) = \int \mathbf{s}\, p(\mathbf{s}|\mathbf{x}, \mathbf{A})\, d\mathbf{s}. \tag{13}$$

Because denoising is certainly the most classical source separation problem, we start this section by showing how two classical denoising techniques, hard-thresholding and Wiener filtering, fit in the piecewise linear separation framework. In the remainder of the section we propose several prior models of the sources, which lead to various choices of local separation matrices. We recover as special cases some known techniques such as binary masking [7,8,11,12], separation by sparse decompositions [9,10] and adaptive Wiener filtering [13–15], but also get new possibilities. In Section 5 we will compare the performance of combinations of some of the decomposition methods proposed in Section 3 with some of the local separation strategies introduced below.

4.1. Denoising

We consider an observed signal $x = s_1 + s_2$ (s_1 is the source of interest, s_2 is the noise) and the corresponding mixing matrix is $\mathbf{A} = [1, 1]$. In the hard-thresholding strategy the source of interest is estimated by : 1/ computing the components $f_j[x] = \langle x, g_j\rangle g_j$ where $\{g_j\}_j$ is an orthonormal basis (*e.g.*, an orthonormal wavelet basis [21]); 2/ deciding whether s_1 is active in $f_j[x]$ by testing whether $|\langle x, g_j\rangle| > \theta$ where θ is a threshold, and applying the local separation matrices $\mathbf{B}_j = [1\ 0]^T$ if s_1 is considered active, $\mathbf{B}_j = [0\ 1]^T$ if it is not. Wiener filtering can also be seen as a piecewise linear separation algorithm : the components $f_j[x] = \langle x, g_j\rangle g_j$ correspond to the Fourier decomposition with $g_j(t) = \frac{1}{\sqrt{T}} \exp(\frac{2i\pi jt}{T})$ and we assume all sources are simultaneously active with known variance $\sigma_{s_i}[j] = E(|\langle s_i, g_j\rangle|^2)$. Local separation matrices are expressed as

$$\mathbf{B}_j = \left[\frac{\sigma_{s_1}^2[j]}{\sigma_{s_1}^2[j] + \sigma_{s_2}^2[j]} \ \frac{\sigma_{s_2}^2[j]}{\sigma_{s_1}^2[j] + \sigma_{s_2}^2[j]} \right]^T$$

4.2. Binary masking

The masking approaches to degenerate BSS [7,8,11,12] estimate $\mathbf{s}^j = (s_1^j, \cdots, s_N^j)^T$ from a piece of the observations $f_j[\mathbf{x}] = (f_j^1[\mathbf{x}], \cdots, f_j^M[\mathbf{x}])^T = \mathbf{A}\mathbf{s}^j + \mathbf{w}$ by : 1/ estimating the index $\widehat{n}_j$ of the only "active" source in the j-th component; 2/ recovering the sources by "masking" the components of one of the observed channels :

$$\widehat{s}_n := \sum_{j|\widehat{n}_j = n} f_j^{m_0}[\mathbf{x}] = \sum_j \delta_{n,\widehat{n}_j} \cdot f_j^{m_0}[\mathbf{x}]. \tag{14}$$

The choice of which channel m_0 is masked is arbitrary. Binary masking can be seen as piecewise linear separation with $\mathbf{B}_j = [\delta_{n,\widehat{n}_j} \cdot \delta_{m_0,m}]_{n,m}$ and the decomposition is either the raw data [7], a STFT [8,12] or adaptive time-frequency representations [9–11].

When only one mixture is available, Roweis [12] proposed to estimate the masks using Hidden Markov Models (HMM); when at least two mixtures are available, they can be estimated by exploiting the so called *spatial diversity* through clustering [7,8,11]. Spatial diversity can still be exploited when more than one source may be active in each component [10], provided that $\text{card}(I_j) \leq M$ where $I_j = \{n | s_n^j \neq 0\}$ is called the *hidden activity state* of the sources on the j-th component. Below we propose different strategies to estimate the hidden activity state depending on the choice of a prior model $p(I_j)$.

