数理解析研究所講究録 1365 巻 2004 年 83-94

Sparseness Theorems and Sparse Representation of Signals

PANDO GEORGIEV and ANDRZEJ CICHOCKI

Brain Science Institute, RIKEN
Lab. for Advanced Brain Signal Processing
2-1, Hirosawa, Wako-shi
Saitama, 351-0198, Japan
{georgiev, cia}@bsp.brain.riken.go.jp

Abstract

We present general sparseness theorems showing that the solutions of various types least square and absolute value optimization problems (linear with respect to l_2 and l_1 norm, non-linear ones) possess sparse solutions. These theorems have direct application to the problem of identification (up to scaling and permutation) of the source signals $\mathbf{S} \in \mathbb{R}^{n\times N}$ and the mixing matrix $\mathbf{A} \in \mathbb{R}^{m\times n}, m \leq n$, knowing only their mixture $\mathbf{X} = \mathbf{AS}$ - this is so called underdetermined *sparse component analysis (SCA)*. We present two new algorithms: for matrix identification (when the sources are very sparse), and for source recovery, improving in such a way the standard basis pursuit method of S. Chen, D. Donoho and M. Sounders (applied when the mixing matrix is known or correctly estimated). We illustrate our algorithms with examples.

1 Introduction

One of the fundamental questions in data analysis, signal processing, neuroscience, etc. is how to represent huge amount of data $\mathbf{X}$ (given in form of a matrix $(m \times N)$), for different tasks. A simple idea 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 $\mathbf{A} \in \mathbb{R}^{m\times n}$ (dictionary) and $\mathbf{S} \in \mathbb{R}^{n\times N}$ (signals) have some specific properties, for instance:

1) the rows of $\mathbf{S}$ 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 sparse representation problem 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), with several potential applications including decomposition of objects into "natural" components, learning the parts of the objects (e.g. learns from set of faces the parts a face consists of, i.e. eyes, nose, mouth, etc.), redundancy and dimensionality reduction, microarray data mining, enhancement of images in nuclear medicine, etc. (see [14]).

There is a large amount of papers devoted to ICA problems (see for instance [7], [11] and references therein) but mostly in the complete case (when the number of sources is equal to the number of sensors). We refer to [12], [4], [13], [2] and reference therein for some recent papers on SCA and underdetermined ICA.

A slightly different problem is so called Blind Source Separation (BSS) problem, in which we know a priory that such a representation like (1) exists and the task is to recover the sources and the mixing matrix as correctly as possible. A fundamental property of BSS problem (which makes it so attractive) under assumptions in 1) and non-Gaussianity of the sources, is that such recovering is possible up to permutation and scaling of the sources.

In this paper we present general sparseness theorems and apply some of them for sparse representation of signals for the underdetermined case (more sources than sensors). So, we consider the BSS problem in the underdetermined case, as additional information compensating the lack of sources is *sparseness*. We describe conditions under which it is possible to estimate the unknown sources $\mathbf{S}$ and the mixing matrix $\mathbf{A}$ uniquely (up to permutation and scaling of the sources, which is usual condition in the complete BSS problems).

We present a new algorithm for identification of the mixing matrix, which works correctly under some conditions (see conditions (i) and (ii) of Theorem 7).

We develop also an improvement of the basis pursuit method of Chen, Donoho and Saunders [8], (which in fact is l_1 norm minimization problem), when the mixing matrix is known or estimated. This improvement is also reduced to a linear programming problem, but we are able to find the sparsest solution of a linear underdetermined system. We present examples which illustrate our methods.

We introduce an optimization problem with nonnegativity constraints with respect to l_1 norm (see Section 4). It appears that it gives also sparse representations and is suitable for large scale problems, since it can be converted to a linear programming problem.

We present several computer simulation examples which confirm the good performance of our algorithms.

2 Nonlinear sparseness theorem

Our first theorem gives even an idea for nonlinear sparse codding. The key idea is from a *night sky* theorem (see Byrne [5]).

Consider the nonlinear least square problem with nonnegativity constraints:

$$\begin{aligned} &\text{minimize} && l(\mathbf{x}) = \|F(\mathbf{x}) - \mathbf{b}\|_2^2, && (2)\\ &\text{subject to} && x_i \geq 0, && (3) \end{aligned}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m, m \leq n$ is a differentiable mapping such that

$$\det\left(\frac{\partial F_i(\mathbf{x})}{\partial x_j}\Big|_{i=1,j\in S}^{m}\right) \neq 0 \tag{4}$$

for every subset S of the set $\{1,...,n\}$ with m elements.

Theorem 1 (Nonlinear sparseness theorem) *Assume that $l_{min} > 0$ (i.e. the equation $F(\mathbf{x}) = \mathbf{b}$ has no nonnegative solution). Then for any solution $\hat{\mathbf{x}}$ of (2), (3) with conditions (4), the subset $S_{\hat{\mathbf{x}}} = \{j \in \{1,...,m\} : \hat{x}_j > 0\}$ has at most $m - 1 - k$ elements, where k is the number of indexes i such that $F_i(\hat{\mathbf{x}}) - b_i = 0$.*

Proof. By the Fritz John theorem [3] there exist Lagrange multipliers $\mu_i \geq 0$, such that

$$2\sum_{i=1}^{m}[F_i(\hat{\mathbf{x}}) - b_i]\frac{\partial F_i(\hat{\mathbf{x}})}{\partial x_j} + \mu_j = 0, \qquad \forall j = 1, ..., n, \tag{5}$$

$$\mu_j \hat{x}_j = 0, \quad \forall j. \tag{6}$$

Assume that $|S_{\hat{\mathbf{x}}}| \geq m - k$. By (6) it follows that $\mu_j = 0$ for $j \in S_{\hat{\mathbf{x}}}$ and by (5), using the non-degeneracy condition (4), we obtain that $F(\hat{\mathbf{x}}) = \mathbf{b}$, a contradiction. Therefore $|S_{\hat{\mathbf{x}}}| \leq m - k - 1$. ∎

3 The linear case

In the linear case the mapping F is represented by a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. In this case the theorem is known (see [5]) and uniqueness of the solution is guaranteed.

Consider the linear least square problem with non-negativity constraints:

$$\text{minimize} \qquad l(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_2^2, \tag{7}$$

$$\text{subject to} \qquad x_i \geq 0, \quad i = 1, ..., n, \tag{8}$$

where $\mathbf{A} \in \mathbb{R}^{m \times N}, m < N$ is a matrix such that any submatrix $(m \times m)$ of it has full rank.

Theorem 2 (Night Sky theorem [5]) *Assume that $l_{min} > 0$ (i.e. the equation $\mathbf{Ax} = \mathbf{b}$ has no nonnegative solution). Then the solution of (7), (8) is unique, say $\hat{\mathbf{x}}$, and contains at most $m - 1 - k$ non-zero elements, where k is the number of indexes i such that $\sum_{j=1}^{n} \hat{x}_j = b_i$.*

In order to apply it we need to verify the condition that the system of linear equations $\mathbf{Ax} = \mathbf{b}$ has no solutions. This condition is equivalent to the condition that $\mathbf{b} \notin \mathbf{A}(\mathbf{K}_1)$, where $\mathbf{K}_1$ means the first octant:

$$\mathbf{b} \notin \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i : \alpha_i \geq 0\}, \tag{9}$$

i.e. this condition says that $\mathbf{b}$ does not belong to the cone generated by the columns of $\mathbf{A}$. This condition can be violated easily with enlarging the dimension of the problem, as we proceed below.

Let $\mathbf{e}$ be a unit length vector and $\hat{\mathbf{x}}_\mathbf{e}$ be a solution of the minimization problem

$$\text{minimize} \quad \mathbf{e}^T\mathbf{x} \quad \text{under constraint} \quad \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}. \tag{10}$$

The following theorem is direct consequence from Theorem 2.

Theorem 3 *For almost all unit length nonnegative vectors $\mathbf{e}$ (in measure sense) the solution $\hat{\mathbf{x}}_\mathbf{e}$ of (10) is unique and sparse (i.e. contains at most m nonzero elements), and*

$$\hat{\mathbf{x}}_\mathbf{e} = \lim_{\varepsilon \to 0_+} \mathbf{x}_\varepsilon,$$

where $\mathbf{x}_\varepsilon$ is the unique solution of the problem

$$\text{minimize} \quad \left\| \begin{pmatrix} \mathbf{A} \\ -\varepsilon \mathbf{e}^T \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix} \right\|_2^2, \tag{11}$$

$$\text{under constraints} \quad \mathbf{x} \geq \mathbf{0}. \tag{12}$$

Basic question: how to find $\mathbf{e}$ such that the solution of (10) (equivalently the solution of (11), (12) is the sparsest possible?

The **basis pursuit** method of Chen, Donoho and Sounders [8]attempts to answer to this question (without constraints on $\mathbf{x}$). It consists of minimization of the l_1 norm of $\mathbf{x}$ under constraint $\mathbf{Ax} = \mathbf{b}$.

In case of nonnegativity constraints, this is a particular case of the minimization problem (10), when all components of the vector $\mathbf{e}$ are equal to one.

