Research Article

Robust Sparse Component Analysis Based on a Generalized Hough Transform

Fabian J. Theis,[1] Pando Georgiev,[2] and Andrzej Cichocki[3, 4]

[1] *Institute of Biophysics, University of Regensburg, 93040 Regensburg, Germany*

[2] *ECECS Department and Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221, USA*

[3] *BSI RIKEN, Laboratory for Advanced Brain Signal Processing, 2-1, Hirosawa, Wako, Saitama 351-0198, Japan*

[4] *Faculty of Electrical Engineering, Warsaw University of Technology, Pl. Politechniki 1, 00-661 Warsaw, Poland*

An algorithm called Hough SCA is presented for recovering the matrix $\mathbf{A}$ in $\mathbf{x}(t) = \mathbf{As}(t)$, where $\mathbf{x}(t)$ is a multivariate observed signal, possibly is of lower dimension than the unknown sources $\mathbf{s}(t)$. They are assumed to be sparse in the sense that at every time instant t, $\mathbf{s}(t)$ has fewer nonzero elements than the dimension of $\mathbf{x}(t)$. The presented algorithm performs a global search for hyperplane clusters within the mixture space by gathering possible hyperplane parameters within a Hough accumulator tensor. This renders the algorithm immune to the many local minima typically exhibited by the corresponding cost function. In contrast to previous approaches, Hough SCA is linear in the sample number and independent of the source dimension as well as robust against noise and outliers. Experiments demonstrate the flexibility of the proposed algorithm.

1. INTRODUCTION

One goal of multichannel signal analysis lies in the detection of underlying sources within some given set of observations. If both the mixture process and the sources are unknown, this is denoted as *blind source separation* (BSS). BSS can be applied in many different fields such as medical and biological data analysis, broadcasting systems, and audio and image processing. In order to decompose the data set, different assumptions on the sources have to be made. The most common assumption currently used is statistical independence of the sources, which leads to the task of *independent component analysis* (ICA); see, for instance, [1, 2] and references therein. ICA very successfully separates data in the linear *complete* case, when as many signals as underlying sources are observed, and in this case the mixing matrix and the sources are identifiable except for permutation and scaling [3, 4]. In the *overcomplete* or *underdetermined* case, fewer observations than sources are given. It can be shown that the mixing matrix can still be recovered [5], but source identifiability does not hold. In order to approximately detect the sources, additional requirements have to be made, usually sparsity of the sources [6–8].

Recently, we have introduced a novel measure for sparsity and shown [9] that based on sparsity alone, we can still detect both mixing matrix and sources uniquely except for trivial indeterminacies (*sparse component analysis (SCA)*). In that paper, we have also proposed an algorithm based on random sampling for reconstructing the mixing matrix and the sources, but the focus of the paper was on the model, and the matrix estimation algorithm turned out to be not very robust against noise and outliers, and could therefore not easily be applied in high dimensions due to the involved combinatorial searches. In the present manuscript, a new algorithm is proposed for SCA, that is, for decomposing a data set $\mathbf{x}(1),\ldots,\mathbf{x}(T) \in \mathbb{R}^m$ modeled by an $(m \times T)$-matrix $\mathbf{X}$ linearly into $\mathbf{X} = \mathbf{AS}$, where the n-dimensional sources $\mathbf{S} = (\mathbf{s}(1),\ldots,\mathbf{s}(T))$ are assumed to be sparse at every time instant. If the sources are of sufficiently high sparsity, the mixtures are clustered along hyperplanes in the mixture space. Based on this condition, the mixing matrix can be reconstructed; furthermore, this property is robust against noise and outliers, which will be used here. The proposed algorithm denoted by *Hough SCA* employs a generalization of the Hough transform in order to detect the hyperplanes in the mixture space, which then leads to matrix and source identification.

The Hough transform [10] is a standard tool in image analysis that allows recognition of global patterns in an image space by recognizing local patterns, ideally a point, in a transformed parameter space. It is particularly useful when the patterns in question are sparsely digitized, contain "holes," or have been taking in noisy environments. The basic idea of this technique is to map parameterized objects such as straight lines, polynomials, or circles to a suitable parameter space. The main application of the Hough transform lies in the field of image processing in order to find straight lines, centers of circles with a fixed radius, parabolas, and so forth in images.

The Hough transform has been used in a somewhat ad hoc way in the field of independent component analysis for identifying two-dimensional sources in the mixture plot in the complete [11] and overcomplete [12] cases, which without additional restrictions can be shown to have some theoretical issues [13]; moreover, the proposed algorithms were restricted to two dimensions and did not provide any reliable source identification method. An application of a time-frequency Hough transform to direction finding within nonstationary signals has been studied in [14]; the idea is based on the Hough transform of the Wigner-Ville distribution [15], essentially employing a generalized Hough transform [16] to find straight lines in the time-frequency plane. The results in [14] again only concentrate on the two-dimensional mixture case. In the literature, overcomplete BSS and the corresponding basis estimation problems have gained considerable interest in the past decade [8, 17–19], but the sparse priors are always used in connection with the assumption of independent sources. This allows for probabilistic sparsity conditions, but cannot guarantee source identifiability as in our case.

The paper is organized as follows. In Section 2, we introduce the overcomplete SCA model and summarize the known identifiability results and algorithms [9]. The following section then reviews the classical Hough transform in two dimensions and generalizes it in order to detect hyperplanes in any dimension. This method is used in section Section 4 to develop an SCA algorithm, which turns out to be highly robust against noise and outliers. We confirm this by experiments in Section 5. Some results of this paper have already been presented at the conference "ESANN 2004" [20].

2.   OVERCOMPLETE SCA

We introduce a strict notion of sparsity and present identifiability results when applying the measure to BSS.

A vector $\mathbf{v} \in \mathbb{R}^n$ is said to be k-sparse if $\mathbf{v}$ has at least k zero entries. An $n \times T$ data matrix is said to be k-sparse if each of its columns is k-sparse. Note that $\mathbf{v}$ is k-sparse, then it is also k'-sparse for $k' \leq k$. The goal of sparse component analysis of level k (k-SCA) is to decompose a given m-dimensional observed signal $\mathbf{x}(t)$, $t = 1, \ldots, T$, into

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \qquad (1)$$

with a real $m \times n$-mixing matrix $\mathbf{A}$ and an n-dimensional

k-sparse sources $\mathbf{s}(t)$. The samples are gathered into corresponding data matrices $\mathbf{X} := (\mathbf{x}(1), \ldots, \mathbf{x}(T)) \in \mathbb{R}^{m \times T}$ and $\mathbf{S} := (\mathbf{s}(1), \ldots, \mathbf{s}(T)) \in \mathbb{R}^{n \times T}$, so the model is $\mathbf{X} = \mathbf{A}\mathbf{S}$. We speak of complete, overcomplete, or undercomplete k-SCA if $m = n$, $m < n$, or $m > n$, respectively. In the following, we will always assume that the sparsity level equals $k = n - m + 1$, which means that at any time instant, fewer sources than given observations are active. In the algorithm, we will also consider additive white Gaussian noise; however, the model identification results are presented only in the noiseless case from (1).

Note that in contrast to the ICA model, the above problem is not translation invariant. However, it is easy to see that if instead of $\mathbf{A}$ we choose an affine linear transformation, the translation constant can be determined from $\mathbf{X}$ only, as long as the sources are nondeterministic. Put differently, this means that instead of assuming k-sparsity of the sources we could also assume that at any fixed time t, only $n - k$ source components are allowed to vary from a previously fixed constant (which can be different for each source). In the following without loss of generality we will assume $m \leq n$: the easier undercomplete (or underdetermined) case can be reduced to the complete case by projection in the mixture space.

The following theorem shows that essentially the mixing model (1) is unique if fewer sources than mixtures are active, that is, if the sources are $(n - m + 1)$-sparse.

Theorem 1 (matrix identifiability). *Consider the k-SCA problem from (1) for $k := n - m + 1$ and assume that every $m \times m$-submatrix of $\mathbf{A}$ is invertible. Furthermore, let $\mathbf{S}$ be sufficiently rich represented in the sense that for any index set of $n - m + 1$ elements $I \subset \{1, \ldots, n\}$ there exist at least m samples of $\mathbf{S}$ such that each of them has zero elements in places with indexes in I and each $m - 1$ of them are linearly independent. Then $\mathbf{A}$ is uniquely determined by $\mathbf{X}$ except for left multiplication with permutation and scaling matrices.*

So if $\mathbf{A}\mathbf{S} = \hat{\mathbf{A}}\hat{\mathbf{S}}$, then $\mathbf{A} = \hat{\mathbf{A}}\mathbf{P}\mathbf{L}$ with a permutation $\mathbf{P}$ and a nonsingular scaling matrix $\mathbf{L}$. This means that we can recover the mixing matrix from the mixtures. The next theorem shows that in this case also the sources can be found uniquely.