Disjoint orthogonal model. The most common assumption on the hidden activity states is that $\text{card}(I_j) = 1$, *i.e.* exactly one source is active in each component (in DUET [8], this is called "W-disjoint orthogonality"). This corresponds to a prior probability $p(I_j = \{n\}) = 1/N,\ 1 \leq n \leq N$. Based on this model, for each hypothesis $I_j = \{n\}$, Eqs. (2)-(3)) give the Maximum Likelihood (ML) estimate of the "would be active" source

$$\arg\min_{s_n^j} ||f_j[\mathbf{x}] - \mathbf{A}_n s_n^j||^2 = \mathbf{A}_n^\dagger f_j[\mathbf{x}] = \frac{\mathbf{A}_n^H f_j[\mathbf{x}]}{||\mathbf{A}_n||^2},$$

where $\mathbf{A}_n$ is the n-th column of the mixing matrix (remember we assume $\mathbf{A}$ is known). Thus, the MAP estimate of the hidden activity state is $\widehat{I}_j = \{\widehat{n}_j\}$ with

$$\widehat{n}_j := \arg\max_n \left|\left|\mathbf{A}_n^H f_j[\mathbf{x}]/||\mathbf{A}_n||\right|\right|^2. \tag{15}$$

Eventually the sources are estimated piecewise linearly as

$$\widehat{s}_n := \sum_{j|\widehat{n}_j = n} \frac{\mathbf{A}_n^H f_j[\mathbf{x}]}{||\mathbf{A}_n||^2} = \sum_j \delta_{n,\widehat{n}_j} \cdot \frac{\mathbf{A}_n^H f_j[\mathbf{x}]}{||\mathbf{A}_n||^2} \tag{16}$$

i.e. with local separation matrices

$$\mathbf{B}_j := \begin{bmatrix} \delta_{1,\widehat{n}_j} \cdot \mathbf{A}_1^H/||\mathbf{A}_1||^2 \\ \dots \\ \delta_{m,\widehat{n}_j} \cdot \mathbf{A}_N^H/||\mathbf{A}_N||^2. \end{bmatrix}. \tag{17}$$

Note that, to the opposite of Eq. (14), there is no need to arbitrarily choose the channel m_0 that is masked : the components of the sources are estimated by *projecting* orthogonally the components of the mixtures onto the direction of the corresponding column of the mixing matrix. We will compare numerically in Section 5 the performance of separation based on Eq. (14) and Eq. (16).

Other models. With any prior $p(I_j)$ on the hidden activity state such that $p(\text{card}(I_j) > M) = 0$, for each hypothesis on I_j, piecewise linear separation is given by Eqs. (2)-(3), hence the MAP estimate of I_j is

$$\widehat{I}_j := \arg\min_{I_j} \left[||f_j[\mathbf{x}] - \mathbf{A}_{I_j}\mathbf{A}_{I_j}^\dagger f_j[\mathbf{x}]||^2 + h(I_j)\right] = \arg\max_{I_j} \left[||\mathbf{A}_{I_j}\mathbf{A}_{I_j}^\dagger f_j[\mathbf{x}]||^2 - h(I_j)\right]$$

with $h(I_j) := -2\sigma_w^2[j] \log p(I_j)$ where $\sigma_w^2[j]$ is the variance of the noise on the j-th component. For example, if $p(I_j = \emptyset) = p$ and $p(I_j = \{n\}) = (1-p)/N,\ 1 \leq n \leq N$, then the usual choice $\widehat{I}_j = \{\widehat{n}_j\}$ given by Eq. (15) is replaced by $\widehat{I}_j = \emptyset$ whenever

$$||\mathbf{A}_{\widehat{n}_j}^H f_j[\mathbf{x}]||^2/||\mathbf{A}_{\widehat{n}_j}||^2 < 2\sigma_w^2[j] \cdot \log\left((1-p)/N \cdot p\right).$$

This is similar to denoising by hard-thresholding [41,42], where small components are considered as pure noise.

In [9,10], BSS algorithms based on sparse decompositions are proposed. It is not difficult to check that these algorithms recover the sources piecewise linearly just as in Eqs. (2)-(3): the hidden activity state $\widehat{I}_j$, which satisfies $\text{card}(\widehat{I}_j) = M$, is estimated by minimizing $\sum_n ||\widehat{s}_n^j||$. The underlying model is Laplacian for the source components and uniform over all hidden activity states of size M.

