**Noname manuscript No.**
(will be inserted by the editor)

# A Geometric Blind Source Separation Method Based on Facet Component Analysis

**Penghang Yin · Yuanchang Sun · Jack Xin**

Received: date / Accepted: date

**Abstract** Given a set of mixtures, blind source separation attempts to retrieve the source signals without or with very little information of the mixing process. We present a geometric approach for blind separation of nonnegative linear mixtures termed *facet component analysis* (FCA). The approach is based on facet identification of the underlying cone structure of the data. Earlier works focus on recovering the cone by locating its vertices (vertex component analysis or VCA) based on a mutual sparsity condition which requires each source signal to possess a stand-alone peak in its spectrum. We formulate alternative conditions so that enough data points fall on the facets of a cone instead of accumulating around the vertices. To find a regime of unique solvability, we make use of both geometric and density properties of the data points, and develop an efficient facet identification method by combining data classification and linear regression. For noisy data, we show that denoising methods may be employed, such as the total variation technique in image processing. We show computational results on nuclear magnetic resonance spectroscopic data to substantiate our method.

**Keywords** Blind source separation · Facet component analysis · Nonnegative matrix factorization · