Chapter 1

OPTIMIZATION ALGORITHMS FOR SPARSE REPRESENTATIONS AND APPLICATIONS

Pando G. Georgiev
Laboratory for Advanced Brain Signal Processing
Brain Science Institute, RIKEN, Wako-shi, Japan
Current address:
ECECS Department, University of Cincinnati, ML 0030, Cincinnati, Ohio 45221-0030, U.S.A.
pgeorgie@ececs.uc.edu

Fabian Theis
Institute of Biophysics, University of Regensburg
D-93040 Regensburg, Germany
fabian@theis.name

Andrzej Cichocki
Laboratory for Advanced Brain Signal Processing
Brain Science Institute, RIKEN, Wako-shi, Japan
cia@bsp.brain.riken.jp

Abstract We consider the following *sparse representation problem*, which is called Sparse Component Analysis: identify the matrices $\mathbf{S} \in \mathbb{R}^{n \times N}$ and $\mathbf{A} \in \mathbb{R}^{m \times n}$ $(m \leq n < N)$ uniquely (up to permutation of scaling), knowing only their multiplication $\mathbf{X} = \mathbf{AS}$, under some conditions, expressed either in terms of $\mathbf{A}$ and sparsity of $\mathbf{S}$ (*identifiability* conditions), or in terms of $\mathbf{X}$ (*Sparse Component Analysis* conditions). A crucial assumption (sparsity condition) is that $\mathbf{S}$ is *sparse of level* k in sense that each column of $\mathbf{S}$ has at least k nonzero elements $(k = 1, 2, ..., m-1)$.

We present two type of optimization problems for such identification. The first one is used for identifying the mixing matrix $\mathbf{A}$: this is a typical clustering type problem aimed to finding hyperplanes in $\mathbb{R}^m$ which contain the columns of $\mathbf{X}$. We present a general algorithm for this clustering problem and a modification of Bradley-Mangasarian's k-

planes clustering algorithm for data allowing reduction of this problem to an orthogonal one.

The second type of problems is those of identifying the source matrix **S**. This corresponds to finding a sparse solution of a linear system. We present a source recovery algorithm, which allows to treat underdetermined case.

Applications include Blind Signal Separation of under-determined linear mixtures of signals in which the sparsity is either given a priori, or obtained with some preprocessing techniques as wavelets, filtering, etc. We apply our orthogonal m-planes clustering algorithm to fMRI analysis.

Keywords : Sparse Component Analysis, Blind Source Separation, underdetermined mixtures

1. Introduction

One of the fundamental questions in data analysis, signal processing, data mining, neuroscience, etc. is how to represent a large data set $\mathbf{X}$ (given in form of a $(m \times N)$-matrix) in different ways. A simple approach is a linear matrix factorization:

$$\mathbf{X} = \mathbf{AS}, \quad \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{S} \in \mathbb{R}^{n \times N}, \tag{1}$$

where the unknown matrices $\mathbf{A}$ (dictionary) and $\mathbf{S}$ (source signals) have some specific properties, for instance:

1) the rows of $\mathbf{S}$ are (discrete) random variables, which are statistically independent as much as possible – this is *Independent Component Analysis* (ICA) problem;

2) $\mathbf{S}$ contains as many zeros as possible – this is the sparse representation or *Sparse Component Analysis* (SCA) problem;

3) the elements of $\mathbf{X}, \mathbf{A}$ and $\mathbf{S}$ are nonnegative – this is *Nonnegative Matrix Factorization* (NMF) (see [13]).

There is a large amount of papers devoted to ICA problems (see for instance [3], [11] and references therein) but mostly for the case $m \geq n$. We refer to [1, 12, 17, 18, 21] and reference therein for some recent papers on SCA and underdetermined ICA ($m < n$).

A related problem is the so called *Blind Source Separation* (BSS) problem, in which we know *a priori* that a representation such as in equation (1) exists and the task is to recover the sources (and the mixing matrix) as accurately as possible. A fundamental property of the complete BSS problem is that such a recovery (under assumptions in 1) and non-Gaussianity of the sources) is possible up to permutation and scaling of the sources, which makes the BSS problem so attractive.

In this paper we consider SCA and BSS problems in the underdetermined case ($m < n$, i.e. more sources than sensors, which is more

challenging problem), where the additional information compensating the limited number of sensors is the *sparseness* of the sources. It should be noted that this problem is quite general and fundamental, since the sources could be not necessarily sparse in time domain. It would be sufficient to find a linear transformation (e.g. wavelet packets), in which the sources are sufficiently sparse.

In the sequel, we present new algorithms for solving the BSS problem: matrix identification algorithm and source recovery algorithm under conditions that the source matrix $\mathbf{S}$ has at most $m-1$ nonzero elements in each column and if the identifiability conditions are satisfied (see Theorem 1). We demonstrate the effectiveness of our general matrix identification algorithm and the source recovery algorithm in the underdetermined case for 7 artificially created sparse source signals, such that the source matrix $\mathbf{S}$ has at most 2 nonzero elements in each column, mixed with a randomly generated (3×7) matrix. For a comparison, we present a recovery using l_1-norm minimization [4], [5], which gives signals that are far from the original ones. This implies that the conditions which ensure equivalence of l_1-norm and l_0-norm minimization [5], Theorem 7, are generally not satisfied for randomly generated matrices. Note that l_1-norm minimization gives solutions which have at most m non-zeros [4], [5]. Another connection with [5] is the fact that our algorithm for source recovery works "with probability one", i.e. for *almost all* data vectors $\mathbf{x}$ (in measure sense) such that the system $\mathbf{x} = \mathbf{As}$ has a sparse solution with less than m nonzero elements, this solution is unique, while in [5] the authors proved that for *all* data vectors $\mathbf{x}$ such that the system $\mathbf{x} = \mathbf{As}$ has a sparse solution with less than $Spark(\mathbf{A})/2$ nonzero elements, this solution is unique. Note that $Spark(\mathbf{A}) \leq m+1$, where $Spark(\mathbf{A})$ is the smallest number of linearly dependent columns of $\mathbf{A}$.