Masking based on structure. When only one channel is available, spatial diversity can no longer be exploited to compute the binary masks. Instead, other prior information has to be exploited. One possibility is to use global priors that may take the form of HMM [12]; another one consists in using the fact that different sources may yield different "types" of components. For instance, on single channel audio signals, transients and sustained parts can be separated based on a Matching Pursuit decomposition with a Gabor dictionary [43,44]; similarly, it is possible to separate edges from textures in images [34] using the fact that each of theses "sources" has a sparse representation in a different basis [33], and the basis corresponding to different sources are "incoherent". Recent results on sparse decompositions and nonlinear approximation with dictionaries [36,37,39,45–49] have given theoretical ground to these techniques which, again, have a piecewise linear form: after a nonlinear decomposition step, each component is used to recover the source to which it belongs.

4.3. Smooth masking

In theory, binary masking techniques can perfectly recover sources that globally linear BSS algorithms cannot recover. In practice, because the underlying model is not perfect, they often introduce artifacts due to the introduction of unnatural zeroes in the representation of the estimated sources. For audio sources, this leads to artifacts similar to the well known "musical noise" or "pipe noise" which is commonly encountered in transform-based coding.

Some authors [13–16] have proposed algorithms for single channel BSS based on "smoother" forms of masking that do not bring in the thresholding effect of binary masks. In these approaches, training data is first used to learn the parameters of generalized Gaussian models [16] (resp. Gaussian mixture models (GMM) [13–15]) of the two sources; separation is done piecewise linearly on new data with $\mathbf{B}_j = [\lambda_j, 1-\lambda_j]^T$, where λ_j is estimated by a MAP approach [16] or by combining different Wiener filters [13] to get the conditional mean (CM) estimator (see Eq. (13)).

When no training data is available, one has to design other techniques that rely on a more generic prior model with few or no parameters to estimate. With this aim, we propose a "smooth masking" strategy which, by analogy with GMM, we call Masking Mixture Model. It is based on hidden activity state models just as in binary masking, but MAP estimation is replaced by CM.

Masking Mixture Models (MMM). Assume we are given a prior $p(I_j)$ on the hidden activity states I_j such that $p(\text{card}(I_j) > M) = 0$. For each hypothesis on I_j, Eqs. (2)-(3) provide implicitly a local separation matrix $\mathbf{B}(I_j)$. The CM estimate of $\mathbf{s}^j$ is $\mathbf{s}^j = \mathbf{B}_j f_j[\mathbf{x}]$ with

$$\mathbf{B}_j := \sum_I p(I_j = I|\mathbf{x}, \mathbf{A})\mathbf{B}(I) \tag{18}$$

where we can compute the likelihood using the fact that

$$\log p(I_j = I|\mathbf{x}, \mathbf{A}) = -\frac{\|f_j[\mathbf{x}] - \mathbf{A}_I \mathbf{A}_I^\dagger f_j[\mathbf{x}]\|^2}{2\sigma_w^2[j]} + \log p(I_j = I) + cst$$

In this strategy, the only parameter to adjust is the variance of the noise. In the limit where $\sigma_w^2 \to 0$ we recover the binary masking strategy, but additional research is necessary to find out good strategies to adjust σ_w.

5. SOME EXPERIMENTAL RESULTS

So far, we have introduced several pieces (decomposition methods and local separation strategies) that can be assembled to make BSS algorithms within our proposed global framework of piecewise linear separation. In this section we gather the results of some experiments of source separation obtained with different combinations of these pieces. The experiments were performed on a stereophonic (M=2) instantaneous mixture of three normalized sources (s_1 =cello,s_2 =drums,s_3 =piano, the sampling rate was $8kHz$ and the number of samples was 19200 which corresponds to a duration of about $2.4s$) with known mixing matrix. In all experiments, we have used explicitly the knowledge of $\mathbf{A}$ to perform the local separation based on the various strategies. Estimating $\mathbf{A}$ is a difficult problem by itself, but the goal of these experiments is to compare the decomposition methods and the local separation strategies independently of the quality of the estimate of $\mathbf{A}$. The dataset (original sources, mixing matrix, mixtures) of the experiment is available online [50] and was already used in the papers [6,11].