Theorem 4 (Donoho, Elad [10]) *The vector* $\mathbf{x}_0$ *is the unique sparsest solution of* $\mathbf{Ax} = \mathbf{b}$ *if* $\|\mathbf{x}_0\| < Spark(\mathbf{A})/2$.

$Spark(\mathbf{A})$ is the minimum number of columns of $\mathbf{A}$ which are linearly dependent.

Observation: For almost all matrices $m \times n$ (in measure sense),

$$Spark(\mathbf{A}) = m + 1.$$

We are interested in the case, when the sparsest solution has less than $(m+1)/2$ nonzero elements.

We propose the following solution of the basic question: solve (10) n times, after setting consequently the coefficients $e_i = 0, i = 1, ..., n$. If no solution has less than $(m+1)/2$ nonzero components, set couple of coefficients $e_i = 0, e_j = 0$, solve (10) and so on, until obtaining the solution with less than $(m+1)/2$ nonzero components.

The following theorem describes conditions under which the l_1-minimization gives the sparsest solution, but in practical problems these conditions are rarely satisfied (as we will see in the sequel).

Theorem 5 (Donoho, Elad [10]) *Suppose that the off-diagonal elements of the matrix* $\mathbf{A}^T\mathbf{A}$ *are bounded by* M. *If* $\mathbf{Ax}_0 = \mathbf{b}$ *and* $\|\mathbf{x}_0\|_0 < (1 + 1/M)/2$, *then* $\mathbf{x}_0$ *is the unique sparsest solution of* $\mathbf{Ax} = \mathbf{b}$ *and is the unique solution of* l_1*-norm minimization problem: minimize* $\|\mathbf{x}\|_1$ *subject to* $\mathbf{Ax} = \mathbf{b}$.

4 Optimization with respect to l_1 norm

Consider the least square problem with nonnegativity constraints with respect to l_1 norm:

$$\text{minimize} \qquad l(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_1 = \sum_{i=1}^{m} \Big| \sum_{j=1}^{n} a_{ij}x_j - b_i \Big|, \tag{13}$$

$$\text{subject to} \qquad x_j \geq 0, \quad j = 1, ..., n, \tag{14}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a matrix, $m < n$, with the following properties:

(P1) if we remove any k columns, where $k \geq n - m$, then the remaining submatrix $\mathbf{A}_k$ is of full column rank;

(P2) if we take $n - k - 1$ rows of $\mathbf{A}_k$, then their linear hull does not contain any sum of the remaining rows multiplied with coefficients ± 1.

It can be proved that most of the matrices (in the measure sense) in $\mathbb{R}^{m \times n}$ have these properties.

Theorem 6 (l_1**-norm sparseness theorem**) *Assume that* $l_{min} > 0$ *(i.e. the equation* $\mathbf{Ax} = \mathbf{b}$ *has no nonnegative solutions). Then for any solution* $\hat{\mathbf{x}}$ *of (13), (14) the set* $S_{\hat{\mathbf{x}}} = \{j \in \{1,...,m\} : \hat{x}_j > 0\}$ *has the properties:*

(i) $S_{\hat{\mathbf{x}}}$ *has at most* $m-1$ *elements;*

(ii) the number of the elements of $S_{\hat{\mathbf{x}}}$ *is less than or equal to the number of indexes* i *for which* $\sum_{j=1}^{n} a_{ij}x_j - b_i = 0$.

Proof. (i) By the necessary conditions for optimality, applied for the problem (13), (14) (see for instance, [6]) there exist Lagrange multipliers $\mu_i \geq 0$, such that

$$2\sum_{i=1}^{m} c_i a_{ij} + \mu_j = 0 \qquad \forall j = 1,...,n \tag{15}$$

$$\mu_j \hat{x}_j = 0 \quad \forall j = 1,...,n, \tag{16}$$

where $c_i \in \partial|\sum_{j=1}^{n} a_{ij}\hat{x}_j - b_i|$, and $\partial|\sum_{j=1}^{n} a_{ij}\hat{x}_j - b_i|$ means the subdifferential of the function $|.|$ at the point $\sum_{j=1}^{n} a_{ij}\hat{x}_j - b_i$. Assume that $|S_{\hat{\mathbf{x}}}| \geq m$. By (16) it follows that $\mu_j = 0$ for $j \in S_{\hat{\mathbf{x}}}$ and by (15), using the non-degeneracy condition (P1), we obtain that $c_i = 0$ for every $i = 1,...,m$. Now we use the following property of the subdifferential:

$$\partial|t| = \begin{cases} [-1,1] & \text{if } t = 0 \\ +1 & \text{if } t > 0 \\ -1 & \text{if } t < 0. \end{cases}$$

Applying this property for $t_i = \sum_{j=1}^{n} a_{ij}\hat{x}_j - b_i$, having in mind that $0 \in \partial|t_i|$, we obtain $t_i = 0$ for every $i = 1,...,m$, a contradiction with the assumption that the system $\mathbf{Ax} = \mathbf{b}$ has no nonnegative solution.

(ii) Let the number of the nonzero elements of $S_{\hat{\mathbf{x}}}$ be k. Again by (15), using now the non-degeneracy condition (P2), we obtain that $c_i \in (0,1)$ for every i from an index set $I \subset \{1,...,m\}$ with k elements, which implies that $\sum_{j=1}^{n} a_{ij}\hat{x}_j - b_i = 0$ for $i \in I$. ∎

We can reduce the optimization problem considered in the previous section to a linear programming problem, by two ways.

(I) minimize $\sum_{i=1}^{m} u_i$

under constraints

$$u_i \geq \sum_{j=1}^{n} a_{ij}x_j - b_i \tag{17}$$

$$u_i \geq -\sum_{j=1}^{n} a_{ij}x_j - b_i \tag{18}$$

$$x_j \geq 0. \tag{19}$$

(II) minimize $\sum_{i=1}^{m} u_i^+ + u_i^-$

under constraints

$$u_i^+ - u_i^- = \sum_{j=1}^{n} a_{ij}x_j - b_i \tag{20}$$

$$u_i^+ \geq 0, \quad u_i^- \geq 0, \quad x_j \geq 0. \tag{21}$$

5 Matrix identification

In this section we describe conditions under which we can identify the mixing matrix in a sparse BSS problem.

Theorem 7 (Identifiability conditions – locally very sparse representation) *Assume that the number of sources is unknown and*

(i) for each index $i = 1, ..., n$ *there are at least two columns of* $\mathbf{S}$, $\mathbf{S}(:, j_1), \mathbf{S}(:, j_2)$ *which have nonzero elements only in position* i *(so each source is uniquely present at least twice), and*

(ii) $\mathbf{X}(:, k) \neq c\mathbf{X}(:, q)$ *for any* $c \in \mathbb{R}$, *any* $k = 1, ..., N$ *and any* $q = 1, ..., N, k \neq q$ *for which* $\mathbf{S}(:, k)$ *has more that one nonzero element.*

Then the number of sources and the matrix $\mathbf{A}$ *are identifiable uniquely up to permutation and scaling.*

Proof. We cluster in groups all nonzero normalized column vectors of $\mathbf{X}$ such that each group consists of vectors which differ only by sign. From conditions (i) and (ii) it follows that the number of the groups containing more that one element is precisely the number of sources n, and that each such group will represent a normalized column of $\mathbf{A}$ (up to sign). ∎

Below we include an algorithm for identification of the mixing matrix in the case of Theorem 7.

Algorithm for identification of the mixing matrix

1) Remove all zero columns of $\mathbf{X}$ (if any) and obtain a matrix $\mathbf{X}_1 \in \mathbb{R}^{m \times N_1}$.

2) Normalize the columns $\mathbf{x}_i, i = 1, \ldots, N_1$ of $\mathbf{X}_1 : \mathbf{y}_i = \mathbf{x}_i / \|\mathbf{x}_i\|$ and put $i = 1, j = 2, k = 1$.

3) if either $y_i = y_j$ or $y_i = -y_j$, then put $a_k = y_i$, increase i, k with 1, put $j = i + 1$ and if $i < N_1$, repeat 3) (otherwise stop). Otherwise: if $j < N_1$, increase j by 1 and repeat 3). If $j = N_1$, increase i by 1, put $j = i + 1$ and repeat 3). Stop when $i = N_1 + 1$.

In a similar way, as Theorem 1, we can prove the following its generalization.

Theorem 8 *Assume that*

(i) for each source $s_i := \mathbf{S}(i, .), i = 1, ..., n$ *there are* $k_i \geq 2$ *time instances when all of the source signals are zero except* s_i *(so each source is uniquely present* k_i *times), and*

(ii) the set $\{j \in \{1, ..., N\} : \mathbf{X}(., p) = c\mathbf{X}(., j)$ *for some* $c \in \mathbb{R}\}$, *contains less than* $\min_{1 \leq i \leq m} k_i$ *elements for any* $p \in \{1, ..., N\}$ *for which* $\mathbf{S}(., p)$ *has more than one nonzero element.*