2. Blind Source Separation

In this section we develop a method for solving the BSS problem if the following assumptions are satisfied:
A1) the mixing matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ has the property that any square $m \times m$ submatrix of it is nonsingular;
A2) each column of the source matrix $\mathbf{S}$ has at most $m-1$ nonzero elements;
A3) the sources are sufficiently rich represented in the following sense: for any index set of $n-m+1$ elements $I = \{i_1, ..., i_{n-m+1}\} \subset \{1, ..., n\}$ there exist at least m column vectors of the matrix $\mathbf{S}$ such that each of

them has zero elements in places with indexes in I and each $m-1$ of them are linearly independent.

2.1. Matrix identification

We describe conditions in the sparse BSS problem under which we can identify the mixing matrix uniquely up to permutation and scaling of the columns. We give two type of such conditions. The first one corresponds to the least sparsest case in which such identification is possible. Further, we consider the most sparsest case (for small number of samples) as in this case the algorithm is much simpler.

2.1.1 General case – full identifiability.

Theorem 1. (Identifiability conditions – general case) *Assume that in the representation* $\mathbf{X} = \mathbf{AS}$ *the matrix* $\mathbf{A}$ *satisfies condition A1), the matrix* $\mathbf{S}$ *satisfies conditions A2) and A3) and only the matrix* $\mathbf{X}$ *is known. Then the mixing matrix* $\mathbf{A}$ *is identifiable uniquely up to permutation and scaling of the columns.*

Proof. It is clear that any column $\mathbf{a}_j$ of the mixing matrix lies in the intersection of all $\binom{n-1}{m-2}$ hyperplanes generated by those columns of $\mathbf{A}$ in which $\mathbf{a}_j$ participates.

We will show that these hyperplanes can be obtained by the columns of the data $\mathbf{X}$ under the condition of the theorem. Let $\mathcal{J}$ be the set of all subsets of $\{1,...,n\}$ containing $m-1$ elements and let $J \in \mathcal{J}$. Note that $\mathcal{J}$ consists of $\binom{n}{m-1}$ elements. We will show that the hyperplane (denoted by H_J) generated by the columns of $\mathbf{A}$ with indexes from J can be obtained by some columns of $\mathbf{X}$. By A2) and A3), there exist m indexes $\{t_k\}_{k=1}^m \subset \{1,...,N\}$ such that any $m-1$ vector columns of $\{\mathbf{S}(:,t_k)\}_{k=1}^m$ form a basis of the $(m-1)$-dimensional coordinate subspace of $\mathbb{R}^n$ with zero coordinates given by $\{1,...,n\}\backslash J$. Because of the mixing model, vectors of the form

$$\mathbf{v}_k = \sum_{j \in J} S(j,t_k)\mathbf{a}_j, \quad k = 1,...,m,$$

belong to the data matrix $\mathbf{X}$. Now, by condition A1) it follows that any $m-1$ of the vectors $\{\mathbf{v}_k\}_{k=1}^{m-1}$ are linearly independent, which implies that they will span the same hyperplane H_J. By A1) and the above, it follows that we can cluster the columns of $\mathbf{X}$ in $\binom{n}{m-1}$ groups $\mathcal{H}_k, k = 1,..., \binom{n}{m-1}$ uniquely such that each group $\mathcal{H}_k$ contains at least

m elements and they span one hyperplane H_{J_k} for some $J_k \in \mathcal{J}$. Now we cluster the hyperplanes obtained in such a way in the smallest number of groups such that the intersection of all hyperplanes in one group gives a single one-dimensional subspace. It is clear that such one-dimensional subspace will contain one column of the mixing matrix, the number of these groups is n and each group consists of $\binom{n-1}{m-2}$ hyperplanes. ∎

The proof of this theorem gives the idea for the matrix identification algorithm.

SCA matrix identification algorithm
Data: samples $\mathbf{x}(1), \ldots, \mathbf{x}(T)$ of $\mathbf{X}$
Result: estimated mixing matrix $\hat{\mathbf{A}}$
Hyperplane identification.

1 Cluster the columns of $\mathbf{X}$ in $\binom{n}{m-1}$ groups $\mathcal{H}_k, k = 1, ..., \binom{n}{m-1}$ such that the span of the elements of each group $\mathcal{H}_k$ produces one hyperplane and these hyperplanes are different.
Matrix identification.
2 Cluster the normal vectors to these hyperplanes in the smallest number of groups $G_j, j = 1, ..., n$ (which gives the number of sources n) such that the normal vectors to the hyperplanes in each group G_j lie in a new hyperplane $\hat{H}_j$.
3 Calculate the normal vectors $\hat{\mathbf{a}}_j$ to each hyperplane $\hat{H}_j, j = 1, ..., n$.
4 The matrix $\hat{\mathbf{A}}$ with columns $\hat{\mathbf{a}}_j$ is an estimation of the mixing matrix (up to permutation and scaling of the columns).

Remark The above algorithm works for data for which we know a priori that they lie on hyperplanes (or near to hyperplanes).

2.2. Identification of sources

Theorem 2. (Uniqueness of sparse representation) *Let $\mathcal{H}$ be the set of all $\mathbf{x} \in \mathbb{R}^m$ such that the linear system $\mathbf{As} = \mathbf{x}$ has a solution with at least $n - m + k$ zero components. If $\mathbf{A}$ fulfills A1), then there exists a subset $\mathcal{H}_0 \subset \mathcal{H}$ with measure zero with respect to $\mathcal{H}$, such that for every $\mathbf{x} \in \mathcal{H} \setminus \mathcal{H}_0$ this system has no other solution with this property.*

Proof. Obviously $\mathcal{H}$ is the union of all $\binom{n}{m-k} = \frac{n!}{(m-k)!(n-m+k)!}$ k-codimensional linear subspaces of $\mathbb{R}^m$ (which are hyperplanes if $k = 1$), produced by taking the linear hull of every subsets of the columns of $\mathbf{A}$ with $m - k$ elements. Let $\mathcal{H}_0$ be the union of all intersections of any two such subspaces. Then $\mathcal{H}_0$ has a measure zero in $\mathcal{H}$ and satisfies the conclusion of the theorem. Indeed, assume that $\mathbf{x} \in \mathcal{H} \setminus \mathcal{H}_0$ and

$\mathbf{As} = \mathbf{A}\bar{\mathbf{s}} = \mathbf{x}$, where $\mathbf{s}$ and $\bar{\mathbf{s}}$ are have at least $n - m + k$ zeros. Since $\mathbf{x} \notin \mathcal{H}_0$, $\mathbf{x}$ belongs to only one k-codimensional linear subspace produced as a linear hull of some $m-k$ columns $\mathbf{a}_{i_1}, ..., \mathbf{a}_{i_{m-k}}$ of $\mathbf{A}$. It means that the vectors $\mathbf{s}$ and $\bar{\mathbf{s}}$ have $n - m + k$ zeros in places with indexes in $\{1, ..., n\} \setminus \{i_1, ..., i_{m-k}\}$. Now from the equation $\mathbf{A}(\mathbf{s} - \bar{\mathbf{s}}) = 0$ it follows that the $m-k$ vector columns $\mathbf{a}_{i_1}, ..., \mathbf{a}_{i_{m-k}}$ of $\mathbf{A}$ are linearly dependent, which is a contradiction with A1). ∎

From Theorem 2 it follows that the sources are identifiable generically, i.e. up to a set with a measure zero, if they have level of sparseness grater than or equal to $n - m + 1$ (i.e., each column of $\mathbf{S}$ has at least $n - m + 1$ elements) and the mixing matrix is known. Below we present an algorithm, based on the observation in Theorem 2. Note that this theorem is used in the sequel only for the case $k = 1$.

Source Recovery Algorithm
Data: samples $\mathbf{x}(1), \ldots, \mathbf{x}(N)$ (vector columns) of the data matrix $\mathbf{X}$, and mixing matrix $\mathbf{A}$
Result: estimated source matrix $\mathbf{S}$

1 Identify the the set of k-codimensional subspaces $\mathcal{H}$ produced by taking the linear hull of every subsets of the columns of $\mathbf{A}$ with $m - k$ elements;
2 Repeat for $i = 1$ to N:
3 Identify the subspace $H \in \mathcal{H}$ containing $\mathbf{x}_i := \mathbf{X}(:, i)$, or, in practical situation with presence of noise, identify the one to which the distance from $\mathbf{x}_i$ is minimal and project $\mathbf{x}_i$ onto H to $\tilde{\mathbf{x}}_i$;
4 if H is produced by the linear hull of column vectors $\mathbf{a}_{i_1}, ..., \mathbf{a}_{i_{m-k}}$, then find coefficients $\lambda_{i,j}$ such that

$$\tilde{\mathbf{x}}_i = \sum_{j=1}^{m-k} \lambda_{i,j} \mathbf{a}_{i_j}.$$

These coefficients are uniquely determined if $\tilde{\mathbf{x}}_i$ doesn't belong to the set $\mathcal{H}_0$ with measure zero with respect to to $\mathcal{H}$ (see Theorem 2);
5 Construct the solution $\mathbf{S}(:, i)$: it contains $\lambda_{i,j}$ in the place i_j for $j = 1, ..., m - k$, the rest of its components are zero.

3. Sparse Component Analysis

In this section we develop a method for the complete solution of the SCA problem. Now the conditions are formulated only in terms of the data matrix $\mathbf{X}$.

Theorem 3. (SCA conditions) *Assume that $m \leq n \leq N$ and the matrix $\mathbf{X} \in \mathbb{R}^{m \times N}$ satisfies the following conditions:*

(i) the columns of $\mathbf{X}$ lie in the union $\mathcal{H}$ of $\binom{n}{m-1}$ different hyperplanes, each column lies in only one such hyperplane, each hyperplane contains at least m columns of $\mathbf{X}$ such that each $m-1$ of them are linearly independent.