Method	$\texttt{SDR}_1$	$\texttt{SDR}_2$	$\texttt{SDR}_3$	$\texttt{SIR}_1$	$\texttt{SIR}_2$	$\texttt{SIR}_3$	$\texttt{SAR}_1$	$\texttt{SAR}_2$	$\texttt{SAR}_3$
best globally linear	-0.05	1.23	11.46	-0.05	1.23	11.46	$+\infty$	$+\infty$	$+\infty$
BB (cosine packets)	5.83	6.25	12.32	18.82	22.21	25.76	6.11	6.39	12.53
BB (Beylkin)	5.27	6.56	12.56	20.82	20.06	26.05	5.43	6.80	12.77
BB (Vaidyanathan)	5.15	6.46	12.84	17.83	22.44	27.49	5.46	6.59	13.00
BB (Battle 3)	4.57	6.34	12.93	18.79	20.27	25.44	4.79	6.56	13.19
BB (Battle 5)	5.78	7.49	13.49	22.09	22.05	28.41	5.91	7.67	13.64
MP (2400 atoms), left masking	4.88	6.54	8.96	26.15	15.39	32.93	4.92	7.27	8.98
MP (2400 atoms), right masking	-0.61	6.43	14.40	27.99	13.56	31.66	-0.60	7.56	14.49
MP (2400 atoms), projection	4.87	6.54	8.97	26.15	15.39	32.89	4.92	7.27	8.99

Table 1. Comparison of the performance of some piecewise linear separation algorithms on an instantaneous stereophonic mixture of three sources. For each decomposition method, we applied the local separation strategy corresponding to Eq. (16) based on the true mixing matrix. We measured $\text{SDR}_i := \text{SDR}(s_i, \widehat{s}_i)$ as well as the corresponding SIR and SAR figures. The performance of the best global linear separation is indicated as a reference.

5.1. Evaluation criteria

It is well known that blind source separation can only recover the sources up to gain and permutation. Letting apart the permutation problem, because of the unknown gain factor, a plain SNR $10\log_{10}||s||^2/||s-\widehat{s}||^2$ does not display correctly the separation performance, and many authors instead use the SNR between $\widehat{s}/||\widehat{s}||$ and $s/||s||$. We showed in [6] that even this modified performance measure does not scale intuitively : in the worst case, when the estimate is orthogonal to the true source, it yields an SNR of only -3 dB. Instead, a slight modification called the Source to Distortion Ratio (SDR) [6] scales from $-\infty$ dB to $+\infty$ dB :

$$\texttt{SDR}(s,\widehat{s}) := 10\log_{10}\frac{\left|\left\langle \widehat{s}, s/||s|| \right\rangle\right|^2}{||\widehat{s}||^2 - \left|\left\langle \widehat{s}, s/||s|| \right\rangle\right|^2}. \tag{19}$$

Moreover, in the degenerate case, efficient BSS algorithms are nonlinear because a perfect globally linear separation is impossible. In [6] we proposed two other measures, the Source to Interference Ratio (SIR) and the Source to Artefacts Ratio (SAR) that measure respectively the amount of distortion due to remaining interferences of the unwanted sources, and the distortion due to nonlinearities in the algorithms, such as the thresholding effects. Thus, any algorithm that performs separation in a globally linear manner produces a SAR of $+\infty$dB. Matlab routines to compute this performance figures are available online [50].

5.2. Experiments

We implemented the multichannel Best Basis algorithm using Wavelab 802 [51] (our source code is available online [52]) and the multichannel MP algorithm using LastWave [53]. We performed a series of experiments where we used the following decomposition methods : every cosine packets and wavelet packets bases available in Wavelab; MP with a Gaussian multiscale Gabor dictionary, with dyadic scales ranging from 4 to 16384. Two local separation strategies were tested : the standard one (see Eq.(14)) where the masks are applied to one channel (the left one or the right one) which is chosen arbitrarily; the one we have derived from the disjoint orthogonal prior model (see Eq. (16)) where a projection on columns of the mixing matrix is used instead of an arbitrary choice. For each combination of decomposition method and local separation strategy we computed the figures SDR^l, SDR^r, SDR^p, SIR^l, SIR^r, SIR^p, SAR^l, SAR^r and SAR^p for each source, where l, r and p stand respectively for the *left*, *right* and *projection* local separation strategies. In addition, to serve as a baseline, we also computed the figures SDR^{lin}, SIR^{lin}, $\text{SAR}^{lin} = +\infty$ corresponding to the best globally linear separation [6].

5.3. Results

The results are summarized in Table 1. The first line of the table displays the best performance that can be expected from a globally linear separation strategy [6], which shows that this BSS problem cannot be solved acceptably without relying on a nonlinear BSS algorithm.