Then the matrix $\mathbf{A}$ *is identifiable up to permutation and scaling.*

6 Identification of sources

Improved basis pursuit (BP) method

The famous basis pursuit method (**BP**) of S.S. Chen, D. Donoho and M. Sounders [8] is rather a principle than an algorithm for decomposing a signal into an "optimal" superposition of dictionary elements, where optimal means having the smallest l_1 norm of coefficients among all such decompositions. So, it consists of finding a minimum l_1-norm solution of a linear

underdetermined system. Such minimality of the l_1 norm ensures sparseness of the coefficients of the solution. Namely, if $\mathbf{A} \in \mathbb{R}^m \times n$, $m < n$, then the minimum l_1-norm solution of the system $\mathbf{As} = \mathbf{x}$ has at most m nonzero elements for almost all $\mathbf{x} \in \mathbb{R}^m$ - a well know fact (see [10] for instance). This problem can be reduced to a linear programming problem as follows:

$$\text{minimize} \qquad \sum_{i=1}^{n} u_i$$

subject to:

$$u_i \geq s_i, \quad u_i \geq -s_i, \quad \mathbf{As} = \mathbf{x}. \tag{22}$$

A disadvantage of this method is that it not always finds the sparsest solution. For a comprehensive discussion of this topic see [10].

Simple Example. Let $\mathbf{s}_* = (0, 2, -3, 0, 0, 0, 0, 0, 0, 0, 0, 0)^T$ be a solution of the system $\mathbf{As} = \mathbf{x}$, where $\mathbf{A} \in \mathbb{R}^{5\times 12}$ is randomly generated. In large number of cases, when we generate random matrix $\mathbf{A}$, the **BP** method doesn't find the sparsest solution $\mathbf{s}_*$.

$$\mathbf{A} = \begin{pmatrix} 0.2974 & 0.5527 & 0.3759 & 0.9200 & 0.1939 & 0.5488 & 0.6273 & 0.8376 & 0.7165 & 0.7006 & 0.1146 & 0.8230 \\ 0.0492 & 0.4001 & 0.0099 & 0.8447 & 0.9048 & 0.9316 & 0.6991 & 0.3716 & 0.5113 & 0.9827 & 0.6649 & 0.6739 \\ 0.6932 & 0.1988 & 0.4199 & 0.3678 & 0.5692 & 0.3352 & 0.3972 & 0.4253 & 0.7764 & 0.8066 & 0.3654 & 0.9994 \\ 0.6501 & 0.6252 & 0.7537 & 0.6208 & 0.6318 & 0.6555 & 0.4136 & 0.5947 & 0.4893 & 0.7036 & 0.1400 & 0.9616 \\ 0.9830 & 0.7334 & 0.7939 & 0.7313 & 0.2344 & 0.3919 & 0.6552 & 0.5657 & 0.1859 & 0.4850 & 0.5668 & 0.0589 \end{pmatrix}.$$

Solution by **BP**:

$$(-0.9977 \quad 0.0000 \quad -0.9640 \quad 0.8411 \quad 0.0000 \quad -0.0000 \quad 0.0000 \quad 0.0000 \quad -0.0000 \quad 0.0000 \quad 0.4042 \quad -0.2228)^T$$

Improved basis pursuit method: BP with zeros

We assume that the matrix $\mathbf{A}$ is known (or estimated correctly) and any $m \times m$ submatrix of it is nonsingular. Assume that the sparsest solution has no more than $m/2$ nonzero components. Recall that in this case (see Theorem 4 and **Observation**) the sparsest solution is unique and has no more than $m/2$ nonzero components, so this is criterion for finding it among all solutions. We propose the following modification of **BP** method, which we call **BP with zeros**:

solve the following minimization problem, where $\mathbf{e} = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$, n times (for $j = 1, ..., n$) as in each time change the $i-$th component of $\mathbf{e}$ with zero:

$$\text{minimize} \qquad \sum_{i=1, i\neq j}^{n} e_i u_i \quad , j = 1, ..., n \tag{23}$$

$$\text{subject to} \qquad u_i \geq s_i, \quad u_i \geq -s_i, \quad \mathbf{As} = \mathbf{x}. \tag{24}$$

If the sparsest solution is not found (which has no more than $m/2$ nonzero components), we replace pairs of the coefficients e_i in (23) with zeros and solve consecutively these problems until obtaining the sparsest solution. If it is not found again, proceed analogically with the triples of zeros and so on. For small m this procedure is effective up to level 2 (i.e. taking pars of zeros in the coefficients of (23). This of course is combinatorial problem which increase computational time but not so dramatic when the solution is very sparse.

The reason why our improvement works, is clear: suppose for instance that the sparsest solution s_* has 2 nonzero elements s_i and s_j. Putting $e_i = e_j = 0$ in some step of the algorithm, it will find this solution, since the minimum of the cost function is zero and it

is obtained exactly at s_*. In most cases the sparsest solution is obtained putting only one coefficient e_i equal to zero.

