BLIND SOURCE SEPARATION AND SPARSE COMPONENT ANALYSIS OF OVERCOMPLETE MIXTURES

Pando Georgiev[†], Fabian Theis[†‡] and Andrzej Cichocki[†]

[†] Brain Science Institute, RIKEN
Lab. for Advanced Brain Signal Processing
2-1, Hirosawa, Wako-shi, Saitama, 351-0198, Japan
[‡] Institute of Biophysics, University of Regensburg
D-93040 Regensburg, Germany
georgiev,cia@bsp.brain.riken.go.jp, fabian@theis.name

ABSTRACT

We formulate conditions (k-SCA-conditions) under which we can represent a given $(m \times N)$-matrix $\mathbf{X}$ (data set) uniquely (up to scaling and permutation) as a multiplication of $m \times n$ and $n \times N$ matrices $\mathbf{A}$ and $\mathbf{S}$ (often called mixing matrix or *dictionary* and source matrix, respectively), such that $\mathbf{S}$ is *sparse of level* $n-m+k$ in sense that each column of $\mathbf{S}$ has at least $n - m + k$ zero elements. We call this the *k-Sparse Component Analysis* problem (k-SCA). Conditions on a matrix $\mathbf{S}$ are presented such that the k-SCA-conditions are satisfied for the matrix $\mathbf{X} = \mathbf{AS}$, where $\mathbf{A}$ is an arbitrary matrix from some class. This is the *Blind Source Separation* problem and the above conditions are called *identifiability conditions*.

We present new algorithms: for matrix identification (under k-SCA-conditions), and for source recovery (under identifiability conditions). The methods are illustrated with examples, showing good separation of the high-frequency part of mixtures of images after appropriate sparsification.

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 $(n \times N)$-matrix) in different ways. A simple idea is a linear representation:

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

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

1) the rows of $\mathbf{S}$ are as statistically independent as possible — this is *Independent Component Analysis* (ICA) problem;

2) $\mathbf{S}$ contains as many zeros as possible — this is the 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*.

Such linear representations have 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, micro-array data mining, enhancement of images in nuclear medicine etc. (see [7], [5]).

There is a large amount of papers devoted to ICA problems (see for instance [3], [6] and references therein) but mostly for the complete case ($m = n$). We refer to [9], [1], [10], [2], [8] and reference therein for some recent papers on SCA and overcomplete ICA ($m < n$).

A slightly different 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 overcomplete case ($m < n$ i.e. more sources than sensors), where the additional information compensating the lack of sensors is the *sparseness* of the sources. The task of the SCA problem is to represent the given data $\mathbf{X}$ as in equation (1) such that the matrix $\mathbf{S}$ (sources) are sparse of level $n-m+k$ (i.e. each column of $\mathbf{S}$ has at least $n - m + k$ zeros). We present conditions on the data matrix $\mathbf{X}$ (*k-SCA- or simply SCA-conditions on the data*), under which the representation in equation (1) is unique up to permutation and scaling of the sources.

The task of BSS problem is to estimate the unknown sources $\mathbf{S}$ (and the mixing matrix $\mathbf{A}$) using the available data matrix $\mathbf{X}$ only. We describe conditions (*identifiability conditions on the sources*) under which this is possible uniquely up to permutation and scaling of the sources, which is the usual condition in the complete BSS problems.

In the sequel, we present new algorithms for solving the BSS problem: matrix identification algorithms and source recovery algorithm, which recovers the sources with level of sparseness at least $n - m + 1$. When the sources are sufficiently sparse (see the conditions of Theorem 2) the matrix identification algorithm is even simpler. We used this simpler form for separation of mixtures of images. After sparsification transformation the algorithm works perfectly in the complete case. In the overcomplete case, the recovery of the matrix is perfect, but the recovery of the sources needs sufficient sparsification of the approximation component of the discrete wavelet transformation. The source recovery algorithm works good, if sufficient sparsification is achieved.

2. BLIND SOURCE SEPARATION

In this section we develop a method for completely solving the BSS problem in the case when 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, and

A2) the sources are sparse of level $n - m + 1$, i.e. each column of $\mathbf{S}$ has at least $n - m + 1$ zero elements.

2.1. Matrix identification

In this section we describe conditions under which we can identify the mixing matrix in the sparse BSS problem. Some proofs are omitted due to lack of space.