We performed a first series of experiments to compare multichannel BB with the 28 libraries of bases (local cosine packets and wavelet packets) that were available in Wavelab. A striking observation was the poor behaviour of the standard local separation strategy compared to the new one based on projection, for all BB decompositions. Over all BB decompositions and all sources, we observed that

$$\begin{aligned} \mathrm{SDR}^p &\geq \max(\mathrm{SDR}^l, \mathrm{SDR}^r) - 0.72\mathrm{dB} \\ \mathrm{SIR}^p &\geq \max(\mathrm{SIR}^l, \mathrm{SIR}^r) \ - 1.56\mathrm{dB} \\ \mathrm{SAR}^p &\geq \max(\mathrm{SAR}^l, \mathrm{SAR}^r) - 0.71\mathrm{dB} \end{aligned}$$

which shows that choosing the projection strategy never degrades significatively the performance compared to the usual method. On the contrary, for every BB decomposition and every source, $\min(\mathrm{SDR}^l, \mathrm{SDR}^r) < \mathrm{SDR}^p$ (similar relations hold for SIR and SAR), hence masking the "wrong" channel in the standard strategy leads to a systematic loss of performance compared to the projection strategy. The systematic loss can be non negligible : in the case of the third source (piano) we observed that, for every BB decomposition

$$\begin{aligned} \min(\mathrm{SDR}^l, \mathrm{SDR}^r) &\leq \mathrm{SDR}^p - 3.49\mathrm{dB} \\ \min(\mathrm{SIR}^l, \mathrm{SIR}^r) &\leq \mathrm{SIR}^p \ - 2.60\mathrm{dB} \\ \min(\mathrm{SAR}^l, \mathrm{SAR}^r) &\leq \mathrm{SAR}^p - 3.45\mathrm{dB}. \end{aligned}$$

The loss of performance was as bad as -5.25dB for the first source (cello) with the SDR measure.

Given these observations, the comparison between various BB decomposition can essentially be made by combining it with the projection strategy Eq. (16) for local separation. The best choice (at least on this dataset) is the wavepacket library based on the Battle 5 filter. In Table 1 we display the results obtained with the five libraries that were most often ranked in the five best performing : the cosine packets, and the wavelet packets based on the Beylkin, Vaidyanathan, Battle 3 and Battle 5 filters. Compared to the best globally linear separation, the improvement in performance achieved through piecewise linear separation with BB decomposition is clear : for the Battle 5 filter wave packets, not only does the SDR increase by up to more than 5dB for the cello and drums sources, but most of all the SIR SIR figures are improved by at least 16dB for the piano and more than 22dB for the other sources. As indicated by the SIR figures, these nonlinear BSS algorithms achieve a good rejection of the unwanted sources, but their nonlinearity introduces artefacts, which are indicated by the SAR figures. In fact, the SDR figures show that the distortion due to artefacts completely dominates the remaining interferences of the unwanted sources, *i.e.* we have SDR $\approx \min(\mathrm{SIR}, \mathrm{SAR}) = \mathrm{SAR}$. We believe the MMM local separation strategy will provide a good compromise between the rejection of the unwanted sources and the nonlinear artefacts.

In addition to the experiments with BB, we made an experiment based on stereo MP [11] with 2400 atoms. For the cello and drums sources, the comparison between the local separation strategy based on the projection method and the standard one leads to the very same observations as with BB, and it seems to be a good choice to rely on Eq. (16). The same observation is true in terms of SIR for the piano source, hence Eq. (16) never noticeably degrades the rejection of unwanted sources. However, for the piano source, the new strategy degraded the SAR (and the SDR which, again, is dominated by artefacts) by about -5.5dB compared to standard binary masking of the right channel. Table 1 displays the results respectively with standard masking applied on the left channel, on the right one and with the projection strategy.

The comparison of BB and MP in Table 1 leads to some interestingly observations. First, these piecewise linear separation algorithms obviously outperform any globally linear separation algorithm both in terms of global separation performance and rejection of the unwanted sources, but they logically introduce nonlinear artefacts. Indeed, the artefacts are currently the limiting factor for their overall performance, and decreasing the level of artefacts should improve the SDR figure. For applications where it is important to minimize the level of artefacts, with this dataset, one would probably perform the decomposition based on BB with the Battle 5 filter (with local separation based on Eq. (16)). If one needs to reject the unwanted sources, at the possible price of a higher level of artefacts, then the Battle 5 choice competes with MP : even if the latter degrades the SIR by about 7dB for the drums source, it increases it by more than 4dB for the cello and piano sources.

6. CONCLUSION

In this paper, we have proposed a new global framework, called *piecewise linear separation*, for blind source separation of possibly degenerate mixtures. The framework is based on a combination of a decomposition of the mixtures into elementary components and local linear separation strategies on each component. We have shown that it covers many existing BSS algorithms, as well as some denoising algorithms. We proposed new multichannel decomposition methods by generalizing the Best Basis (BB) and the Matching Pursuit (MP) algorithms, and new local separation strategies besides the classical binary masking strategy. We have performed experiments with various combinations of the proposed decomposition methods and local separation strategies, and showed that a new local separation strategy generally performs better than the standard one. On the dataset used for the experiments, it seems that BB with either cosine packets of wavelet packets (Beylkin, Vaidyanathan, Battle 3 or Battle 5 filter) are the best choices in terms of overall performance because they introduce a relatively low level of artefacts in the estimation of the sources; MP introduces slightly more artefacts, but can improve the rejection of the unwanted sources.

While this paper has focussed on the issue of recovering the sources from a degenerate mixture *assuming we know the mixing matrix* $\mathbf{A}$, this assumption cannot be made in true BSS problems. We are currently investigating the possibility to estimate $\mathbf{A}$ and the sources in a iterative way : given the current estimate $\widehat{\mathbf{A}}$, piecewise linear separation would provide estimates of the sources and a measure of likelihood of the most likely set of hidden activity states. An EM algorithm would be used to update the estimate of $\mathbf{A}$.

Piecewise linear separation improves the separation performance compared to globally linear separation, but its overall performance is limited by the artefacts it introduces. We believe it is possible to improve the overall performance (SDR) by reducing the level of artefacts (SAR) without degrading too much the ability to reject unwanted sources (SIR). With this aim, the use smooth forms of local separation such as Masking Mixture Models and Gaussian Mixture Models, is under investigation. Eventually, we believe it is worth building models of dependencies between the hidden activity states of different components : when N instruments are mixed in a commercial musical recording, the sources are temporally synchronized (the musicians play together) and have common harmonics (they are tuned together). Moreover, transients and harmonic lines give rise to quite structured time-frequency representations [35,54,55]. Hidden Markov Chains [12] and Hidden Markov Trees [56] are probably good models to investigate.

6.1. Acknowledgments

Many thanks to Thomas Reulos for his precious help in the implementation of multichannel MP with LastWave [53] and in the numerical experiments. The author is also very grateful to Laurent Benaroya, Frédéric Bimbot and Guillaume Gravier for their useful suggestions, and to Emmanuel Vincent, Cédric Févotte and for their comments on an early version of this paper.

REFERENCES

1. J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE. Special issue on blind identification and estimation* **9**, pp. 2009–2025, Oct. 1998.
2. E. Oja and M. Plumbley, "Blind separation of positive sources using non-negative PCA," in *Proc. 4th Int. Symp. on Independent Component Anal. and Blind Signal Separation (ICA2003)*, pp. 11–16, (Nara, Japan), Apr. 2003.
3. J.-F. Cardoso, "On the performance of orthogonal source separation algorithms," in *Proc. EUSIPCO*, pp. 776–779, (Edinburgh), Sept. 1994.
4. A. Taleb and C. Jutten, "On underdetermined source separation," in *Proc. ICASSP'99*, pp. 2089–2092, (Phoenix (AR, USA)), may 1999.
5. O. Bermond and J.-F. Cardoso, "Méthodes de séparation de sources dans le cas sous-déterminé," in *Proc. GRETSI, Vannes, France*, pp. 749–752, 1999.
6. R. Gribonval, L. Benaroya, E. Vincent, and C. Févotte, "Proposals for performance measurement in source separation," in *Proc. 4th Int. Symp. on Independent Component Anal. and Blind Signal Separation (ICA2003)*, pp. 763–768, (Nara, Japan), Apr. 2003.
7. M. Van Hulle, "Clustering approach to square and non-square blind source separation," in *IEEE Workshop on Neural Networks for Signal Processing (NNSP99)*, pp. 315–323, Aug. 1999.

