Robust overcomplete matrix recovery for sparse sources using a generalized Hough transform

Fabian J. Theis, Pando Georgiev and Andrzej Cichocki
Brain Science Institute, RIKEN, Lab. for Adv. Brain Signal Processing
2-1, Hirosawa, Wako-shi, Saitama, 351-0198, Japan

Abstract. We propose an algorithm for recovering the matrix $\mathbf{A}$ in $\mathbf{X} = \mathbf{AS}$ where $\mathbf{X}$ is a random vector of lower dimension than $\mathbf{S}$. $\mathbf{S}$ is assumed to be sparse in the sense that $\mathbf{S}$ has less nonzero elements than the dimension of $\mathbf{X}$ at any given time instant. In contrast to previous approaches, the computational time of the presented algorithm is linear in the sample number and independent of source dimension, and the algorithm is robust against noise. Experiments confirm these theoretical results.

1 Introduction

A new algorithm for decomposing a data set $\mathbf{X}$ in $\mathbf{X} = \mathbf{AS}$ is proposed where the sources $\mathbf{S}$ are assume to be sparse (Sparse Component Analysis, SCA) [2]. We use the Hough transform to detect hyperplanes in $\mathbf{X}$, which leads to matrix and source identification. The Hough transform has been used in a somewhat ad hoc way in the field of independent component analysis by [4] for identifying two dimensional sources in the mixture plot, and in [5], again in the overcomplete case but with mixture dimension two as a direct application to the mixtures (without additional restrictions), which can be shown to be problematic theoretically. In the literature, there are quite a few more overcomplete BSS and basis estimation algorithms (for example [6, 7] and references therein), which also use sparse priors (sources), but in addition also assume independence of the sources. This allows for less strict sparsity conditions, but cannot guarantee source identifiability as in our case.

2 Overcomplete SCA

Definition 2.1. *A vector* $\mathbf{v} \in \mathbb{R}^n$ *is said to be* k-sparse *if* $\mathbf{v}$ *has at least* k *zero entries. A random vector* $\mathbf{S} : \Omega \to \mathbb{R}^n$ *is said to be* k-sparse *if* $\mathbf{S}(\omega)$ *is* k*-sparse for every event* $\omega \in \Omega$.

The goal of *Sparse Component Analysis* of level k (k-SCA) is to decompose a given m-dimensional random vector $\mathbf{X}$ into $\mathbf{X} = \mathbf{AS}$ with a real $m \times n$-matrix

$\mathbf{A}$ and an n-dimensional k-sparse random vector $\mathbf{S}$. $\mathbf{S}$ is called the *source vector*, $\mathbf{X}$ the *mixtures* and $\mathbf{A}$ the *mixing matrix*. We speak of *complete*, *overcomplete* or *undercomplete* k-SCA if $m = n$, $m < n$ or $m > n$ respectively.

In the following without loss of generality we will assume $m \leq n$: the undercomplete case can be reduced to the complete case by projection of $\mathbf{X}$.

Theorem 2.2 (Matrix identifiability). *Consider the k-SCA problem for $k := n - m + 1$ and assume that every $m \times m$-submatrix of $\mathbf{A}$ is invertible. If $\mathbf{S}$ is sufficiently rich then $\mathbf{A}$ is uniquely determined by $\mathbf{X}$ except for left-multiplication with permutation and scaling matrices.*

This means that we can recover the mixing matrix from the mixtures. Sufficiently rich means that all combinations of non-zeros must occur in sufficiently many samples [2].

Theorem 2.3 (Source identifiability). *Let $\mathcal{H}$ be the set of all $\mathbf{x} \in \mathbb{R}^m$ such that the linear system $\mathbf{A}\mathbf{s} = \mathbf{x}$ has a $(n-m+1)$-sparse solution i.e. one with at least $n - m + 1$ zero components. If $\mathbf{A}$ fulfills the condition from theorem 2.2, 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.*

For proofs of these theorems together with the corresponding algorithms, we refer to [2].

3 Hough transform

The Hough transform is a well-known method for finding shapes in images, widely used in the field of image processing [1, 3]. It is robust to noise and occlusions and is used for extracting lines, circles or other shapes from images. In addition to these nonlinear extensions, it can also be made more robust to noise using anti-aliasing techniques.

The mixing matrix $\mathbf{A}$ in the case of $(n-m+1)$-sparse SCA can be recovered by finding all 1-codimensional subvectorspaces in the mixture data set. The algorithm presented here uses a generalized version of the Hough transform in order to determine hyperplanes through 0 as follows:

Vectors $\mathbf{x} \in \mathbb{R}^m$ lying on such a hyperplane H can be described by the equation $\mathbf{n}^\top \mathbf{x} = 0$, where $\mathbf{n}$ is a nonzero vector orthogonal to H. After normalization (i.e. $\mathbf{n}$ lies on the $(n-1)$-sphere) $\mathbf{n}$ is uniquely determined by H if we additionally require $\mathbf{n}$ to lie on one hemisphere. In terms of spherical coordinates

$$\mathbf{n} = \begin{pmatrix} \cos\varphi \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{m-2} \\ \sin\varphi \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{m-2} \\ \cos\theta_1 \sin\theta_2 \ldots \sin\theta_{m-2} \\ \vdots \quad \ddots \quad \vdots \\ \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{m-2} \end{pmatrix} \tag{1}$$

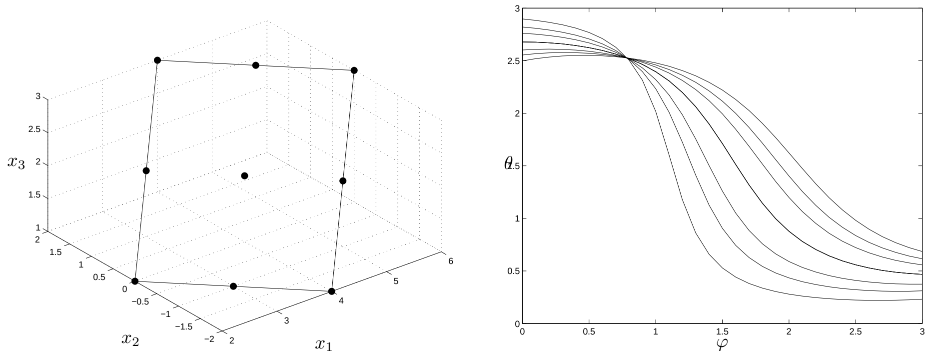


Figure 1: Illustration of the 'hyperplane detecting' Hough transform in three dimensions: A point (x_1, x_2, x_3) in the data space (left) is mapped onto the curve $\{(\varphi, \theta) | \theta = \arctan(x_1 \cos\varphi + x_2 \sin\varphi) + \frac{\pi}{2}\}$ in the parameter space $[0, \pi)^2$ (right). The Hough curves of points belonging to one plane in data space intersect in precisely one point (φ, θ) in the data space — and the data points lie on the plane given by the normal vector $(\cos\varphi \sin\theta, \sin\varphi \sin\theta, \cos\theta)$.

with $(\varphi, \theta_1, \ldots, \theta_{m-2}) \in [0, 2\pi) \times [0, \pi)^{m-2}$ uniqueness of $\mathbf{n}$ can be achieved by requiring $\varphi \in [0, \pi)$.

Vectors $\mathbf{x}$ in the hyperplane given by $\mathbf{n}$ are orthogonal to it, so in spherical coordinates this gives $\cot\theta_{m-2} = -\sum_{i=1}^{m-1} \nu_i(\varphi, \theta_1, \ldots, \theta_{m-3}) \frac{x_i}{x_m}$ for $\mathbf{x} \in \mathbb{R}^m$ with $x_m \neq 0$ and

$$\nu_i := \begin{cases} \cos\varphi \prod_{j=1}^{m-3} \sin\theta_j & i = 1 \\ \sin\varphi \prod_{j=1}^{m-3} \sin\theta_j & i = 2 \\ \prod_{j=1}^{i-2} \cos\theta_j \prod_{j=i-1}^{m-3} \sin\theta_j & i > 2 \end{cases}$$

With $\cot(\theta + \frac{\pi}{2}) = -\tan(\theta)$ we finally get $\theta_{m-2} = \arctan\left(\sum_{i=1}^{m-1} \nu_i \frac{x_i}{x_m}\right) + \frac{\pi}{2}$. Note that continuity is achieved if we set $\theta_{m-2} := 0$ for $x_m = 0$. We can then define the *generalized ('hyperplane detecting') Hough transform* as

$$\begin{aligned} \eta : \mathbb{R}^m &\longrightarrow \mathcal{P}([0, \pi)^{m-1}) \\ \mathbf{x} &\longmapsto \left\{ (\varphi, \theta_1, \ldots, \theta_{m-2}) \in [0, \pi)^{m-1} | \theta_{m-2} = \arctan(\sum_{i=1}^{m-1} \nu_i \frac{x_i}{x_m}) + \frac{\pi}{2} \right\} \end{aligned} \tag{2}$$

where $\mathcal{P}(U)$ denotes the set of all subsets of U.

Points lying on the same hyperplane are mapped to surfaces that intersect in precisely one point in $[0, \pi)^{m-1}$. This is demonstrated in the case $m = 3$ in figure 1. The hyperplane structures of a data set $\bar{\mathbf{X}} = \{\mathbf{x}(1), \ldots, \mathbf{x}(T)\}$ can be analyzed by finding clusters in $\eta(\bar{\mathbf{X}})$.

4 Hough SCA Algorithm

The SCA matrix detection algorithm [2] consists of two steps. In the first step, $d := \binom{n}{m-1}$ hyperplanes given by their normalvectors $\mathbf{n}^{(1)}, \ldots, \mathbf{n}^{(d)}$ are constructed such that the mixture data lies as good as possible in the union of these hyperplanes. In the second step, mixture matrix columns $\mathbf{a}_i$ are identified as generators of the n lines lying at the intersections of $\binom{n-1}{m-2}$ hyperplanes. We replace the first step by the following *Hough SCA algorithm*.

The idea is to first plot the Hough curves $\eta(\mathbf{x}(t))$ corresponding to the samples $\mathbf{x}(t)$ of $\mathbf{X}$ in a discretized parameter space (often called *accumulator*). Plotting in the accumulator is sometimes denoted as *voting* for each bin, similar to histogram generation. Maxima analysis of the accumulator gives the hyperplanes in the parameter space.

The Hough SCA algorithm is described in algorithm 1. We see that only the hyperplane identification step is different from the original hyperplane detection algorithm [2], the matrix identification is the same.

Algorithm 1: Hough SCA algorithm

Data : Samples $\mathbf{x}(1), \ldots, \mathbf{x}(T)$ of the random vector $\mathbf{X}$

Result: Estimated mixing matrix $\hat{\mathbf{A}}$

Hyperplane identification.

Fix the number β of bins (can be separate for each angle).

Initialize the 'accumulator' array $\alpha \in \mathbb{R}^{\beta^{m-1}}$ with zeros.

for $t \leftarrow 1, \ldots, T$ **do**

    for $\varphi, \theta_1, \ldots, \theta_{m-3} \leftarrow 0, \frac{\pi}{\beta}, \ldots, \frac{(\beta-1)\pi}{\beta}$ **do**

        $\theta_{m-2} \leftarrow \arctan(\sum_{i=1}^{m-1} \nu_i(\varphi, \ldots, \theta_{m-3}) \frac{x_i(t)}{x_m(t)}) + \frac{\pi}{2}$

        Increase ('vote for') the accumulator value of α in bin corresponding to $(\varphi, \theta_1, \ldots, \theta_{m-2})$ by one.

    end

end

The $d := \binom{n}{m-1}$ largest local maxima of α correspond to the d hyperplanes present in the data set.

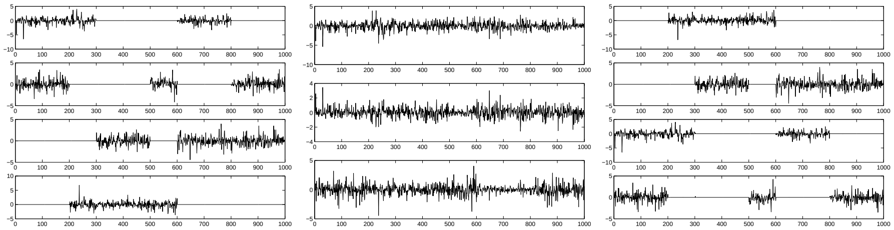
Backtransformation as in equation 1 gives the corresponding normal vectors $\mathbf{n}^{(1)}, \ldots, \mathbf{n}^{(d)}$ to those hyperplanes.

Matrix identification.

Clustering of hyperplanes generated by $(m-1)$-tuples in $\{\mathbf{n}^{(1)}, \ldots, \mathbf{n}^{(d)}\}$ gives n separate hyperplanes.

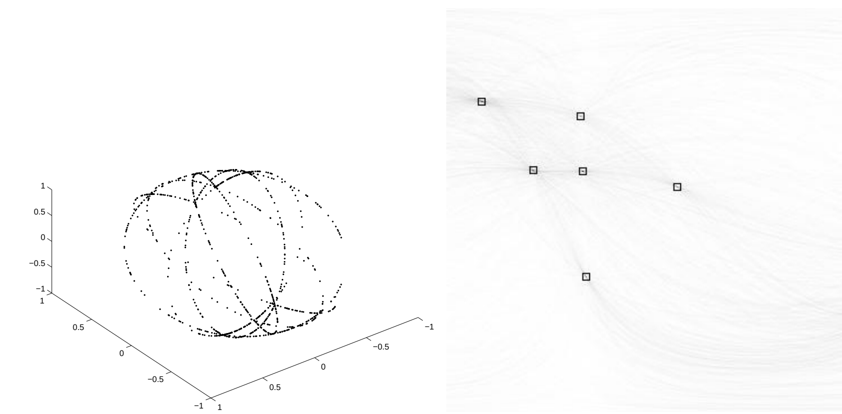
Their normal vectors are the n columns of the estimated mixing matrix $\hat{\mathbf{A}}$.