Theorem 2 (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 an $(n-m+1)$-sparse solution, that is, one with at least $n - m + 1$ zero components. If $\mathbf{A}$ fulfills the condition from Theorem 1, 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 we refer to [9]. The above two theorems show that in the case of overcomplete BSS using k-SCA with $k = n - m + 1$, both the mixing matrix and the sources can uniquely be recovered from $\mathbf{X}$ except for the omnipresent permutation and scaling indeterminacy. The essential idea of both theorems as well as a possible algorithm is

(a) Three hyperplanes span$\{\mathbf{a}_i,$ $\mathbf{a}_j\}$ for $1 \le i < j \le 3$ in the 3×3 case

(b) Hyperplanes from (a) visualized by intersection with the sphere

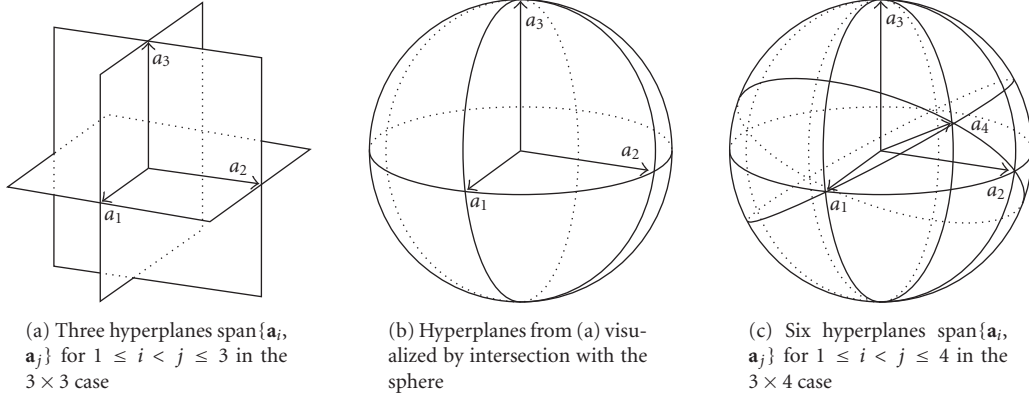(c) Six hyperplanes span$\{\mathbf{a}_i,$ $\mathbf{a}_j\}$ for $1 \le i < j \le 4$ in the 3×4 case

FIGURE 1: Visualization of the hyperplanes in the mixture space $\{\mathbf{x}(t)\} \subset \mathbb{R}^3$. Due to the source sparsity, the mixtures are generated by only two matrix columns $\mathbf{a}_i$, $\mathbf{a}_j$, and are hence contained in a union of hyperplanes. Identification of the hyperplanes gives mixing matrix and sources.

Data: samples $\mathbf{x}(1),\dots,\mathbf{x}(T)$
Result: estimated mixing matrix $\hat{\mathbf{A}}$

Hyperplane identification.

(1) Cluster the samples $\mathbf{x}(t)$ in $\binom{n}{m-1}$ groups such that the span of the elements of each group produces one distinct hyperplane H_i.

Matrix identification.

(2) Cluster the normal vectors to these hyperplanes in the smallest number of groups G_j, $j = 1,\dots,n$ (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 $\widehat{H}_j$.

(3) Calculate the normal vectors $\hat{\mathbf{a}}_j$ to each hyperplane $\widehat{H}_j$, $j = 1,\dots,n$.

(4) The matrix $\hat{\mathbf{A}}$ with columns $\hat{\mathbf{a}}_j$ is an estimate of the mixing matrix (up to permutation and scaling of the columns).

ALGORITHM 1: SCA matrix identification algorithm.

Data: samples $\mathbf{x}(1),\dots,\mathbf{x}(T)$ and estimated mixing matrix $\hat{\mathbf{A}}$
Result: estimated sources $\hat{\mathbf{s}}(1),\dots,\hat{\mathbf{s}}(T)$

(1) Identify the set of hyperplanes $\mathcal{H}$ produced by taking the linear hull of every subsets of the columns of $\hat{\mathbf{A}}$ with $m - 1$ elements

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

(2)   Identify the hyperplane $H \in \mathcal{H}$ containing $\mathbf{x}(t)$, or, in the presence of noise, identify the one to which the distance from $\mathbf{x}(t)$ is minimal and project $\mathbf{x}(t)$ onto H to $\tilde{\mathbf{x}}$.

(3)   If H is produced by the linear hull of column vectors $\hat{\mathbf{a}}_{i_1},\dots,\hat{\mathbf{a}}_{i_{m-1}}$, find coefficients $\lambda_{i(j)}$ such that $\tilde{\mathbf{x}} = \sum_{j=1}^{m-1} \lambda_{i(j)}\hat{\mathbf{a}}_{i(j)}$.

(4)   Construct the solution $\hat{\mathbf{s}}(t)$: it contains $\lambda_{i(j)}$ at index $i(j)$ for $j = 1,\dots,m-1$, the other components are zero.

end

ALGORITHM 2: SCA source identification algorithm.

---

illustrated in Figure 1: by assuming sufficiently high sparsity of the sources, the mixture space clusters along a union of hyperplanes, which uniquely determine both mixing matrix and sources.

The matrix and source identification algorithm from [9] are recalled in Algorithms 1 and 2. We will present a modification of the matrix identification part—the same source identification algorithm (Algorithm 2) will be used in the experiments. The "difficult" part of the matrix identification algorithm lies in the hyperplane detection; in Algorithm 1, a random sampling and clustering technique is used. Another more efficient algorithm for finding the hyperplanes containing the data has been developed by Bradley and Mangasarian [21], essentially by extending k-means batch clustering. Their so-called *k-plane clustering algorithm* in the special case of hyperplanes containing 0 is shown in Algorithm 3. The

finite termination of the algorithm is proven in [21, Theorem 3.7]. We will later compare the proposed Hough algorithm with the k-hyperplane algorithm. The k-hyperplane algorithm has also been extended to a more general, orthogonal k-subspace clustering method [22, 23] thus allowing a search not only for hyperplanes but also for lower-dimensional subspaces.

3. HOUGH TRANSFORM

The Hough transform is a classical method for locating shapes in images, widely used in the field of image processing; see [10, 24]. 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 antialiasing techniques.

Data: samples $\mathbf{x}(1), \ldots, \mathbf{x}(T)$
Result: estimated k hyperplanes H_i given by the normal
    vectors $\mathbf{u}_i$
(l) Initialize randomly $\mathbf{u}_i$ with $|\mathbf{u}_i| = 1$ for $i = 1, \ldots, k$.
   do
      Cluster assignment.
      for $t \leftarrow 1, \ldots, T$ **do**
(2)   Add $\mathbf{x}(t)$ to cluster $\mathbf{Y}^{(i)}$, where i is chosen to
        minimize $|\mathbf{u}_i^\top \mathbf{x}(t)|$ (distance to hyperplane H_i).
      end
(3)   Exit if the mean distance to the hyerplanes is smaller
      than some preset value.
      Cluster update.
      for $i \leftarrow 1, \ldots, k$ **do**
(4)   Calculate the i-th cluster correlation $\mathbf{C} := \mathbf{Y}^{(i)} \mathbf{Y}^{(i)\top}$.
(5)   Choose an eigenvector $\mathbf{v}$ of $\mathbf{C}$ corresponding to
        a minimal eigenvalue.
(6)   Set $\mathbf{u}_i \leftarrow \mathbf{v}/|\mathbf{v}|$.
      end
   end

ALGORITHM 3: k-hyperplane clustering algorithm.

3.1. Definition

Its main idea can be described as follows: consider a parameterized object

$$M_\mathbf{a} := \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{f}(\mathbf{x}, \mathbf{a}) = 0\} \tag{2}$$

for a fixed parameter set $\mathbf{a} \in U \subset \mathbb{R}^p$—here $U \subset \mathbb{R}^p$ is the *parameter space*, and the *parameter function* $\mathbf{f} : \mathbb{R}^n \times U \to \mathbb{R}^m$ is a set of m equations describing our types of objects (manifolds) $M_\mathbf{a}$ for different parameters $\mathbf{a}$. We assume that the equations given by $\mathbf{f}$ are *separating* in the sense that if $M_\mathbf{a} \subset M_{\mathbf{a}'}$, then already $\mathbf{a} = \mathbf{a}'$. A simple example is the set of unit circles in $\mathbb{R}^2$; then $f(\mathbf{x}, \mathbf{a}) = |\mathbf{x} - \mathbf{a}| - 1$. For a given $\mathbf{a} \in \mathbb{R}^2$, $M_\mathbf{a}$ is the circle of radius 1 centered at $\mathbf{a}$. Obviously f is separated. Other object manifolds will be discussed later. A nonseparated object function is, for example, $f(x, a) := 1 - 1_{[0,a]}(x)$ for $(x, a) \in \mathbb{R} \times [0, \infty)$, where the characteristic function $1_{[0,a]}(x)$ equals 1 if and only if $x \in [0, a]$ and 0 otherwise. Then $M_1 = [0, 1] \subset [0, 2] = M_2$ but the parameters are different.