Theorem 1 (Identifiability conditions - general case) *Assume that for every $j \in \{1, ..., n\}$, there exists m group of samples $\{t_{q_1,p}\}_{p=1}^m, ..., \{t_{q_m,p}\}_{p=1}^m$ such that $s_j(t_{q_k,1}) \neq 0$, $\forall k = 1, ..., m$, the vectors $\{\mathbf{s}(t_{q_k,p})\}_{p=1}^m$ have zeros at the same indexes, and any $m - 1$ of them are linearly independent. Then the mixing matrix $\mathbf{A}$ is identifiable from any mixture $\mathbf{X} = \mathbf{AS}$, provided $\mathbf{A}$ and $\mathbf{S}$ satisfy conditions A1) and A2) respectively.*

Algorithm for identification of the mixing matrix 1

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.

2) Cluster the normal vectors to these hyperplanes in n groups G_j, $j = 1, ..., n$ such that the normal vectors to the hyperplanes in each group G_j lie in a new hyperplane $\hat{H}_j$, and these hyperplanes $\hat{H}_j$ are different.

3) Calculate the normal vectors $\hat{\mathbf{a}}_j$ to each hyperplane $\hat{H}_j$, $j = 1, ..., n$. Note that the one-dimensional subspace spanned by $\hat{\mathbf{a}}_j$ is the intersection of all hyperplanes in G_j. The matrix $\hat{\mathbf{A}}$ with columns $\{\hat{\mathbf{a}}_j\}_{j=1}^n$ is an estimation of the mixing matrix (up to permutation and scaling of the columns).

Theorem 2 (Identifiability conditions - sparse instances) *Assume that the number of sources is unknown and*

1) for each source $s_i := \mathbf{S}(i, .)$ there are at least two time instances when all the signals are zero except s_i (so each source is uniquely present at least twice), and

2) $\mathbf{A}(\mathbf{S}(., k) - M\mathbf{S}(., q)) \neq 0$ for any $M \in \mathbb{R}$, 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 is recoverable and the matrix $\mathbf{A}$ is identifiable up to permutation and scaling.

Here we used the notation $\mathbf{S}(i, .)$ and $\mathbf{S}(., j)$ for the i-th row respectively j-th column of $\mathbf{S}$.

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 1) and 2) 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 2.

Algorithm for identification of the mixing matrix 2

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

2) Normalize the columns $\mathbf{x}_i, i = 1, ..., N_1$ 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 $i = N_1 + 1$.

2.2. Identification of sources

Theorem 3 (Source recovery) *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 when $k = 1$), 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, because if $\mathbf{As} = \mathbf{A\bar{s}}$ then due to A1) and sparseness, $\mathbf{s}$ and $\mathbf{\bar{s}}$ lie in different subspaces, so $\mathbf{x} \in \mathcal{H}_0$. ∎

From Theorem 3 it follows that the sources are identifiable generically (or with probability one), i.e. up to a set with measure zero, if they have level of sparseness grater than or equal to $n - m + 1$, and the mixing matrix is known. Below is the algorithm, based on the observation in Theorem 3.

Source recovery algorithm:

1. Identify the the set of k-codimensional subvectorspaces $\mathcal{H}$ 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:

2.1. Identify the space $H \in \mathcal{H}$ 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 $\tilde{\mathbf{x}}_i$.

2.2. If H is spanned by 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 $\mathcal{H}$ (see Theorem 3).

2.3. Construct solution $\mathbf{s}_i = \mathbf{S}(., i)$: it contains $\lambda_{i,j}$ in the place i_j for $j = 1, ..., m - k$, the other components are zero.

3. SPARSE COMPONENT ANALYSIS

In this section we develop a method for the complete solution of the SCA problem in the case of A1) and A2) from the previous section.

Theorem 4 *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 hyperplanes, each column lies in only one such hyperplane, each hyperplane contains at least m columns of $\mathbf{X}$ such that any $m - 1$ of them are linearly independent.

(ii) for each $i \in \{1, ..., n\}$ there exist at least m different hyperplanes $\{H_{i,j}\}_{j=1}^m$ in $\mathcal{H}$ such that their intersection $L_i = \cap_{j=1}^m H_{i,j}$ 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 the rows of $\mathbf{S}$) in the form (1), and $\mathbf{A}$, $\mathbf{S}$ satisfy conditions A1) and A2) respectively.