(ii) for each $i \in \{1, ..., n\}$ there exist $p = \binom{n-1}{m-2}$ different hyperplanes $\{H_{i,j}\}_{j=1}^{p}$ in $\mathcal{H}$ such that their intersection $L_i = \cap_{k=1}^{p} H_{i,j}$ is one dimensional subspace.

(iii) any m different L_i span the whole $\mathbb{R}^m$.

Then the matrix $\mathbf{X}$ is representable uniquely (up to permutation and scaling of the columns of $\mathbf{A}$ and $\mathbf{S}$) in the form $\mathbf{X} = \mathbf{AS}$, where the matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{S} \in \mathbb{R}^{n \times N}$ satisfy the conditions A1) and A2), A3) respectively.

Proof. Let L_i be spanned by $\mathbf{a}_i$ and set $\mathcal{A} = \{\mathbf{a}_i\}_{i=1}^{n}$. Condition (iii) implies that any hyperplane from $\mathcal{H}$ contains at most $m-1$ vectors from $\mathcal{A}$. By (i) and (ii) it follows that these vectors are exactly $m-1$: only in this case the calculation of the number of all hyperplanes by (ii) will give the number in (i): $n\binom{n-1}{m-2}/(m-1) = \binom{n}{m-1}$. Let $\mathbf{A}$ be a matrix whose column vectors are all vectors from $\mathcal{A}$ (taken in an arbitrary order). Since every column vector $\mathbf{x}$ of $\mathbf{X}$ lies only in one hyperplane from $\mathcal{H}$, the linear system $\mathbf{As} = \mathbf{x}$ has unique solution, which has at least $n - m + 1$ zeros (see the proof of Theorem 2). Let $\{\mathbf{x}_i\}_{i=1}^{m}$ be m column vectors from $\mathbf{X}$, which span one hyperplane from $\mathcal{H}$, and $m-1$ of them are linearly independent (such vectors exist by (i)). Then we have: $\mathbf{As}_i = \mathbf{x}_i$, for some uniquely determined vectors $\mathbf{s}_i, i = 1, ..., m-1$, which are linearly independent and have at least $n - m + 1$ zeros in the same coordinates. In such a way we can write: $\mathbf{X} = \mathbf{AS}$ for some uniquely determined matrix $\mathbf{S}$, which satisfies A2) and A3). ∎

4. Skeletons of a finite set of points

Let X be a finite set of points represented by the columns $\{\mathbf{x}_j\}_{j=1}^{N}$ of the matrix $\mathbf{X} \in \mathbb{R}^{m \times N}$. The solution $\{(\mathbf{n}_i^0, b_i^0)\}_{i=1}^{n}$ of the following minimization problem:

$$\text{minimize} \sum_{j=1}^{N} \min_{1 \leq i \leq k} |\mathbf{n}_i^T \mathbf{x}_j - b_i| \quad \text{subject to} \quad \|\mathbf{n}_i\| = 1, b_i \in \mathbb{R}, i = 1, ..., n, \tag{2}$$

defines $n^{(1)}$-*skeleton* of $\mathbf{X}$ (introduced in [16]). It consists of a union of n hyper-planes $H_i = \{x \in \mathbb{R}^m : \mathbf{n}_i^T \mathbf{x} = b_i\}, i = 1, ..., k$, such that the sum of minimum distances of every point $\mathbf{x}_j$ to them is minimal.

Analogically, the solution of the following minimization problem:

$$\text{minimize} \sum_{j=1}^{N} \min_{1 \le i \le k} |\mathbf{n}_i^T \mathbf{x}_j - b_i|^2 \quad \text{subject to} \quad \|\mathbf{n}_i\| = 1, b_i \in \mathbb{R}, i = 1, ..., n, \tag{3}$$

defines $n^{(2)}$-*skeleton* of $\mathbf{X}$ (introduced in [2]). It consists of union of n hyper-planes $\{H_i\}_{i=1}^n$ (defined as above) such that the sum of squared minimum distances of every point $\mathbf{x}_j$ to them is minimal.

Assuming that the matrixes $\mathbf{A}$ and $\mathbf{S}$ satisfy conditions A1) and A3), it is clear by Theorem 1 that the representation $\mathbf{X} = \mathbf{AS}$ is sparse (in sense that each column of $\mathbf{S}$ contains at most $m - 1$ non-zero elements, i.e. it satisfies the condition A2)) if and only if the above defined two skeletons coincide and the data points (columns of $\mathbf{X}$) lie on them.

5. Orthogonal m-planes clustering algorithm

In this section we propose a modification of the k-plane clustering algorithm of Bradley and Mangasarian [2]. The idea is to reduce the problem of finding the m-skeleton of $\mathbf{X}$ to an orthogonal problem: requiring that the hyperplanes of it are orthogonal, i.e. defined by an orthonormal matrix $\mathbf{W} \in \mathbb{R}^{m \times m}$. This can be done in the following way, if we we assume that the source matrix $\mathbf{S}$ after normalization is semi-orthogonal, i.e. $\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T = \mathbf{I}$.