Given a separating parameter function $\mathbf{f}(\mathbf{x}, \mathbf{a})$, its *Hough transform* is defined as

$$\begin{aligned} \eta[\mathbf{f}] : \mathbb{R}^n &\longrightarrow \mathcal{P}(U), \\ \mathbf{x} &\longmapsto \{\mathbf{a} \in U \mid \mathbf{f}(\mathbf{x}, \mathbf{a}) = 0\}, \end{aligned} \tag{3}$$

where $\mathcal{P}(U)$ denotes the set of all subsets of U. So $\eta[\mathbf{f}]$ maps a point $\mathbf{x}$ onto the set of all parameters describing objects containing $\mathbf{x}$. But an object $M_\mathbf{a}$ as a set is mapped onto a single point $\{\mathbf{a}\}$, that is,

$$\bigcap_{\mathbf{x} \in M_\mathbf{a}} \eta[\mathbf{f}](\mathbf{x}) = \{\mathbf{a}\}. \tag{4}$$

This follows because if $\bigcap_{\mathbf{x} \in M_\mathbf{a}} \eta[\mathbf{f}](\mathbf{x}) = \{\mathbf{a}, \mathbf{a}'\}$, then for all $\mathbf{x} \in M_\mathbf{a}$ we have $\mathbf{f}(\mathbf{x}, \mathbf{a}') = 0$, which means that $M_\mathbf{a} \subset M_{\mathbf{a}'}$; the

parameter function $\mathbf{f}$ is assumed to be separating, so $\mathbf{a} = \mathbf{a}'$. Hence, objects $M_\mathbf{a}$ in a data set $\overline{\mathbf{X}} = \{\mathbf{x}(1), \ldots, \mathbf{x}(T)\}$ can be detected by analyzing clusters in $\eta[\mathbf{f}](\overline{\mathbf{X}})$.

We will illustrate this concept for line detection in the following section before applying it to the hyperplane identification needed for our SCA problem.

3.2. Classical Hough transform

The *(classical) Hough transform* detects lines in a given two-dimensional data space as follows: an affine, nonvertical line in $\mathbb{R}^2$ can be described by the equation $x_2 = a_1 x_1 + a_2$ for fixed $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2$. If we define

$$f_L(\mathbf{x}, \mathbf{a}) := a_1 x_1 + a_2 - x_2, \tag{5}$$

then the above line equals the set $M_\mathbf{a}$ from (2) for the unique parameter $\mathbf{a}$, and $\mathbf{f}$ is clearly separating. Figures 2(a) and 2(b) illustrate this idea.

In practice, polar coordinates are used to describe the line in Hessian normal form; this allows to also detect vertical lines ($\theta = \pi/2$) in the data set, and moreover guarantees for an isotropic error in contrast to the parametrization (5). This leads to a parameter function

$$f_P(\mathbf{x}, \theta, \rho) = x_1 \cos(\theta) + x_2 \sin(\theta) - \rho = 0 \tag{6}$$

for parameters $(\theta, \rho) \in U := [0, \pi) \times \mathbb{R}$. Then points in data space are mapped to sine curves given by f; see Figure 2(c).

3.3. Generalization

The mixing matrix $\mathbf{A}$ in the case of $(n - m + 1)$-sparse SCA can be recovered by finding all 1-codimensional subvector spaces 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

$$f_h(\mathbf{x}, \mathbf{n}) := \mathbf{n}^\top \mathbf{x} = 0, \tag{7}$$

where $\mathbf{n}$ is a nonzero vector orthogonal to H. After normalization $|\mathbf{n}| = 1$, the normal vector $\mathbf{n}$ is uniquely determined by H if we additionally require $\mathbf{n}$ to lie on one hemisphere of the unit sphere $S^{n-1} := \{\mathbf{x} \in \mathbb{R}^n \mid |\mathbf{x}| = 1\}$. This means that the parametrization f_h is separating. In terms of spherical coordinates of S^{n-1}, $\mathbf{n}$ can be expressed as

$$\mathbf{n} = \begin{pmatrix} \cos \varphi & \sin \theta_1 \sin \theta_2 & \cdots & \sin \theta_{m-2} \\ \sin \varphi & \sin \theta_1 \sin \theta_2 & \cdots & \sin \theta_{m-2} \\ & \cos \theta_1 \sin \theta_2 & \cdots & \sin \theta_{m-2} \\ & \vdots & \ddots & \vdots \\ & \cos \theta_1 \cos \theta_2 & \cdots & \cos \theta_{m-2} \end{pmatrix} \tag{8}$$

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 $\varphi \in [0, \pi)$. Plugging $\mathbf{n}$ in spherical coordinates into (7) gives

$$\cot \theta_{m-2} = -\sum_{i=1}^{m-1} \nu_i(\varphi, \theta_1, \ldots, \theta_{m-3}) \frac{x_i}{x_m} \tag{9}$$

for $\mathbf{x} \in \mathbb{R}^m$ with $x_m \neq 0$ and

$$
v_i := \begin{cases}
\cos \varphi \prod_{j=1}^{m-3} \sin \theta_j, & i = 1, \\
\sin \varphi \prod_{j=1}^{m-3} \sin \theta_j, & i = 2, \\
\prod_{j=1}^{i-2} \cos \theta_j \prod_{j=i-1}^{m-3} \sin \theta_j, & i > 2.
\end{cases}
\tag{10}
$$

With $\cot(\theta + \pi/2) = -\tan(\theta)$ we finally get $\theta_{m-2} = \arctan \left(\sum_{i=1}^{m-1} v_i x_i / x_m \right) + \pi/2$. 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

$$
\eta[f_h] : \mathbb{R}^m \longrightarrow \mathcal{P}\left([0,\pi)^{m-1} \right),
$$

$$
\mathbf{x} \longmapsto \Bigg\{ (\varphi, \theta_1, \ldots, \theta_{m-2})
$$

$$
\in [0,\pi)^{m-1} \mid \theta_{m-2} = \arctan \left(\sum_{i=1}^{m-1} v_i \frac{x_i}{x_m} \right) + \frac{\pi}{2} \Bigg\}.
\tag{11}
$$

The parametrization f_h is separating, so 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 3. The hyperplane structures of a data set $\overline{\mathbf{X}} = \{\mathbf{x}(1), \ldots, \mathbf{x}(T)\}$ can be analyzed by finding clusters in $\eta[f_h](\overline{\mathbf{X}})$.

Let $\mathbb{RP}^{m-1}$ denote the $(m-1)$-dimensional *real projective space*, that is, the manifold of all 1-dimensional subspaces of $\mathbb{R}^m$. There is a canonical diffeomorphism between $\mathbb{RP}^{m-1}$ and the Grassmanian manifold of all $(m-1)$-dimensional subspaces of $\mathbb{R}^m$, induced by the scalar product. Using this diffeomorphism, we can reformulate our aim of identifying hyperplanes as finding elements of $\mathbb{RP}^{m-1}$. So, the Hough transform $\eta[f_h]$ maps $\mathbf{x}$ onto a subset of $\mathbb{RP}^{m-1}$, which is topologically equivalent to the upper hemisphere in $\mathbb{R}^m$ with identifications along the boundary. In fact, in (11) we simply have constructed a coordinate map of $\mathbb{RP}^{m-1}$ using spherical coordinates.

4. HOUGH SCA ALGORITHM

The SCA matrix detection algorithm (Algorithm 1) consists of two steps. In the first step, $d := \binom{n}{m-1}$ hyperplanes given by their normal vectors $\mathbf{n}^{(1)}, \ldots, \mathbf{n}^{(d)}$ are constructed such that the mixture data lies in the union of these hyperplanes—in the case of noise this will hold only approximately. 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*.
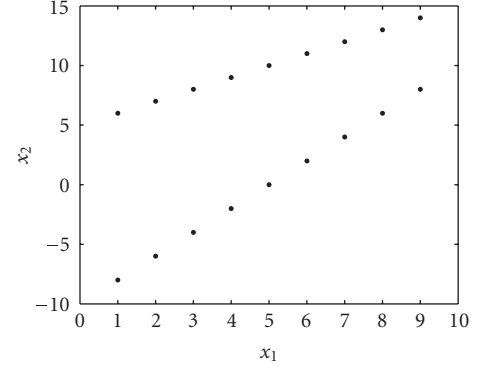
(a) Data space
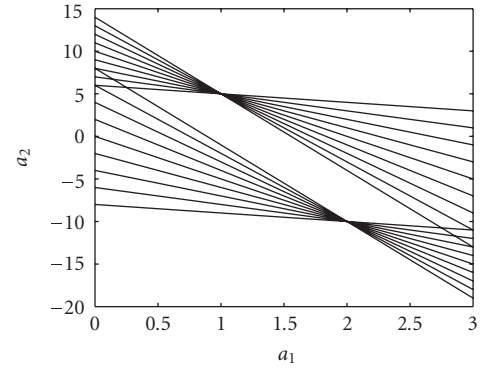
(b) Linear Hough space
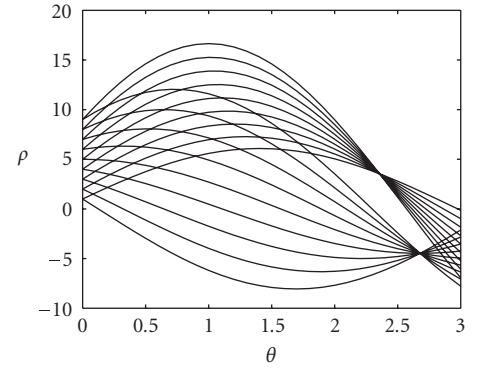
(c) Polar Hough space

FIGURE 2: Illustration of the "classical" Hough transform: a point (x_1, x_2) in the data space (a) is mapped (b) onto the line $\{(a_1, a_2) \mid a_2 = -a_1 x_1 + x_2\}$ in the linear parameter space $\mathbb{R}^2$ or (c) onto a translated sine curve $\{(\theta, \rho) \mid \rho = x_1 \cos \theta + x_2 \sin \theta\}$ in the polar parameter space $[0, \pi) \times \mathbb{R}_0^+$. The Hough curves of points belonging to one line in data space intersect in precisely one point $\mathbf{a}$ in the Hough space—and the data points lie on the line given by the parameter $\mathbf{a}$.