In case when the sources are nonnegative, a faster algorithm is proposed in Theorem 3.

7 Computer simulation examples

7.1 Complete case

In this example for the complete case ($m = n$) of instantaneous mixtures, we demonstrate the effectiveness of our algorithm for identification of the mixing matrix in the special case considered in Theorem 7. We mixed 3 images of landscapes (shown in Fig.1) with a 3-dimensional Hilbert matrix $\mathbf{A}$ and transformed them by a 2-D discrete Haar wavelet transform. As a result, since this transformation is linear, the high frequency components of the source signals become very sparse and they satisfy the conditions of Theorem 7. We use only one row (320 points) from the diagonal coefficients of the wavelet transformed mixture, which is enough to recover very precisely the ill conditioned mixing matrix $\mathbf{A}$. Fig. 3 shows the recovered mixtures.

Figure 1: Original images

Figure 2: Mixed (observed) images

7.2 Underdetermined case

First example. We generated artificially sources, shown in Fig.4 (left). They have level of sparseness 2 (at most two are nonzero at any time instant) and each source is uniquely active (achieves nonzero value while at the same time the rest of the signals are zero) at only 10 time instants. For instance, $s_4(k) = 0$ for $k = 211, ..., 220$, as unique nonzero source in this period is s_3, but with very small amplitude. Nevertheless, our algorithm is capable to

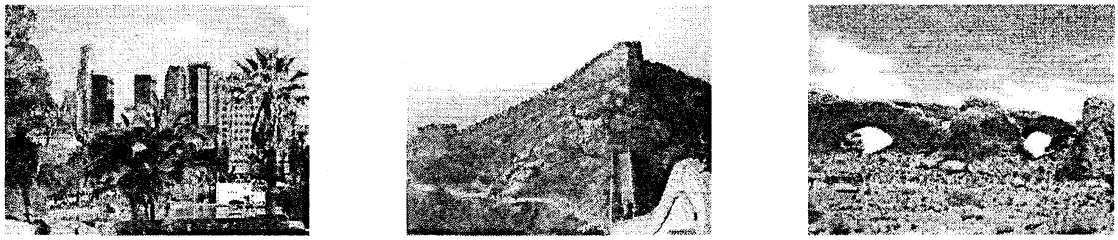

Figure 3: Estimated normalized images using the estimated matrix.

estimate precisely any randomly chosen matrix after the linear mixture of the sources. For instance, we generated randomly a matrix $\mathbf{A}46 \in \mathbb{R}^{4\times 6}$, and mixed the sources by it. The mixed sources are shown in Fig. 4 (right). We run our algorithm for estimating the mixing matrix (shown below as $\mathbf{A}$) and run the original **BP** method - the results of separation are shown in Fig. 5 (right). Our method **basis pursuit with zeros** gives excellent results (shown in Fig. 5 (left)), much better than those obtained by the standard **BP** method.

Initial matrix:

$$\mathbf{A}46 = \begin{pmatrix} 1.6777 & 0.3630 & 0.4840 & -1.8402 & 0.1751 & 0.4269 \\ 1.9969 & -0.5670 & -0.1938 & -1.6282 & 0.2294 & 1.4548 \\ 0.6970 & -1.0442 & -0.3781 & -1.1738 & -1.2409 & -0.5102 \\ -1.3664 & 0.6971 & -0.8864 & -0.4154 & 0.7000 & -0.0067 \end{pmatrix}$$

Normalized initial matrix: A46N

$$\mathbf{A}46N = \begin{pmatrix} 0.5545 & 0.2548 & 0.4418 & -0.6681 & 0.1204 & 0.2668 \\ 0.6600 & -0.3980 & -0.1768 & -0.5911 & 0.1578 & 0.9094 \\ 0.2303 & -0.7329 & -0.3451 & -0.4261 & -0.8536 & -0.3189 \\ -0.4516 & 0.4893 & -0.8090 & -0.1508 & 0.4815 & -0.0042 \end{pmatrix}$$

Estimated matrix (normalized)

$$\mathbf{A} = \begin{pmatrix} -0.2668 & -0.1204 & 0.6681 & -0.4418 & 0.2548 & 0.5545 \\ -0.9094 & -0.1578 & 0.5911 & 0.1768 & -0.3980 & 0.6600 \\ 0.3189 & 0.8536 & 0.4261 & 0.3451 & -0.7329 & 0.2303 \\ 0.0042 & -0.4815 & 0.1508 & 0.8090 & 0.4893 & -0.4516 \end{pmatrix}$$

Second example. The original sources are shown in Fig. 6 (lest). In this case the level of sparseness of the sources is 1, i.e. they are not overlapping or they are disjointly sparse. They were mixed with a randomly generated matrix, which after normalization is

$$\mathbf{A}24 = \begin{pmatrix} 0.5994 & 0.2458 & 0.0977 & 0.4335 \\ 0.8005 & 0.9693 & 0.9952 & 0.9011 \end{pmatrix}$$

The mixed signals are shown in Fig 7.