8. A. Jourjine, S. Rickard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: Demixing n sources from 2 mixtures," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP'00)*, **5**, pp. 2985–2988, (Istanbul, Turkey), June 2000.
9. M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computations* **13**(4), pp. 863–882, 2001.
10. P. Kisilev, M. Zibulevsky, Y. Y. Zeevi, and B. A. Pearlmutter, "Multiresolution framework for blind source separation," Tech. Rep. CCIT Report # 317, Technion University, June 2001.
11. R. Gribonval, "Sparse decomposition of stereo signals with matching pursuit and application to blind separation of more than two sources from a stereo mixture," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP'02)*, (Orlando, Florida), May 2002.
12. S. Roweis, "One microphone source separation," in *Advances in Neural Information Processing Systems*, **13**, pp. 793–799, 2001.
13. L. Benaroya and F. Bimbot, "Wiener-based source separation with HMM/GMM using a single sensor," in *Proc. 4th Int. Symp. on Independent Component Anal. and Blind Signal Separation (ICA2003)*, (Nara, Japan), apr 2003.
14. L. Benaroya, L. McDonagh, F. Bimbot, and R. Gribonval, "Non negative sparse representation for wiener based source separation with a single sensor," in *Proc. IEEE Intl. Conf. Acoust. Speech Signal Process (ICASSP'03)*, pp. 613–616, (Hong-Kong), Apr. 2003.
15. L. Benaroya, *Séparation de plusieurs sources sonores avec un seul microphone*. PhD thesis, Université de Rennes I, Rennes, France, June 2003.
16. G.-J. Jang, L. T.-W., and Y.-H. Oh, "Single channel signal separation using maximum likelihood subspace projections," in *Proc. 4th Int. Symp. on Independent Component Anal. and Blind Signal Separation (ICA2003)*, pp. 529–534, (Nara, Japan), apr 2003.
17. F. J. Theis, C. Puntonet, and E. W. Lang, "A histogram-based overcomplete ICA algorithm," in *Proc. 4th Int. Symp. on Independent Component Anal. and Blind Signal Separation (ICA2003)*, pp. 1071–1076, (Nara, Japan), Apr. 2003.
18. K. Waheed and F. M. Salem, "Algebraic overcomplete independent component analysis," in *Proc. 4th Int. Symp. on Independent Component Anal. and Blind Signal Separation (ICA2003)*, pp. 1077–1082, (Nara, Japan), Apr. 2003.
19. L. De Lathauwer, B. De Moor, J. Vandewalle, and J.-F. Cardoso, "Independent component analysis of largely underdetermined mixtures," in *Proc. 4th Int. Symp. on Independent Component Anal. and Blind Signal Separation (ICA2003)*, pp. 29–33, (Nara, Japan), Apr. 2003.
20. L. Albera, A. Ferreol, P. Comon, and P. Chevalier, "Sixth order blind identification of underdetermined mixtures (BIRTH) of sources," in *Proc. 4th Int. Symp. on Independent Component Anal. and Blind Signal Separation (ICA2003)*, pp. 909–914, (Nara, Japan), Apr. 2003.
21. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1998.
22. R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory* **38**, pp. 713–718, Mar. 1992.
23. S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Process.* **41**, pp. 3397–3415, Dec. 1993.
24. S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing* **20**, pp. 33–61, Jan. 1999.
25. S. Rickard and O. Yilmaz, "On the approximate w-disjoint orthogonality of speech," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP'02)*, (Orlando, Florida), May 2002.
26. R. Balan and J. Rosca, "Statistical properties of STFT ratios for two channel systems and applications to blind source separation," in *Proc. of the Int'l. Workshop on Independent Component Analysis and Blind Signal Separation (ICA 2000)*, pp. 429–434, (Helsinki, Finland), June 2000.
27. D. Field and B. Olshausen, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature* **381**, pp. 607–609, 1996.
28. A. Bell and T. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vision Research* **37**(23), pp. 3327–3338, 1997.
29. A. Bell and T. Sejnowski, "Learning the higher-order structure of a natural sound," *Network* **7**(2), 1996.
30. M. Lewicki, "Efficient coding of natural sounds," *Nature Neuosci.* **5**(4), pp. 356–363, 2002.