4.1. Definition

The idea is to first gather the Hough curves $\eta[f_h](\mathbf{x}(t))$ corresponding to the samples $\mathbf{x}(t)$ in a discretized parameter space, in this context often called *Hough accumulator*. Plotting these curves in the accumulator is sometimes denoted as *voting* for each bin, similar to histogram generation. According to the previous section, all points $\mathbf{x}$ from some
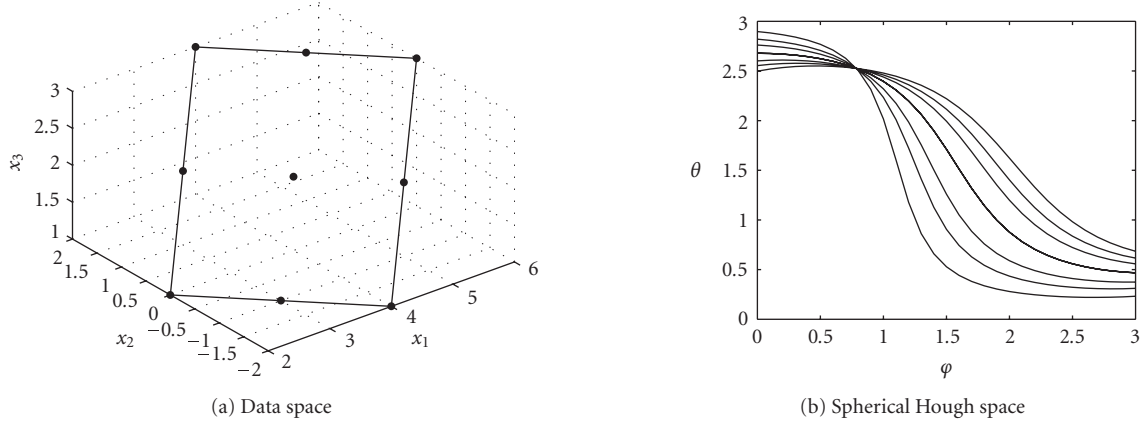
(a) Data space

(b) Spherical Hough space

FIGURE 3: Illustration of the "hyperplane detecting" Hough transform in three dimensions: a point (x_1, x_2, x_3) in the data space (a) is mapped onto the curve $\{(\varphi, \theta) \mid \theta = \arctan(x_1 \cos \varphi + x_2 \sin \varphi) + \pi/2\}$ in the parameter space $[0, \pi)^2$ (b). The Hough curves of points belonging to one plane in data space intersect in precisely one point (φ, θ) in the Hough space and the points lie on the plane given by the normal vector $(\cos \varphi \sin \theta, \sin \varphi \sin \theta, \cos \theta)$.

hyperplane H given by a normal vector with angles $(\varphi, \boldsymbol{\theta})$ are mapped onto a parameterized object that contains $(\varphi, \boldsymbol{\theta})$ for all possible $\mathbf{x} \in H$. Hence, the corresponding angle bin will contain votes from all samples $\mathbf{x}(t)$ lying in H, whereas other bins receive much less votes. Therefore, maxima analysis of the accumulator gives the hyperplanes in the parameter space. This idea corresponds to clustering all possible normal vectors of planes through $\mathbf{x}(t)$ on $\mathbb{RP}^{m-1}$ for all t. The resulting *Hough SCA* algorithm is described in Algorithm 4. We see that only the hyperplane identification step is different from Algorithm 1, the matrix identification is the same.

The number β of bins is also called the *grid resolution*. Similar to histogram-based density 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; see next section. Note that Hough SCA performs a global search hence it is expected to be much slower than local update algorithms such as Algorithm 3, but also much more robust. In the following, its properties will be discussed; applications are given in the example in Section 5.

4.2. Complexity

We will only discuss the complexity of the hyperplane estimation because the matrix identification is performed on a data set of size d being typically much smaller than the sample size T.

The angle θ_{m-2} has to be calculated $T\beta^{m-2}$ times. Due to the fact that only discrete values of the angles are of interest, the trigonometric functions as well as the ν_i can be precalculated and stored in exchange for speed. Then each calculation of θ_{m-2} involves $2m - 1$ operations (sum and product/division). The voting (without taking "lookup" costs in the accumulator into account) costs an additional operation. Altogether the accumulator can be filled with $2T\beta^{m-2}m$

Data: Samples $\mathbf{x}(1), \dots, \mathbf{x}(T)$ of the random vector $\mathbf{X}$
Result: Estimated mixing matrix $\hat{\mathbf{A}}$

Hyperplane identification.

(1) Fix the number β of bins (can be separate for each angle).
(2) Initialize the $\beta \times \cdots \times \beta$ ($m - 1$ terms) array $\alpha \in \mathbb{R}^{\beta^{m-1}}$ with zeros (accumulator).
   for $t \leftarrow 1, \dots, T$ **do**
       for $\varphi, \theta_1, \dots, \theta_{m-3} \leftarrow 0, \pi/\beta, \dots, (\beta - 1)\pi/\beta$ **do**
(3) $\theta_{m-2} \leftarrow \arctan(\sum_{i=1}^{m-1} \nu_i(\varphi, \dots, \theta_{m-3})x_i(t)/x_m(t)) + \pi/2$
(4)   Increase (vote for) the accumulator value of α in bin corresponding to $(\varphi, \theta_1, \dots, \theta_{m-2})$ by one.
       end
   end
(5) The $d := \binom{n}{m-1}$ largest local maxima of α correspond to the d hyperplanes present in the data set.
(6) Back transformation as in (8) gives the corresponding normal vectors $\mathbf{n}^{(1)}, \dots, \mathbf{n}^{(d)}$ to those hyperplanes.

Matrix identification.

(7) Clustering of hyperplanes generated by $(m - 1)$-tuples in $\{\mathbf{n}^{(1)}, \dots, \mathbf{n}^{(d)}\}$ gives n separate hyperplanes.
(8) Their normal vectors are the n columns of the estimated mixing matrix $\hat{\mathbf{A}}$.

ALGORITHM 4: Hough SCA algorithm for mixing matrix identification.

---

operations. This means that the algorithm depends linearly on the sample size and is polynomial in the grid resolution and exponential in the mixture dimension. The maxima search involves $\mathcal{O}(\beta^{m-1})$ operations, which for small to medium dimensions can be ignored in comparison to the accumulator generation because usually $\beta \ll T$.

So the main part of the algorithm does not depend on the source dimension n but only on the mixture dimension m. This means for applications that n can be quite large but

hyperplanes will still be found if the grid resolution is high enough. Increase of the grid resolution (in polynomial time) results in increased accuracy also for higher source dimensions n. The memory requirement of the algorithm is dominated by the accumulator size, which is β^{m-1}. This can limit the grid resolution.

4.3. Resolution error

The choice of the grid resolution β in the algorithm induces a systematic *resolution error* in the estimation of $\mathbf{A}$ (as trade-off for robustness and speed). This error is calculated in this section.

Let $\mathbf{A}$ be the unknown mixing matrix and $\hat{\mathbf{A}}$ its estimate, constructed by the Hough SCA algorithm (Algorithm 4) with grid resolution β. Let $\mathbf{n}^{(1)}, \ldots, \mathbf{n}^{(d)}$ be the normal vectors of hyperplanes generated by $(m-1)$-tuples of columns of $\mathbf{A}$ and let $\hat{\mathbf{n}}^{(1)}, \ldots, \hat{\mathbf{n}}^{(d)}$ be their corresponding estimates. Ignoring permutations, it is sufficient to only describe how $\hat{\mathbf{n}}^{(i)}$ differs from $\mathbf{n}^{(i)}$.