The estimated (normalized) matrix by our algorithm is

$$\mathbf{A} = \begin{pmatrix} 0.5994 & -0.2458 & -0.4335 & 0.0977 \\ 0.8005 & -0.9693 & -0.9011 & 0.9952 \end{pmatrix}$$

The estimated sources are shown in Fig. 6 (right). We should mention that in this example the results by the standard **BP** method are almost the same (which is due to the fact that the sources are not overlapping).

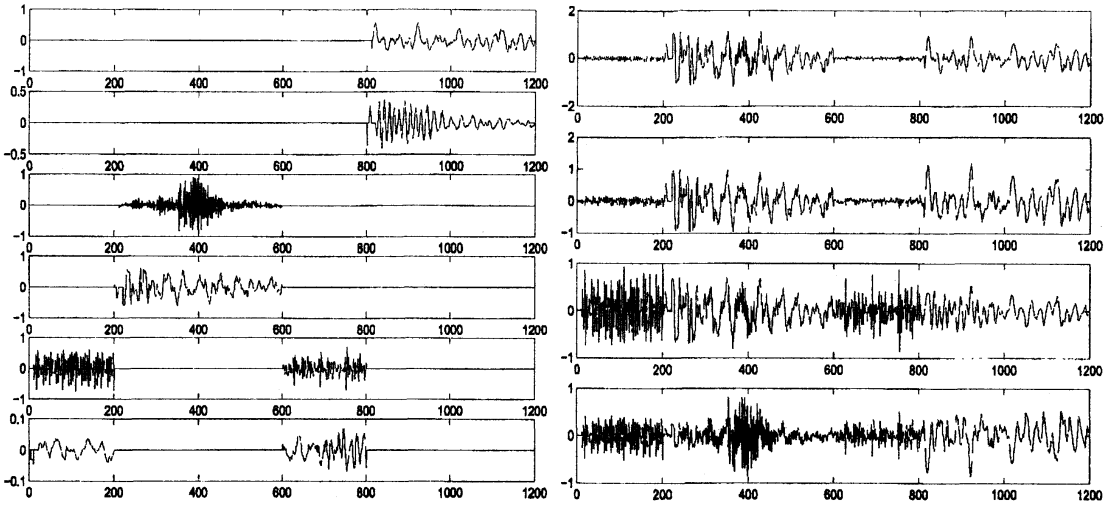

Figure 4: Original sources (left) and mixed sources (right)

8 Conclusion

We presented general theorems guaranteeing sparse solutions of nonlinear least square problems and of linear ones with respect to l_2-norm and least absolute value l_1-norm. We considered underdetermined sparse component analysis in the case when the sources are very sparse and presented two new algorithms: for matrix identification and for source recovery. When the sources are nonnegative, we propose a faster algorithm, based on a sparseness theorem for linear least square problems defined by l_2-norm. We presented several examples showing good performance of our algorithms.

References

[1] S. Amari and A. Cichocki "Adaptive blind signal processing - neural network approaches". *Proceedings IEEE* (invited paper), 86(10): 2016-2048, 1998.

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

[3] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization. Theory and Examples*, CMS Books in Mathematics, Springer, 2000.

[4] A.M. Bronstein, M.M. Bronstein, M. Zibulevsky, Y. Y. Zeevi, " Blind Separation of Reflections using Sparse ICA", in Proc. Int. Conf. ICA2003, Nara, Japan, pp.227-232.

[5] C. Byrne, Mathematics in Signal Processing, http://faculty.uml.edu/cbyrne/master.pdf

[6] F. Clarke, Optimization and Nonsmooth Analysis, Wiley and sons, 1984.