31. S. Abdallah and M. Plumbley, "If edges are the independent components of natural images, what are the independent components of natural sounds?," in *Proc. of the Int. Conf. on Indep. Component Anal. and Blind Signal Separation (ICA2001)*, pp. 534–539, (San Diego, California), Dec. 2001.
32. K. Kreutz-Delgado, B. Rao, K. Engan, T.-W. Lee, and T. Sejnowski, "Convex/schur-convex (csc) log-priors and sparse coding," in *6th Joint Symposium on Neural Computation*, pp. 65–71, (Institute for Neural Computation), May 1999.
33. L. Benaroya, R. Gribonval, and F. Bimbot, "Représentations parcimonieuses pour la séparation de sources avec un seul capteur," in *GRETSI 2001*, (Toulouse, France), 2001. Article ♯ 434.
34. J.-L. Starck, M. Elad, and D. Donoho, "Image decomposition : Separation of textures from piecewise smooth content." presented at the GDR ISIS workshop on Texture, jun 2003.
35. L. Daudet, *Représentations structurelles de signaux audiophoniques : méthodes hybrides pour des applications à la compression*. PhD thesis, Université de Provence (Aix-Marseille I), 2000.
36. A. Gilbert, S. Muthukrishnan, and M. Strauss, "Approximation of functions over redundant dictionaries using coherence," in *The 14th ACM-SIAM Symposium on Discrete Algorithms (SODA'03)*, Jan. 2003.
37. J. Tropp, "Greed is good : Algorithmic results for sparse approximation," tech. rep., Texas Institute for Computational Engineering and Sciences, 2003. In preparation.
38. R. Gribonval and M. Nielsen, "Approximation with highly redundant dictionaries," in *Wavelets: Applications in Signal and Image Processing X, Proc. SPIE '03*, M. Unser, A. Aldroubi, and A. F. Laine, eds., **5207**, (San Diego, CA), aug 2003.
39. R. Gribonval and P. Vandergheynst, "Exponential convergence of Matching Pursuit in quasi-incoherent dictionaries," tech. rep., IRISA, 2003. in preparation.
40. D. Leviatan and V. Temlyakov, "Simultaneous approximation by greedy algorithms," Tech. Rep. 0302, IMI, Dept of Mathematics, University of South Carolina, Columbia, SC 29208, 2003.
41. D. Donoho and I. Johnstone, "Ideal denoising in an orthonormal basis chosen from a library of bases," *Comptes-Rendus Acad. Sci. Paris Série I* **319**, pp. 1317–1322, 1994.
42. D. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Inform. Theory* **41**, pp. 613–627, May 1995.
43. R. Gribonval, E. Bacry, S. Mallat, P. Depalle, and X. Rodet, "Analysis of sound signals with high resolution matching pursuit," in *Proc. IEEE Conf. Time-Freq. and Time-Scale Anal. (TFTS'96)*, pp. 125–128, (Paris, France), June 1996.
44. R. Gribonval, *Temps-fréquence : concepts et outils*, ch. Analyse temps-fréquence linéaire I : Reprsentations type Fourier. Information–Commande–Communication (IC2), Hermès, 2003. to appear.
45. D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decompositions," *IEEE Trans. Inform. Theory* **47**, pp. 2845–2862, Nov. 2001.
46. M. Elad and A. Bruckstein, "A generalized uncertainty principle and sparse representations in pairs of bases," *IEEE Trans. Inform. Theory* **48**, pp. 2558–2567, Sept. 2002.
47. R. Gribonval and M. Nielsen, "Sparse decompositions in unions of bases," Tech. Rep. 1499, IRISA, Nov. 2002. submitted to IEEE Trans. Inf. Th.
48. D. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Aca. Sci.* **100**, pp. 2197–2202, Mar. 2003.
49. J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inf. Th.* , 2003. submitted.
50. Action Jeunes Chercheurs du GDR ISIS (CNRS), "Ressources pour la séparation de signaux audiophoniques." `http://www.ircam.fr/anasyn/ISIS/`.
51. D. Donoho, M. Duncan, X. Huo, O. Levi, *et al.*, *Wavelab 802 for Matlab5.x*. http://www-stat.stanford.edu/ wavelab/.
52. R. Gribonval, *Best Basis Blind Source Separation (B3S2) software*. http://www.irisa.fr/metiss/gribonval.
53. E. Bacry, *LastWave software (GPL license)*. http://wave.cmap.polytechnique.fr/soft/LastWave/.
54. G. García, P. Depalle, and X. Rodet, "Tracking of partial for additive sound synthesis using hidden Markov models," in *Proc. Int. Computer Music Conf. (ICMC'93)*, pp. 94–97, (Tokyo), 1993.
55. R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Process.* **51**, pp. 101–111, jan 2003.
56. M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based signal processing using hidden markov models," *IEEE Trans. Signal Process.(Special Issue on Wavelets and Filterbanks)* **46**, pp. 886–902, Apr. 1998.