Assume that the maxima of the accumulator are correctly estimated, but due to the discrete grid resolution, an average error of $\pi/2\beta$ is made when estimating the precise maximum position, because the size of one bin is π/β. How is this error propagated into $\hat{\mathbf{n}}^{(i)}$? By assumption each estimate $\hat{\varphi}, \hat{\theta}_1, \ldots, \hat{\theta}_{m-2}$ differs from $\varphi, \theta_1, \ldots, \theta_{m-2}$ maximally by $\pi/2\beta$. As we are only interested in an upper boundary, we simply calculate the deviation of each component of $\hat{\mathbf{n}}^{(i)}$ from $\mathbf{n}^{(i)}$. Using the fact that sine and cosine are bounded by one, (8) then gives us estimates $|\hat{n}_j^{(i)} - n_j^{(i)}| \le (m-1)\pi/(2\beta)$ for coordinate j, so altogether

$$\left| \hat{\mathbf{n}}^{(i)} - \mathbf{n}^{(i)} \right| \le \frac{(m-1)\sqrt{m}\pi}{2\beta}. \tag{12}$$

This estimate may be improved by using the Jacobian of the spherical coordinate transformation and its determinant, but for our purpose this boundary is sufficient. In summary, we have shown that the grid resolution contributes to a β^{-1}-perturbation in the estimation of $\mathbf{A}$.

4.4. Robustness

Robustness with regard to additive noise as well as outliers is important for any algorithm to be used in the real world. Here an outlier is roughly defined to be a sample far away from other observations, and indeed some researchers define outliers to be sample further away from the mean than say 5 standard deviations. However, such definitions do necessarily depend on the underlying random variable to be estimated, so most books only give examples of outliers, and indeed no consistent, context-free, precise definition of outliers exists [25]. In the following, given samples of a fixed random variable of interest, we denote a sample as *outlier* if it is drawn from another sufficiently different distribution.

Data fitting of only one hyperspace to the data set can be achieved by linear regression namely by minimizing the squared distance to such a possible hyperplane. These least

squares fitting algorithms are well known to be sensitive to outliers, and various extensions of the LS method such as least median of squares and reweighted least squares [26] have been developed to overcome this problem. The *breakdown point* of the latter is 0.5, which means that the fit parameters are only stably estimated for data sets with less than 50% outliers. The other techniques typically have much lower breakdown points, usually below 0.3. The classical Hough transform, albeit no regression method, is comparable in terms of breakdown with robust fitting algorithms such as the reweighted least squares algorithm [27]. In the experiments we will observe similar results for the generalized method presented above. Namely, we achieve breakdown levels of up to 0.8 in the low-noise case, which considerably decrease with increasing noise.

From a mathematical point of view, the "classical" Hough transform as an estimator (and extension of linear regression) as well as regarding algorithmic and implementational aspects has been studied quite extensively; see, for example, [28] and references therein. Most of the presented theoretical results in the two-dimensional case could be extended to the more general objective presented here, but this is not within the scope of this manuscript. Simulations giving experimental evidence that the robustness also holds in our case are shown in Section 5.

4.5. Extensions

The following possible extensions to the Hough SCA algorithm can be employed to increase its performance.

If the noise level is known, *smoothing of the accumulator* (antialiasing) will help to give more robust results in terms of noise. For smoothing (usually with a Gaussian), the smoothing radius must be set according to the noise level. If the noise level is not known, smoothing can still be applied by gradually increasing the radius until the number of clearly detectable maxima equals d.

Furthermore, an additional *fine-tuning step* is possible: the estimated plane norms are slightly deteriorated by the systematic resolution error as shown previously. However, after application of Hough SCA, the data space can now be clustered into data points lying close to corresponding hyperplanes. Within each cluster linear regression (or some more robust version of it; see Section 4.4) can now be applied to improve the hyperplane estimate—this is actually the idea used locally in the k-hyperplane clustering algorithm (Algorithm 3). Such a method requires additional computational power, but makes the algorithm less dependent on the grid resolution, which is only needed for the hyperplane clustering step. However, it is expected that this additional fine-tuning step may decrease robustness especially against biased noise and outliers.

5. SIMULATIONS

We give a simulation example as well as batch runs to analyze the performance of the proposed algorithm.

(a) Source signals

(b) Mixture signals

(c) Normalized mixture scatter plot
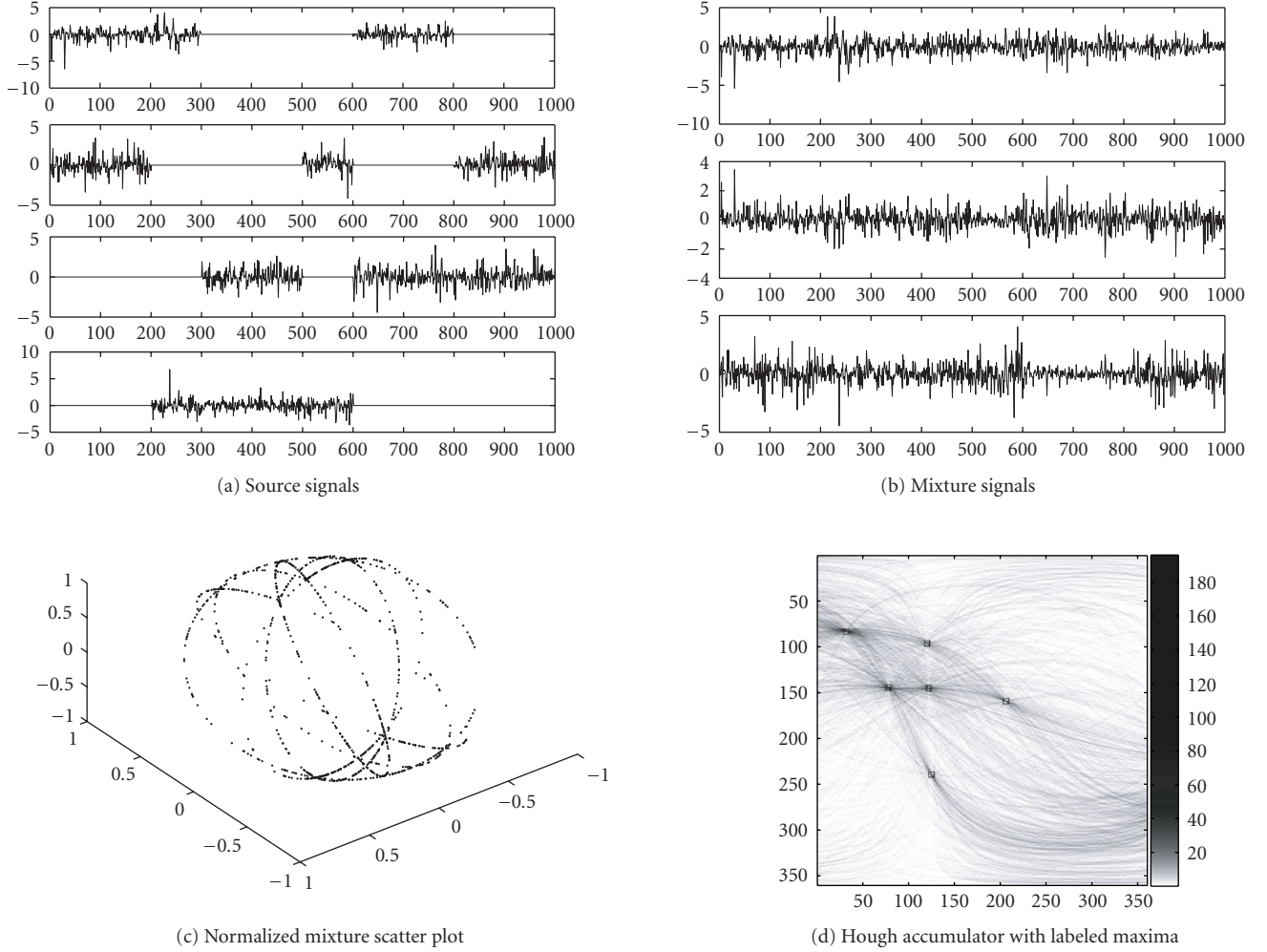

(d) Hough accumulator with labeled maxima

Figure 4: Example: (a) shows the 2-sparse, sufficiently rich represented, 4-dimensional source signals, and (b) the randomly mixed, 3-dimensional mixtures. The normalized mixture scatter plot $\{\mathbf{x}(t)/|\mathbf{x}(t)| \mid t = 1, \dots, T\}$ is given in (c), and the generated Hough accumulator in (d); note that the color scale in (d) was chosen to be nonlinear ($\gamma_{\mathrm{new}} := (1 - \gamma/\max)^{10}$) in order to visualize structure in addition to the strong maxima.

5.1. Explicit example

In the first experiment, we consider the case of source dimensions $n = 4$ and mixing dimension $m = 3$. The 4-dimensional sources have been generated from i.i.d. samples (two Laplacian and two Gaussian sequences) followed by setting some entries to zero in order to fulfill the sparsity constraints; see Figure 4(a). They 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 is needed by 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 4(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, as indicated by the normalized scatter plot in Figure 4(c), similar to the illustration from Figure 1(c).