Let $\mathbf{XX}^T = \mathbf{ULU}^T$ be the eigenvalue decomposition of the matrix $\mathbf{XX}^T$. Assume that the diagonal elements of $\mathbf{L}$ are positive. Then, denoting $\mathbf{W} = \mathbf{L}^{-1/2}\mathbf{U}^T\mathbf{A}$ and $\mathbf{Y} = \mathbf{L}^{-1/2}\mathbf{U}^T\mathbf{X}$, we have

$$\mathbf{Y} = \mathbf{W}\tilde{\mathbf{S}}, \quad \mathbf{WW}^T = \mathbf{I}, \quad \mathbf{YY}^T = \mathbf{I}. \tag{4}$$

Then the cluster update steps in Bradley-Mangasarian algorithm [2] can be unified in the following optimization problem with orthogonality constraints:

$$\text{minimize} \quad \sum_{i=1}^{m} \mathbf{w}_i^T \mathbf{Y}^{(i)} (\mathbf{Y}^{(i)})^T \mathbf{w}_i \tag{5}$$

$$\text{under constraints} \quad \mathbf{w}_i \mathbf{w}_j^T = \delta_{ij}, \tag{6}$$

where $\mathbf{Y}^{(i)}$ is the matrix with vector columns, which are elements of the i-th cluster.

Orthogonal m-planes clustering algorithm
Data: samples $\mathbf{x}(1), \ldots, \mathbf{x}(T)$ of $\mathbf{X}$
Result: estimated orthonormal mixing matrix $\mathbf{W}$ in (4)

1 Initialize randomly $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$ – orthonormal matrix.
for $j \leftarrow 1, \ldots, j_0$
Cluster assignment.
for $t \leftarrow 1, \ldots, T$
2 Add $\mathbf{x}(t)$ to cluster $\mathbf{Y}^{(i)}$, where i is chosen to minimize $|\mathbf{w}_i^\top \mathbf{x}(t)|$ (distance to hyperplane given by the i-th column of $\mathbf{W}$).
end
3 Exit if the mean distance to hyperplanes is smaller than some preset value.
Matrix update.
for $k \leftarrow 1, \ldots, n$
4 Define projection matrix $\mathbf{P}$ with rows consisting of an orthonormal basis of the orthogonal complement of $\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}$.
5 Calculate projected cluster covariance $\mathbf{C} \leftarrow \mathbf{P}\mathbf{Y}^{(i)}(\mathbf{Y}^{(i)})^\top \mathbf{P}^\top$
6 Choose eigenvector $\mathbf{v}_k$ of $\mathbf{C}$ corresponding to a minimal eigenvalue.
7 Set $\mathbf{w}_k \leftarrow \mathbf{P}^\top \mathbf{v}_k$.
end
end

The constant j_0 is chosen in practice sufficiently large. The finite termination of the algorithm is proved in [2], Theorem 3.7.

6. Applications

6.1. Computer simulation examples: Underdetermined case

We consider a mixture of 7 artificially created sources (see Fig. 1.2 left) – sparsified randomly generated signals with at least 5 zeros in each column – with a randomly generated mixing matrix with dimension 3×7.

Figure 1.1 gives the mixed signals together with a normalized scatterplot of the mixtures – the data lies in $21 = \binom{7}{2}$ hyperplanes.

Applying the underdetermined matrix recovery algorithm (Algorithm 1) to the mixtures gives the recovered mixing matrix perfectly well, up to permutation and scaling. Applying the source recovery algorithm (Algorithm 4) we recover the source signals up to permutation and scaling

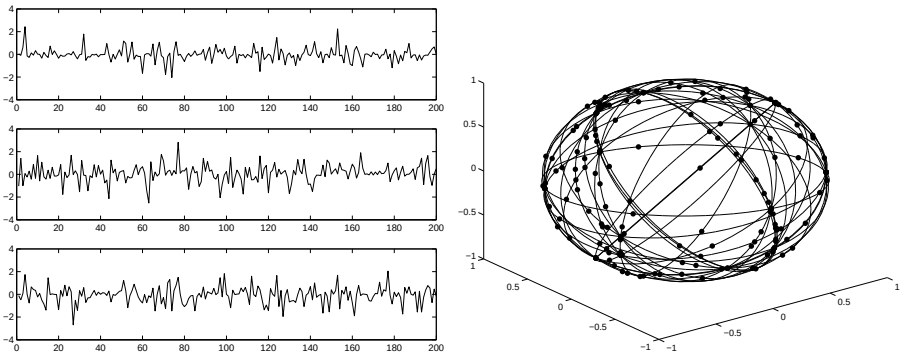


Figure 1.1. Mixed signals (left) and normalized scatter plot (density) of the mixtures (right) together with the 21 data set hyperplanes, visualized by their intersection with the unit sphere in $\mathbb{R}^3$.

(see Fig. 1.2, middle). This figure (right) shows also that the recovery by l_1-norm minimization (known as Basis Pursuit method of S. Chen, D. Donoho and M. Saunders [4]) does not perform well, even if the mixing matrix is perfectly known.

6.2. Complete case – orthogonal m-planes clustering algorithm applied to fMRI data

We now analyze the performance of the orthogonal m-planes clustering algorithm when applied to functional *magnetic resonance imaging (fMRI)* measurements.