4. COMPUTER SIMULATION EXAMPLES

4.1. 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 2. We mixed 3 images of landscapes (shown in Fig.1) with a 3-dimensional Hilbert matrix $\mathbf{H}$ and transformed them by a discrete wavelet transform using the Haar wavelet. As a result, high frequency components become very sparse and they satisfy the conditions of theorem 2. We use only one row (320 points) of the transformed mixture, which is enough to recover very precisely the mixing matrix ($\det(\mathbf{H}) = 0.00046$). Fig. 3 shows the recovered mixtures.

Fig. 1. Original sources

Fig. 2. Mixed (observed) signals

Fig. 3. Estimated normalized images using the estimated matrix. The signal-to-noise ratios with the sources from figure 1 are 232, 239 and 228 respectively.

4.2. Overcomplete case

Here simulations for the overcomplete case are presented both using artificial and real data.

We consider the mixture of four artificially created sources — sparsified sine signals with different wavelengths and at least 2 zeros in each column — with the semiorthogonal mixing matrix

$$\mathbf{A} = \left(\begin{array}{cccc} -0.52 & 0.20 & 0.56 & -0.88 \\ -0.76 & -0.45 & -0.80 & 0.12 \\ -0.38 & 0.87 & -0.20 & 0.45 \end{array} \right).$$

Figure 4 gives a normalized scatterplot of the mixtures — the data clearly lie in 6 hyperplanes, which confirms that the sources are
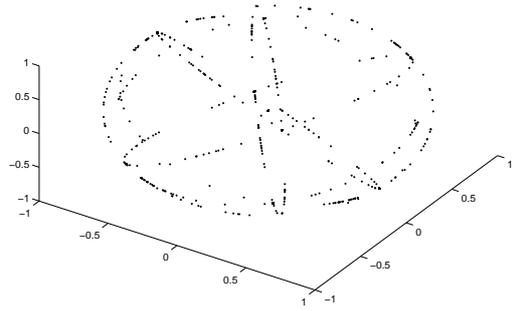

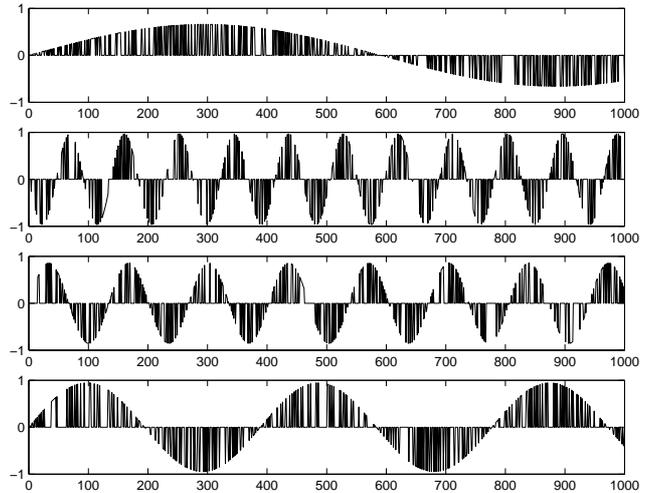**Fig. 4**. Normalized scatter plot (density) of the mixtures.

Fig. 5. Recovered source signals. The signal-to-noise ratio between the original sources and the recoveries is very high with 292, 282, 305 and 291 dB after permutation and normalization.

at least sparse of level 2. According to Theorem 4 SCA can be used. Applying the overcomplete matrix recovery algorithm 1 to the mixtures gives the recovered mixing matrix

$$\mathbf{B} = \left(\begin{array}{cccc} 0.56 & -0.20 & -0.88 & -0.53 \\ -0.80 & 0.45 & 0.12 & -0.76 \\ -0.20 & -0.87 & 0.45 & -0.38 \end{array} \right).$$

The generalized crosstalking error [8] is very low with $E(\mathbf{A}, \mathbf{B}) = 6.7 \cdot 10^{-16}$, which confirms the matrix identification stated in theorem 1. Applying sparse source recovery algorithm to the mixtures with this matrix gives perfectly recovered signals, see figure 5. Again this is a confirmation of the assertion of source identification in Theorem 3.