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

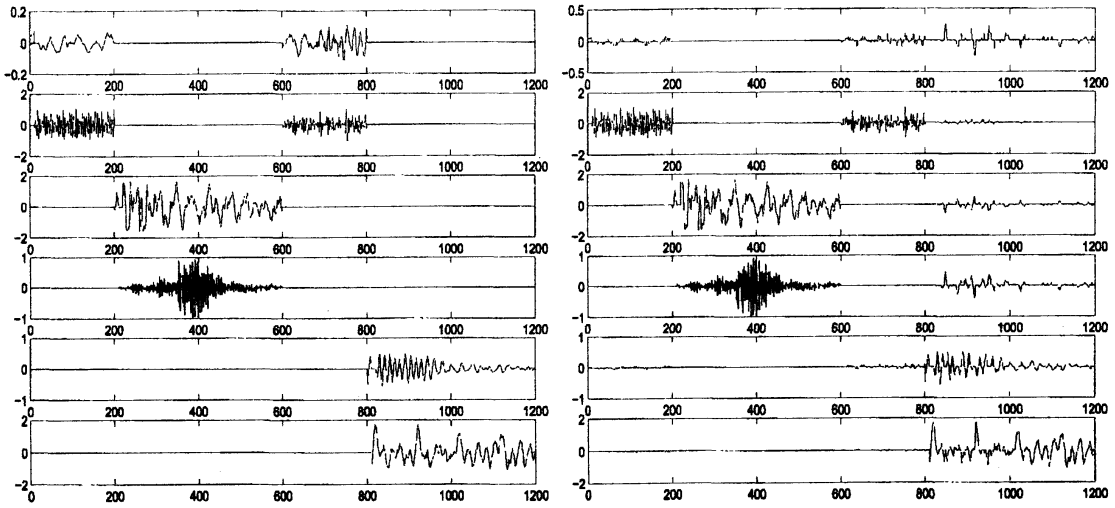

Figure 5: Left: Estimated sources by the new method: **basis pursuit with zeros** using the estimated matrix. Right: Estimated sources by the basis pursuit using the estimated matrix

[8] S. Chen, D. Donoho, M. Saunders, " Atomic decomposition by basis pursuit", SIAM J. Sci. Comput. 20 (1998), no. 1, 33–61.

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

[10] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization", Proc. Nat. Acad. Sci., March 4, 2003, vol.100, no.5, 2197-2202.

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

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

[13] M. Zibulevsky, and B. A. Pearlmutter, Blind source separation by sparse decomposition, Neural Comp. 13(4), 2001.

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

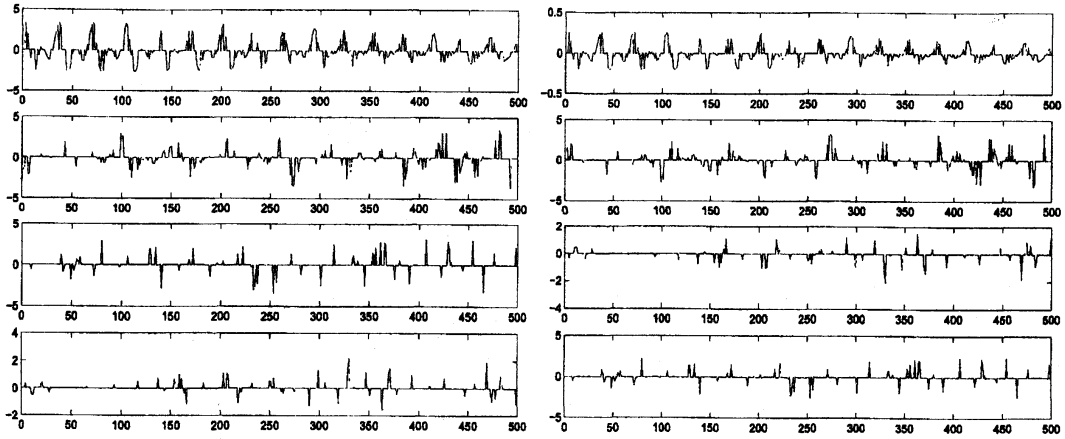

Figure 6: Left: Original sources (second example). Right: Estimated sources

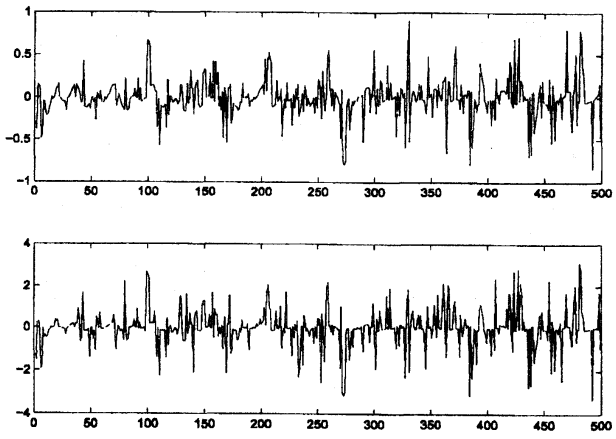

Figure 7: Mixed signals (second example)