The typical setup of fMRI experiments is the following: NMR brain imaging techniques are used to record brain activity data over a certain span of time, during which the subject is asked to perform some kind of task (e.g. 5 seconds of activity in the motor cortex followed by 5 seconds of activity in the visual cortex; this iterative procedure is often called *block diagram*). The brain recordings show areas of high and of low brain activity (using the *BOLD effect*). Analysis is performed on the 2d-image slices recorded at the discrete time steps. General linear model (GLM) approaches or ICA-based fMRI analysis then decompose this data set into a certain set of *component maps* i.e. sets of (independent) images that are active at certain time steps corresponding to the block diagram.

fMRI data were recorded from six subjects (3 female, 3 male, age 20–37) performing a visual task. In five subjects, five slices with 100 images (TR/TE = 3000/60 msec) were acquired with five periods of rest and

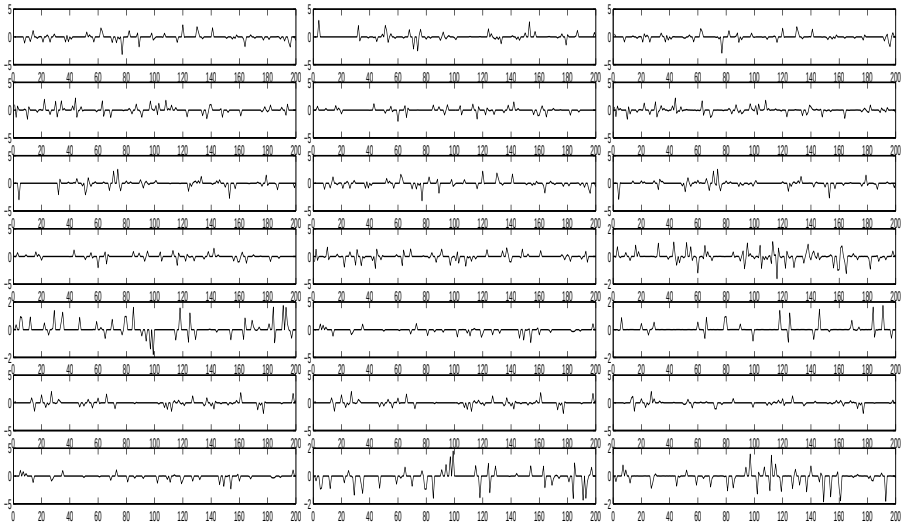


Figure 1.2. The original source signals are shown in the left column. The middle column gives the recovered source signals — the signal-to-noise ratio between the original sources and the recoveries is very high (above 278 dB after permutation and normalization). Note that only 200 samples are enough for excellent separation. The right column shows the recovered source signals using l_1-norm minimization and known mixing matrix. Simple comparison confirms that the recovered signals are far from the original ones – the signal-to-noise ratio is only around 4 dB.

five photic simulation periods with rest. Simulation and rest periods comprised 10 repetitions each, i.e. 30s. Resolution was $3\times3\times4$ mm. The slices were oriented parallel to the calcarine fissure. Photic stimulation was performed using an 8 Hz alternating checkerboard stimulus with a central fixation point and a dark background with a central fixation point during the control periods [19]. The first scans were discarded for remaining saturation effects. Motion artifacts were compensated by automatic image alignment (AIR, [20]).

Blind Signal Separation, mainly based on ICA, nowadays is a quite common tool in fMRI analysis (see for example [14], [15]). Here, we analyze the fMRI data set using as a separation criterion a spatial decomposition of fMRI data images to sparse component maps. Such an approach we consider as very reasonable and advantageous when the stimuli are sparse and dependent, and therefore the ICA methods couldn't give good results. Due to the availability of fMRI data, it appears that the results of our SCA method and ICA method give similar results, which itself

we consider as a surprising fact. Here we use our orthogonal m-planes clustering algorithm for matrix identification.

Figure 1.3 shows the performance of SCA method; see figure text for interpretation. Using only the first 9 principal components, our orthogonal m-planes clustering algorithm could recover the stimulus component as well as detect additional components. It performs equally well as fastICA [11], figure 1.4, which is interesting in itself: apparently the two different criteria, sparseness and independence, lead to similar results in this setting. This can be partially explained by noting that all components, mainly the stimulus component, have high kurtoses i.e. strongly peaked densities.

Acknowledgements

The authors would like to thank Dr. Dorothee Auer from the Max Planck Institute of Psychiatry in Munich, Germany, for providing the fMRI data, and Oliver Lange from the Department of Clinical Radiology, Ludwig-Maximilian University, Munich, Germany, for data preprocessing and visualization.

7. Conclusion

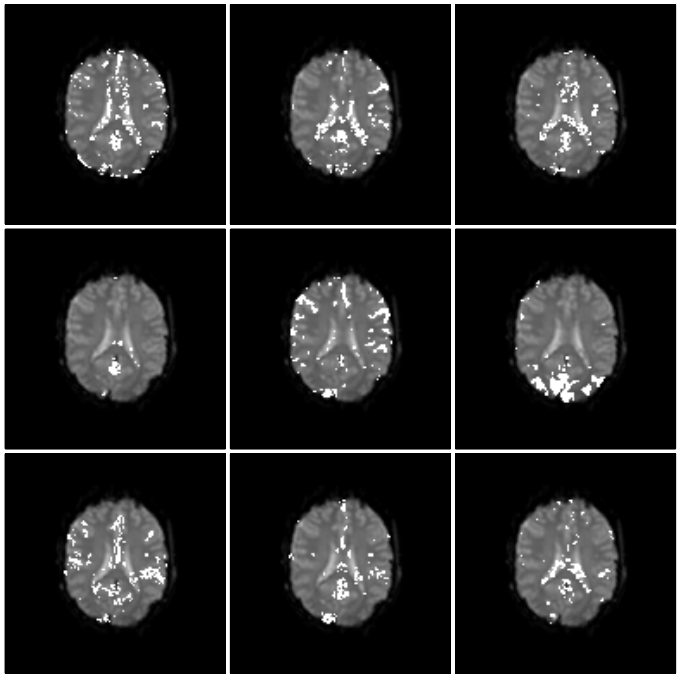
We defined rigorously Blind Signal Separation problem (BSS) and Sparse Component Analysis problem (SCA) of sparse signals and presented sufficient conditions for their solving. The main theoretical contributions are:

- Identifiability conditions for BSS – especially condition A3)
- Uniqueness of sparse representations up to a set with measure zero
- SCA conditions for sparse representation.