β is also called the *grid resolution*. Similar to histogram-based density



(a) sources (b) mixtures (c) recovered sources

Figure 2: Example signals.



(a) (b)

Figure 3: A sphered (normalized) scatter plot is shown in order to depict the density 3(a) together with the mixture space after Hough transform ('accumulator') with 360 discretization steps, where the found maxima are marked with little squares 3(b).

estimation the choice of β can seriously effect the algorithm performance — if chosen too small possible maxima cannot be resolved, and if chosen too large, the sensitivity of the algorithm increases and the computational burden in terms of speed and memory grows considerably.

5 Experiments

In the first experiment we consider the case of 4 sources and 3 mixtures. The 4-dimensional sources as plotted in figure 2(a) are 2-sparse and consist of 1000 samples. Obviously all combinations (i, j), $i < j$ of active sources are present in the data set — this condition (sufficiently rich sources) is needed in the matrix recovery step. The sources were mixed using a mixing matrix with randomly (uniform in $[-1, 1]$) chosen coefficients to give mixtures as shown in figure 2(b). The mixture density clearly lies in 6 disjoint hyperplanes, spanned by pairs $(\mathbf{a}_i, \mathbf{a}_j)$, $i < j$ of mixture matrix columns.

In order to detect the planes in the data space, we apply the generalized Hough transform as explained in section 3. Figure 3(b) shows the Hough image with $\beta = 360$. Each sample results in a curve, and clearly 6 intersection points are visible, which correspond to the 6 hyperplanes in question. Maxima analysis retrieves these points (in Hough space) as shown in the same figure. After transforming these points back into $\mathbb{R}^3$ with the inverse Hough transform, we get 6 normalized vectors corresponding to the 6 planes. We then recover the 4 matrix columns combinatorially from the plane norm vectors, see algorithm 1 to get the recovered mixing matrix $\hat{\mathbf{A}}$. The deviation from the original mixing matrix in the overcomplete case can be measured by the *generalized cross talking error* [6], which here is very low with $E(\mathbf{A}, \hat{\mathbf{A}}) = 0.040$. Then, the sources are recovered using the source recovery algorithm from [2] with the approximated mixing matrix $\hat{\mathbf{A}}$. The (normalized) SNRs of the recovered sources with the original ones, figure 2, are high with 36, 38, 36 and 37 dB respectively.

In a second experiment, we add 1% Gaussian white noise to the above mixture and apply the generalized Hough transform to the mixture data, however because of the noise we choose a more coarse discretization (180 bins). Curves in Hough space corresponding to a single plane do not any more intersect in precisely one point due to the noise; a low-resolution Hough space however fuses these intersections in one point, so that our simple maxima detection still achieves good results. We recover the mixing matrix $\hat{\mathbf{A}}'$ with generalized cross talking error is low with $E(\mathbf{A}, \hat{\mathbf{A}}') = 0.12$. We get good SNRs of the recovered sources (average 20 dB), which is quite satisfactory considering the additive noise and the overcomplete mixture situation.

In a final example, we use again the same mixtures, but replace 80% of the samples by outliers (drawn from a 4-dimensional normal distribution). We find that the Hough SCA algorithm is very robust against outliers. This is confirmed by the recovered mixing matrix $\hat{\mathbf{A}}''$ with excellent generalized cross talking error of $E(\mathbf{A}, \hat{\mathbf{A}}'') = 0.040$. Furthermore, the sparse recovery method can detect outliers by measuring distance from the hyperplanes, so outlier rejection is possible.

References

[1] R.O. Duda and P.E. Hart. Use of hough transform to detect lines and curves in pictures. *Comm. ACM*, 15:204–208, 1972.

[2] P. Georgiev, F.J. Theis, and A. Cichocki. Blind source separation and sparse component analysis of overcomplete mixtures. In *Proc. of ICASSP 2004*, Montreal, Canada, 2004.

[3] P.V.C. Hough. Machine analysis of bubble chamber pictures. In *International Conference on High Energy Accelerators and Instrumentation*, CERN, 1959.

[4] J. Lin, D. Grier, and J. Cowan. Feature extraction approach to blind source separation. In *IEEE Workshop on Neural Networks for Signal Processing*, pages 398–405, 1997.

[5] H. Shindo and Y. Hirai. An approach to overcomplete-blind source separation using geometric structure. In *2001 Annual Conference of Japanese Neural Network Society*, pages 95–96, Nara-Machi Center, 2001. in Japanese.

[6] F.J. Theis, E.W. Lang, and C.G. Puntonet. A geometric algorithm for overcomplete linear ICA. *Neurocomputing*, 56:381–398, 2004.

[7] M. Zibulevsky and B.A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computation*, 13(4):863–882, 2001.