Now an application to real data sets is presented. We consider the mixtures of natural images. Namely, four 320×240 grey-level natural scenes are linearly mixed to three images using the semiorthogonal mixing matrix $\mathbf{A}$ from above. The mixed images were sparsified using iterative two-dimensional discrete wavelet decomposition with the Haar wavelet. Application of the overcomplete matrix recovery algorithm 2 to only five rows of the sparse

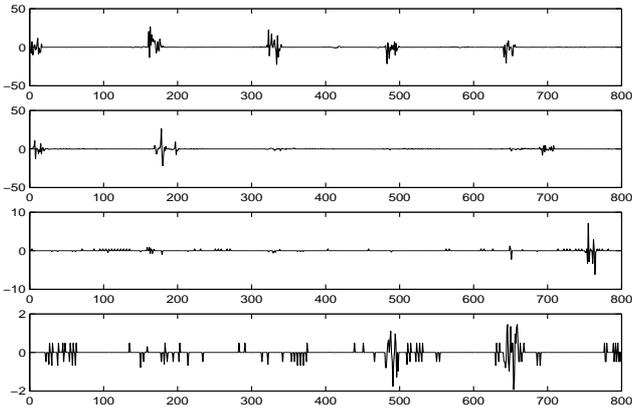**Fig. 6**. Three mixed images are shown, containing four independent source images.

Fig. 7. Recovered diagonal wavelet coefficients using sparseness. The signal-to-noise ratios of the original source coefficients with the recovered signals are 11.4, 8.0, 23.7 and 19.7 respectively. During the signal recovery, 24 out of 800 points were found which did not lie on any hyperplane spanned by columns of **A**. This is due to the fact that the sources are not fully sparse of level 2, which accounts for the (small) recovery error.

first diagonal wavelet coefficients gives the recovered mixing matrix

$$\mathbf{B} = \begin{pmatrix} 0.88 & 0.56 & 0.20 & -0.53 \\ -0.12 & -0.80 & -0.45 & -0.76 \\ -0.45 & -0.20 & 0.87 & -0.38 \end{pmatrix}.$$

Again, the generalized cross talking error between those two matrices is very low $E(\mathbf{A}, \mathbf{B}) = 1.1 \cdot 10^{-13}$, which confirms the good matrix recovery. Applying the overcomplete source recovery algorithm to the very sparse wavelet coefficients gives very good performance, as is demonstrated in figure 7.

In order to then find the source images, the overcomplete source recovery algorithm is applied to each of the wavelet coefficients separately. Only the non-sparse approximation images could not be recovered in this way; for them, best results were achieved by applying the pseudo-inverse of **B**. This corresponds to 2-norm minimization under the constraint $\mathbf{x} = \mathbf{Bs}$ i.e. to a Gaussian prior [8], which represents a better approximation to the unknown low-frequency density of the approximation images than the sparse models used for the wavelet coefficients. Figure 8 gives a plot of the recovered images. It can be seen that the high-frequency part is well recovered although low-frequency obstructions due to indeterminacies of 2-norm minimization did not allow for recoveries with higher SNRs.
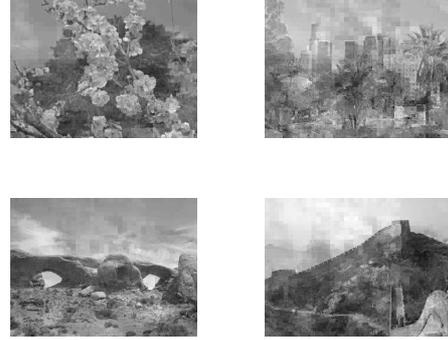

Fig. 8. Recovered images using overcomplete SCA. Sign and scaling was chosen appropriately for visualization. The signal-to-noise ratios of the sources with the recovered images are 6.8, 4.1, 11.0 and 11.8 respectively.

5. CONCLUSION

We developed rigorously a theory for Blind Signal Separation of overcomplete linear mixtures of sparse signals, presenting sufficient conditions for that and developing new algorithms for identification of the mixing matrix and for source recovery. Our main assumptions for sparseness (i.e. each column of the source matrix has less non-zero elements than the mixing dimension) allows us to obtain uniqueness for the source recovery by our algorithm. The presented experiment for separation of artificially created signals with sufficient level of sparseness give excellent results. The experiments for the separation of mixtures of images perform very well in the complete case and sufficiently well in the overcomplete case (due to presence of low-frequency components).

6. REFERENCES

[1] 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.

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

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

[4] Chen, Scott Shaobing; Donoho, David L.; Saunders, Michael A., " Atomic decomposition by basis pursuit", SIAM J. Sci. Comput. 20 (1998), no. 1, 33–61.

[5] 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

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

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

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

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

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