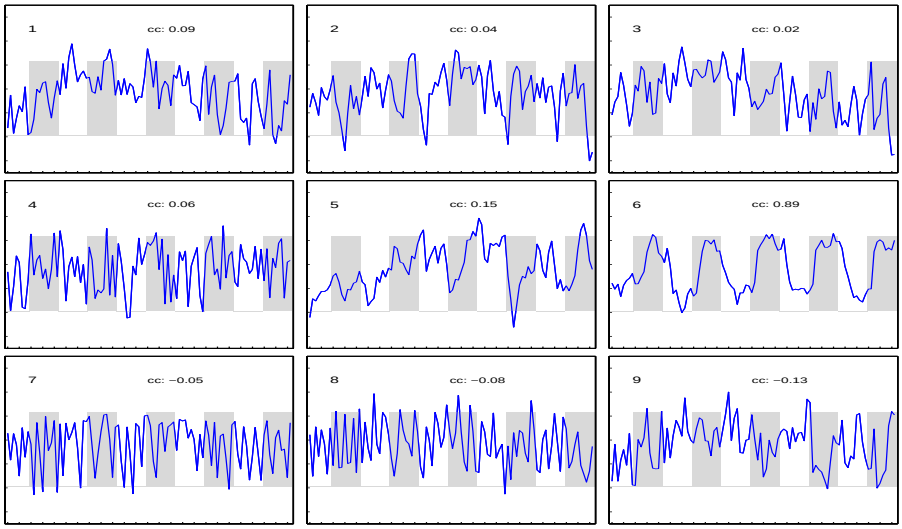
These theoretical results are supported by algorithmic implementations: SCA matrix identification algorithm and Source recovery algorithm.

The k-planes clustering algorithm of Bradley and Mangasarian is modified to an orthogonal one, which has superior performance.

Our SCA methods are illustrated with examples: computer simulated ones for the underdetermined case, and fMRI data analysis by our Orthogonal clustering algorithm.