In order to detect the planes in the data space, we apply the generalized Hough transform as explained in Section 3.3. Figure 4(d) 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. Considering intersections of the hyperplanes, we notice that only 4 of them intersect in precisely 3 planes, and these 4 intersection lines are spanned by the matrix columns $\mathbf{a}_i$. For practical reasons, we recover these combinatorially from the plane norm vectors; see Algorithm 4. The deviation of the recovered mixing matrix $\widehat{\mathbf{A}}$ from the original mixing matrix $\mathbf{A}$ in the overcomplete case can be measured by the *generalized crosstalking error* [8] defined as $E(\mathbf{A}, \widehat{\mathbf{A}}) := \min_{\mathbf{M} \in \Pi} \|\mathbf{A} - \widehat{\mathbf{A}}\mathbf{M}\|$, where the minimum is taken over the group Π of all invertible real

$n \times n$-matrices where only one entry in each column differs from 0; $\|\cdot\|$ denotes a fixed matrix norm. In our case the generalized crosstalking error is very low with $E(\mathbf{A}, \hat{\mathbf{A}}) = 0.040$. This essentially means that the two matrices, after permutation, differ only by 0.04 with respect to the chosen matrix norm, in our case the (squared) Frobenius norm. Then, the sources are recovered using the source recovery algorithm (Algorithm 2) with the approximated mixing matrix $\hat{\mathbf{A}}$. The normalized signal-to-noise ratios (SNRs) of the recovered sources with respect to the original ones are high with 36, 38, 36, and 37 dB, respectively.

As a modification of the previous example, we now consider also additive noise. We use sources $\mathbf{S}$ (which have unit covariance) and mixing matrix $\mathbf{A}$ from above, but add 1% random white noise to the mixtures $\mathbf{X} = \mathbf{AS} + 0.01\mathbf{N}$ where $\mathbf{N}$ is a normal random vector. This corresponds to a still high mean SNR of 38 dB. When considering the normalized scatter plot, again the 6 planes are visible, but the additive noise deteriorates the clear separation of each plane. We apply the generalized Hough transform to the mixture data; however, because of the noise we choose a more coarse discretization ($\beta = 180$ bins). Curves in Hough space corresponding to a single plane do not intersect any more 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 similar to the above to get a low generalized crosstalking error of $E(\mathbf{A}, \hat{\mathbf{A}}) = 0.12$. The sources are recovered well with mean SNRs of 20 dB, which is quite satisfactory considering the noisy, overcomplete mixture situation.

The following example demonstrates the good performance in higher source dimensions. Consider 6-dimensional 2-sparse sources that are mixed again by matrix $\mathbf{A}$ with coefficients drawn uniformly from $[-1, 1]$. Application of the generalized Hough transform to the mixtures retrieves the plane norm vectors. The recovered mixing matrix has a low generalized crosstalking error of $E(\mathbf{A}, \hat{\mathbf{A}}) = 0.047$. However, if the noise level increases, the performance considerably drops because many maxima, in this case 15, have to be located in the accumulator. After recovering the sources with this approximated matrix $\hat{\mathbf{A}}$, we get SNRs of only 11, 8, 6, 10, 12, and 11 dB. The rather high source recovery error is most probably due to the sensitivity of the source recovery to slight perturbations in the approximated mixing matrix.

5.2. Outliers

We will now perform experiments systematically analyzing the robustness of the proposed algorithm with respect to outliers in the sense of model-violating samples.

In the first explicit example we consider the sources from Figure 4(a), but 80% of the samples have been replaced by outliers (drawn from a 4-dimensional normal distribution). Due to the high percentage of outliers, the mixtures, mixed by the same random 3×4 matrix $\mathbf{A}$ as before, do not obviously exhibit any clear hyperplane structure. As discussed in

Section 4.4, the Hough SCA algorithm is very robust against outliers. Indeed, in addition to a noisy background within the Hough accumulator, the intersection maxima are still noticeable, and local maxima detection finds the correct hyperplanes (cf. Figure 4(d)), although 80% of the data is corrupted. The recovered mixing matrix has an excellent generalized crosstalking error of $E(\mathbf{A}, \hat{\mathbf{A}}) = 0.040$. Of course the sparse source recovery from above cannot recover the outlying samples. Applying the corresponding algorithms, we get SNRs of the corrupted sources with the recovered ones of around 4 dB; source recovery with the pseudo-inverse of $\hat{\mathbf{A}}$ corresponding to maximum-likelihood recovery with a Gaussian prior gives somewhat better SNRs of around 6 dB. But the sparse recovery method has the advantage that it can detect outliers by measuring distance from the hyperplanes. So outlier rejection is possible. Note that we get similar results when the outliers are not added in the source space but only in the mixture space, that is, only after the mixing process.

We now perform a numerical comparison of the number of outliers versus the algorithm performance for varying noise level; see Figure 5. The rationale behind this is that already small noise levels in addition to the outliers might be enough to destroy maxima in the accumulator thus deteriorating the SCA performance. The same (uncorrupted) sources and mixing matrix from above are used. Numerically, we get breakdown points of 0.8 for the no-noise case, and values of 0.5, 0.3, and 0.1 with increasing noise levels of 0.1% (58 dB), 0.5% (44 dB), and 1% (38 dB). Better performances at higher noise levels could be achieved by applying antialiasing techniques before maxima detection as described in Section 4.5.

5.3. Grid resolution

In this section we will demonstrate numerical examples to confirm the linear dependence of the algorithm performance with the inverse grid resolution β^{-1}. We consider 4-dimensional sources $\mathbf{S}$ with 1000 samples, in which for each sample two source components were drawn out of a distribution uniform in $[-1, 1]$, and the other two were set to zero, so $\mathbf{S}$ is 2-sparse. For each grid resolution β we perform 50 runs, and in each run a new set of sources is generated as above. These are then mixed using a 3×4 mixing matrix $\mathbf{A}$ with random coefficients uniformly out of $[-1, 1]$. Application of the Hough SCA algorithm gives an estimated matrix $\hat{\mathbf{A}}$. In Figure 6 we plot the mean generalized crosstalking error $E(\mathbf{A}, \hat{\mathbf{A}})$ for each grid resolution. With increasing β the accuracy increases—a logarithmic plot indeed confirms the linear dependence on β^{-1} as stated in Section 4.3. Furthermore we see that for example for $\beta = 360$, among all $\mathbf{S}$ and $\mathbf{A}$ as above we get a mean crosstalking error of 0.23 ± 0.5.

5.4. Batch runs and comparison with hyperplane k-means

In the last example, we consider the case of $m = n = 4$, and do compare the proposed algorithm (now with a
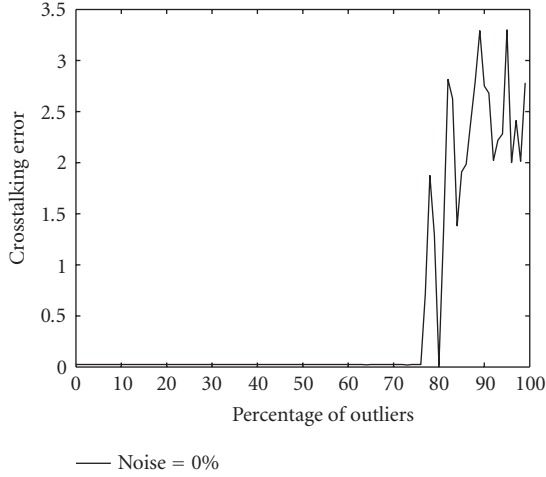
(a) Noiseless breakdown analysis with respect to outliers
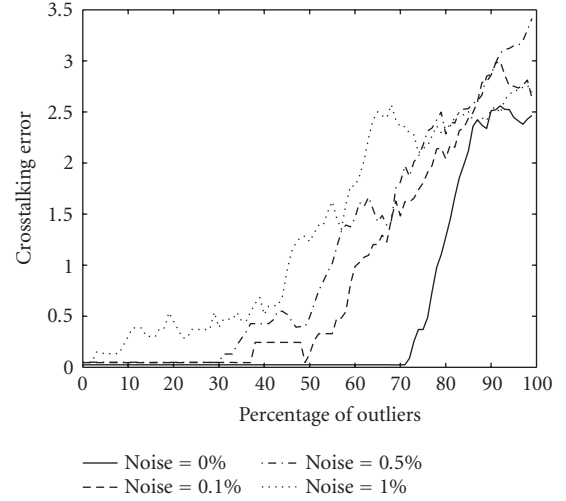
(b) Breakdown analysis for varying noise level