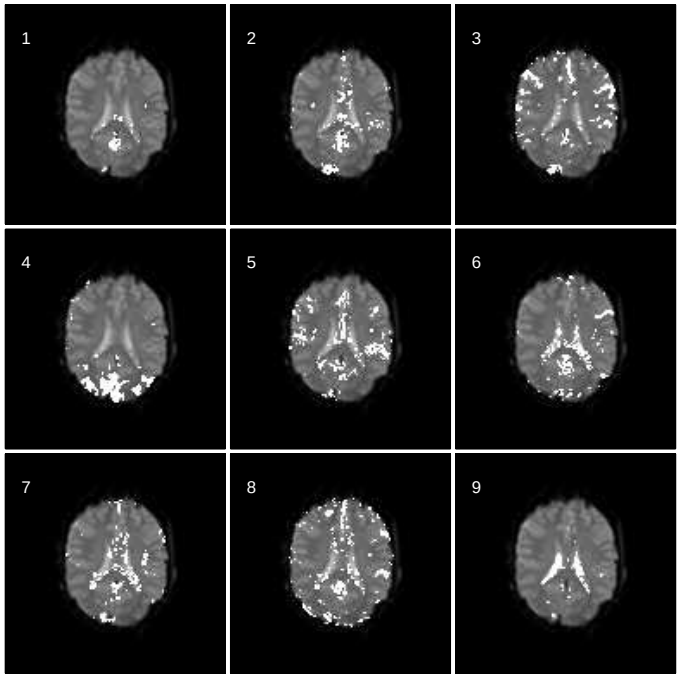


(a) component maps

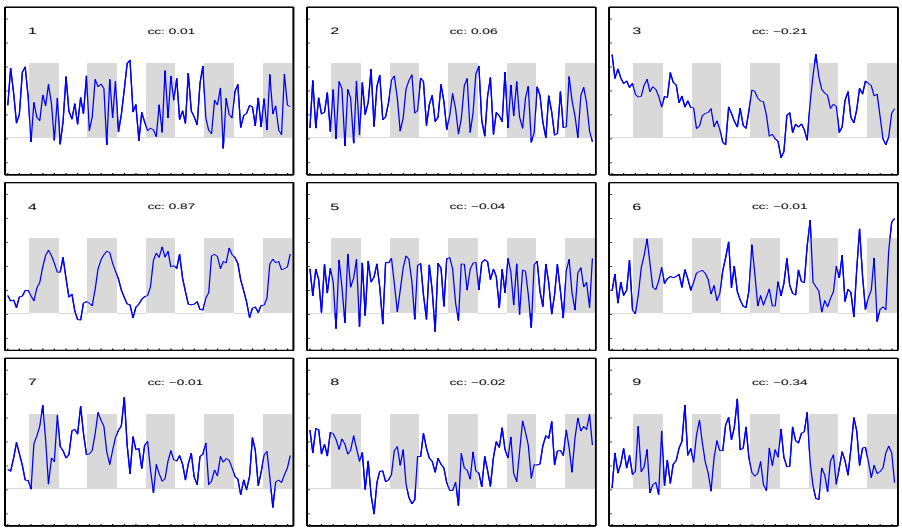


(b) time courses

Figure 1.3. fMRI analysis by our orthogonal m-planes clustering algorithm. The data was reduced to the first 9 principal components. (a) shows the recovered component maps (white points indicate values stronger than 3 standard deviations), and (b) their time courses. The stimulus component is given in component 6 (indicated by the high crosscorrelation $cc = 0.89$ with the stimulus time course, delayed by roughly 2 seconds due to the BOLD effect), which is strongly active in the visual cortex as expected.



(a) component maps



(b) time courses

Figure 1.4. FastICA result during fMRI analysis of the same data set as in figure 1.3. The stimulus component is given in component 4 with high stimulus cross-correlation $cc = 0.87$.

References

[1] P. Bofill and M. Zibulevsky, " Underdetermined Blind Source Separation using Sparse Representation", Signal Processing, vol. 81, no. 11, pp. 2353-2362, 2001.

[2] P.S. Bradley and O. L. Mangasarian, "k-Plane Clustering", *J. Global Optim.*, 16, (2000), no.1, 23-32.

[3] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. John Wiley, Chichester, 2002.

[4] S. Chen, D. Donoho and M. Saunders, " Atomic decomposition by basis pursuit", *SIAM J. Sci. Comput.*, Vol. 20, no. 1, pp. 33–61, 1998.

[5] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization", *Proc. Nat. Acad. Sci.*, vol.100, no.5, pp. 2197–2202, 2003.

[6] D. Donoho and V. Stodden, "When Does Non-Negative Matrix Factorization Give a Correct Decomposition into Parts?", Neural Information Processing Systems (NIPS) 2003 Conference, http://books.nips.cc

[7] P. G. Georgiev and A. Cichocki, "Sparse component analysis of overcomplete mixtures by improved basis pursuit method", in Proc. 2004 IEEE International Symposium on Circuits and Systems (ISCAS 2004), Vancouver, Canada, May 23–26.

[8] P.G. Georgiev, F. Theis and A. Cichocki, "Blind Source Separation and Sparse Component Analysis of overcomplete mixtures", in Proc. of International Conference on Acoustics and Statistical Signal Processing (ICASSP2004), Montreal, Canada, May 17-21, 2004.

[9] Georgiev P., Cichocki A., Bakardjian H., "Optimization Techniques for Independent Component Analysis with Applications to EEG

Data", In: Pardalos et al, editors, Quantitative Neuroscience: Models, Algorithms, Diagnostics, and Therapeutic Applications, Kluwer Academic Publishers, 2004, pp. 53-68.

[10] Gorodnitsky I, Belouchrani A., " Joint cumulant and correlation based signal separation with application of EEG data analysis" , in Proc. 3-rd Int. Conf. on Independent Component Analysis and Signal Separation, San Diego, California, Dec. 9-13, 2001, pp.475-480.

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

[12] T.-W. Lee, M.S. Lewicki, M. Girolami, T.J. Sejnowski, "Blind sourse separation of more sources than mixtures using overcomplete representaitons", *IEEE Signal Process. Lett.*, Vol. 6, no. 4, pp. 87–90, 1999.

[13] D. D. Lee and H. S. Seung "Learning the parts of objects by non-negative Matrix Factorization", *Nature*, Vol. 40, pp. 788–791, 1999.

[14] McKewon, M., Jung, T., Makeig, S., Brown, G., Kindermann, S., Bell, A., Sejnowski, T., "Analysis of fMRI data by blind separation into independent spatial components", *Human Brain Mapping* 6, 1998, pp. 160–188.

[15] M. McKeown, L. Hansen and T. Sejnowski, "Independent component analysis of functional MRI: what is signal and what is noise? *Current Opinion in Neurobiology* 2003, 13, pp. 620-629.

[16] A. M. Rubinov and J. Ugon, "Skeletons of finite sets of points", Research working paper 03/06, 2003, School of Information Technology & Mathematical Sciences, Univ. of Ballarat.

[17] F.J. Theis, E.W. Lang, and C.G. Puntonet, A geometric algorithm for overcomplete linear ICA. *Neurocomputing*, in print, 2003.

[18] K. Waheed, F. Salem, "Algebraic Overcomplete Independent Component Analysis", in *Proc. Int. Conf. ICA2003*, Nara, Japan, pp. 1077–1082.

[19] A. Wismüller, O. Lange, D. Dersch, G. Leinsinger, K. Hahn, B. Pütz and D. Auer, "Cluster Analysis of Biomedical Image Time–Series", *International Journal on Computer Vision*, Vol. 46, 2, 2002, pp.102–128.

[20] R. Woods and S. Cherry and J. Mazziotta, "Rapid automated algorithm for aligning and reslicing PET images", *Journal of Computer Assisted Tomography*, Vol. 16, 8, 1992, pp. 620–633.

[21] M. Zibulevsky, and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary", *Neural Comput.*, Vol. 13, no. 4, pp. 863–882, 2001.