FIGURE 5: Performance of Hough SCA with increasing number of outliers. Plotted is the percentage of outliers in the source data versus the matrix recovery performance (measured by the generalized crosstalking error). For each 1%-step one calculation was performed; in (b) the plots have been smoothed by taking average over ten 1%-steps. In the no-noise case 360 bins were used, 180 bins in all other cases.
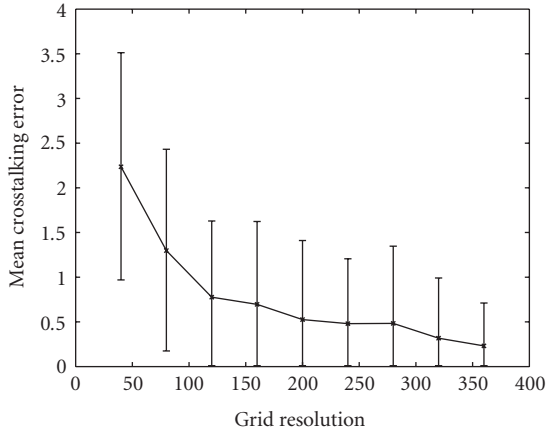
(a) Mean performance versus grid resolution
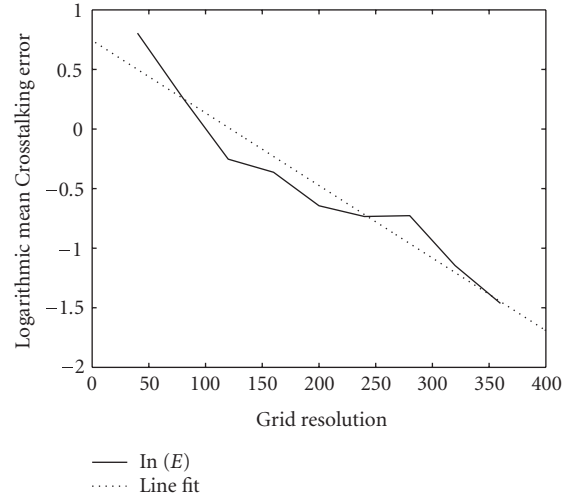
(b) Fit of logarithmic mean performance

FIGURE 6: Dependence of Hough SCA performance (a) on the grid resolution β; mean has been taken over 50 runs. With a logarithmic y-axis (b), a least squares line fit confirms the linear dependence of performance and β^{-1}.

three-dimensional accumulator) with k-hyperplane clustering algorithm (Algorithm 3). For this, random sources with $T = 10^5$ samples are uniformly drawn from $[-1, 1]$ uniform distribution, and a single coordinate is randomly set to zero, thus generating 1-sparse sources **S**. In 100 batch runs, a random 4×4 mixing matrix **A** with coefficients uniformly drawn from $[-1, 1]$, but columns normalized to 1 are constructed. The resulting mixtures $\mathbf{X} := \mathbf{AS}$ are then separated both by the proposed Hough k-SCA algorithm as well as the Bradley-Mangasarian k-hyperplane clustering algorithm (with 100 iterations, and without restart).

The resulting median crosstalking error $E(\mathbf{A}, \widehat{\mathbf{A}})$ of the Hough algorithm is 3.3 ± 2.3, and hence considerably lower than the k-hyperplane clustering result of 5.5 ± 1.9. This confirms the well-known fact that k-means and its extension exhibit local convergence only and are therefore susceptible to local minima, as seems to be the case in our example. A possible solution would be to use many restarts, but global convergence cannot be guaranteed. For practical applications, we therefore suggest using a rather rough (low grid resolution β) global search by Hough SCA followed by a finer local search using k-hyperplane clustering; see Section 4.5.

(a) Source signals

(b) Hough accumulator with three labeled maxima

(c) Recovered sources
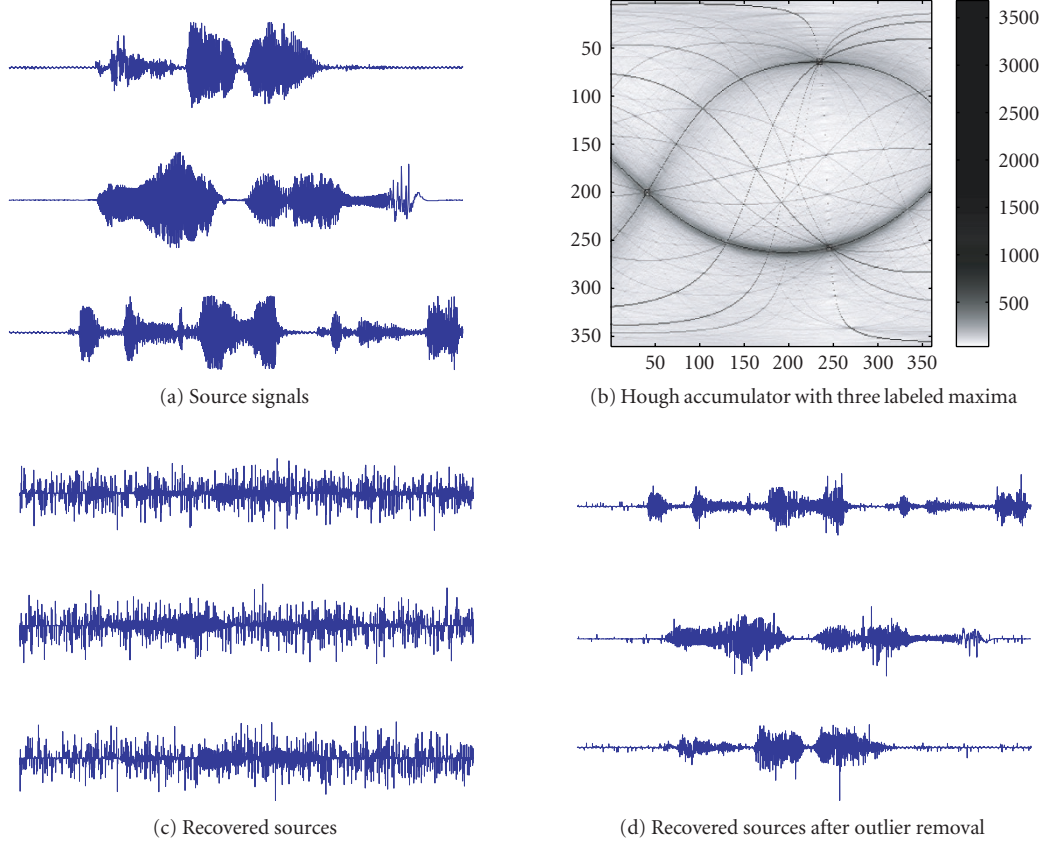

(d) Recovered sources after outlier removal

FIGURE 7: Application to speech signals: (a) shows the original speech sources ("peace and love," "hello, how are you," and "to be or not to be"), and (b) the Hough accumulator when trained to mixtures of (a) with 20% outliers. A nonlinear gray scale $\gamma_{new} := (1 - \gamma/\max)^{10}$ was chosen for better visualization. (c) and (d) present the recovered sources, without and with outlier removal. They coincide with (a) up to permutation (reversed order) and scaling.

5.5. Application to the separation of speech signals

In order to illustrate that the SCA assumptions are also valid for real data sets, we shortly present an application to audio source separation, namely, to the instantaneous, robust BSS of speech signals—a problem of importance in the field of audio signal processing. In the next section, we then refer to other works applying the model to biomedical data sets.

We consider three speech signals **S** of length 2.2s, sampled at 22000 Hz; see Figure 7(a). They are spoken by the same person, but may still be assumed to be independent. The signals are mixed by a randomly chosen mixing matrix **A** (coefficients uniform from $[-1, 1]$) to yield mixtures $\mathbf{X} = \mathbf{AS}$, but 20% outliers are introduced by replacing 20% of the samples of **X** by i.i.d. Gaussian samples. Without the outliers, more classical BSS algorithms such as ICA would have been able to perfectly separate the mixtures; however, in this noisy setting, ICA performs very poorly: application of the popular fastICA algorithm [29] yields only a poor estimate $\hat{\mathbf{A}}_f$ of the mixing matrix **A**, with high crosstalking error of $E(\mathbf{A}, \hat{\mathbf{A}}_f) = 3.73$.

Instead, we apply the complete-case Hough-SCA algorithm to this model with $\beta = 360$ bins—the sparseness assumption now means that we are searching for sources,

which have samples with at least one zero (quiet) source component. The Hough accumulator exhibits very nicely three strong maxima; see Figure 7(b). And indeed, the crosstalking error of the corresponding estimated mixing matrix $\hat{\mathbf{A}}$ with the original one is very low at $E(\mathbf{A}, \hat{\mathbf{A}}) = 0.020$. This experimentally confirms that speech signals obey an $(m-1)$-sparse signal model, at least if $m = n$. An explanation for this fact is that in typical speech data sets, considerable pauses are common, so with high probability we may find samples in which at least one source vanishes, and all such permutations occur—which is necessary for identifying the mixing matrix according to Theorem 1. We are dealing with a complete-case problem, so inverting $\hat{\mathbf{A}}$ directly yields recovered sources $\hat{\mathbf{S}}$. But of course due to the outliers, the SNR of $\hat{\mathbf{S}}$ with the original sources is low with only -1.35 dB. We therefore apply a simple outlier removal scheme by scanning each estimated source using a window of size $w = 10$ samples. An adjacent sample to the window is identified as outlier if its absolute value is larger than 20% of the maximal signal amplitude, but the window sample variance is lower than half of the variance when including the sample. The outliers are then replaced by the window average. This rough outlier-detection algorithm works satisfactorily well, see Figure 7(d);

the perceptual audio quality increased considerably, see also the differences between Figures 7(c) and 7(d), although the nominal SNR increase is only roughly 4.1 dB. Altogether, this example illustrates the applicability of the Hough SCA algorithm and its corresponding SCA model to audio data sets also in noisy settings, where ICA algorithms perform very poorly.

5.6. Other applications

We are currently studying several biomedical applications of the proposed model and algorithm, including the separation of functional magnetic resonance imaging data sets as well as surface electromyograms. For results on the former data set, we refer to the detailed book chapters [22, 23].

The results of the k-SCA algorithm applied to the latter signals are shortly summarized in the following. An electromyogram (EMG) denotes the electric signal generated by a contracting muscle; its study is relevant to the diagnosis of motoneuron diseases as well as neurophysiological research. In general, EMG measurements make use of invasive, painful needle electrodes. An alternative is to use surface EMGs, which are measured using noninvasive, painless surface electrodes. However, in this case the signals are rather more difficult to interpret due to noise and overlap of several source signals. When applying the k-SCA model to real recordings, Hough-based separation outperforms classical approaches based on filtering and ICA in terms of a greater reduction of the zero-crossings, a common measure to analyze the unknown extracted sources. The relative sEMG enhancement was $24.6 \pm 21.4\%$, where the mean was taken over a group of 9 subjects. For a detailed analysis, comparing various sparse factorization models both on toy and on real data, we refer to [30].

6. CONCLUSION

We have presented an algorithm for performing a global search for overcomplete SCA representations, and experiments confirm that Hough SCA is robust against noise and outliers with breakdown points up to 0.8. The algorithm employs hyperplane detection using a generalized Hough transform. Currently, we are working on applying the SCA algorithm to high-dimensional biomedical data sets to see how the different assumption of high sparsity contributes to the signal separation.

ACKNOWLEDGMENTS

REFERENCES

[1] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, John Wiley & Sons, New York, NY, USA, 2002.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, NY, USA, 2001.

[3] P. Comon, "Independent component analysis. A new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[4] F. J. Theis, "A new concept for separability problems in blind source separation," *Neural Computation*, vol. 16, no. 9, pp. 1827–1850, 2004.

[5] J. Eriksson and V. Koivunen, "Identifiability and separability of linear ica models revisited," in *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Source Separation (ICA '03)*, pp. 23–27, Nara, Japan, April 2003.

[6] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[7] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2197–2202, 2003.

[8] F. J. Theis, E. W. Lang, and C. G. Puntonet, "A geometric algorithm for overcomplete linear ICA," *Neurocomputing*, vol. 56, no. 1–4, pp. 381–398, 2004.

[9] P. Georgiev, F. J. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 992–996, 2005.

[10] P. V. C. Hough, "Machine analysis of bubble chamber pictures," in *International Conference on High Energy Accelerators and Instrumentation*, pp. 554–556, CERN, Geneva, Switzerland, 1959.

[11] J. K. Lin, D. G. Grier, and J. D. Cowan, "Feature extraction approach to blind source separation," in *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing (NNSP '97)*, pp. 398–405, Amelia Island, Fla, USA, September 1997.

[12] H. Shindo and Y. Hirai, "An approach to overcomplete-blind source separation using geometric structure," in *Proceedings of Annual Conference of Japanese Neural Network Society (JNNS '01)*, pp. 95–96, Naramachi Center, Nara, Japan, 2001.

[13] F. J. Theis, C. G. Puntonet, and E. W. Lang, "Median-based clustering for underdetermined blind signal processing," *IEEE Signal Processing Letters*, vol. 13, no. 2, pp. 96–99, 2006.

[14] L. Cirillo, A. Zoubir, and M. Amin, "Direction finding of nonstationary signals using a time-frequency Hough transform," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, pp. 2718–2721, Philadelphia, Pa, USA, March 2005.

[15] S. Barbarossa, "Analysis of multicomponent LFM signals by a combined Wigner-Hough transform," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1511–1515, 1995.

[16] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognition*, vol. 13, no. 2, pp. 111–122, 1981.

[17] T.-W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Processing Letters*, vol. 6, no. 4, pp. 87–90, 1999.

[18] K. Waheed and F. Salem, "Algebraic overcomplete independent component analysis," in *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Source Separation (ICA '03)*, pp. 1077–1082, Nara, Japan, April 2003.

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

[20] F. J. Theis, P. Georgiev, and A. Cichocki, "Robust overcomplete matrix recovery for sparse sources using a generalized Hough transform," in *Proceedings of 12th European Symposium on Artificial Neural Networks (ESANN '04)*, pp. 343–348, Bruges, Belgium, April 2004, d-side, Evere, Belgium.

[21] P. S. Bradley and O. L. Mangasarian, "*k*-plane clustering," *Journal of Global Optimization*, vol. 16, no. 1, pp. 23–32, 2000.

[22] P. Georgiev, P. Pardalos, F. J. Theis, A. Cichocki, and H. Bakardjian, "Sparse component analysis: a new tool for data mining," in *Data Mining in Biomedicine*, Springer, New York, NY, USA, 2005, in print.

[23] P. Georgiev, F. J. Theis, and A. Cichocki, "Optimization algorithms for sparse representations and applications," in *Multiscale Optimization Methods*, P. Pardalos, Ed., Springer, New York, NY, USA, 2005.

[24] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 204–208, 1972.

[25] R. Dudley, Department of Mathematics, MIT, course 18.465, 2005.

[26] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, John Wiley & Sons, New York, NY, USA, 1987.

[27] P. Ballester, "Applications of the Hough transform," in *Astronomical Data Analysis Software and Systems III*, J. Barnes, D. R. Crabtree, and R. J. Hanisch, Eds., vol. 61 of *ASP Conference Series*, 1994.

[28] A. Goldenshluger and A. Zeevi, "The Hough transform estimator," *Annals of Statistics*, vol. 32, no. 5, pp. 1908–1932, 2004.

[29] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[30] F. J. Theis and G. A. García, "On the use of sparse signal decomposition in the analysis of multi-channel surface electromyograms," *Signal Processing*, vol. 86, no. 3, pp. 603–623, 2006.

Fabian J. Theis obtained his M.S. degree in mathematics and physics from the University of Regensburg, Germany, in 2000. He also received the Ph.D. degree in physics from the same university in 2002 and the Ph.D. degree in computer science from the University of Granada in 2003. He worked as a Visiting Researcher at the Department of Architecture and Computer Technology (University of Granada, Spain), at the RIKEN Brain Science Institute (Wako, Japan), at FAMU-FSU (Florida State University, USA), and at TUAT's Laboratory for Signal and Image Processing (Tokyo, Japan). Currently, he is heading the Signal Processing & Information Theory Group at the Institute of Biophysics at the University of Regensburg and is working on his habilitation. He serves as an Associate Editor of "Computational Intelligence and Neuroscience," and is a Member of IEEE, EURASIP, and ENNS. His research interests include statistical signal processing, machine learning, blind source separation, and biomedical data analysis.

Pando Georgiev received his M.S., Ph.D., and "Doctor of Mathematical Sciences" degrees in mathematics (operations research) from Sofia University "St. Kl. Ohridski," Bulgaria, in 1982, 1987, and 2001, respectively. He has been with the Department of Probability, Operations Research, and Statistics at the Faculty of Mathematics and Informatics, Sofia University "St. Kl. Ohridski," Bulgaria, as an Assistant Professor (1989–1994), and since 1994, as an Associate Professor. He was a Visiting Professor at the University of Rome II, Italy (CNR grants, several one-month visits), the International Center for Theoretical Physics, Trieste, Italy (ICTP grant, six months), the University of Pau, France (NATO grant, three months), Hirosaki University, Japan (JSPS grant, nine months), and so forth. He has been working for four years (2000–2004) as a research scientist at the Laboratory for Advanced Brain Signal Processing, Brain Science Institute, the Institute of Physical and Chemical Research (RIKEN), Wako, Japan. After that and currently he is a Visiting Scholar in ECECS Department, University of Cincinnati, USA. His interests include machine learning and computational intelligence, independent and sparse component analysis, blind signal separation, statistics and inverse problems, signal and image processing, optimization, and variational analysis. He is a Member of AMS, IEEE, and UBM.

Andrzej Cichocki was born in Poland. He received the M.S. (with honors), Ph.D., and Habilitate Doctorate (Dr.Sc.) degrees, all in electrical engineering, from the Warsaw University of Technology (Poland) in 1972, 1975, and 1982, respectively. He is the co-author of three international and successful books (two of them were translated to Chinese): *Adaptive Blind Signal and Image Processing* (John Wiley, 2002) *MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems* (Springer, 1989), and *Neural Networks for Optimization and Signal Processing* (J. Wiley and Teubner Verlag, 1993/1994) and the author or coauthor of more than three hundred papers. He is the Editor-in-Chief of the Journal Computational Intelligence and Neuroscience and an Associate Editor of IEEE Transactions on Neural Networks. Since 1997, he has been the Head of the Laboratory for Advanced Brain Signal Processing in the Riken Brain Science Institute, Japan.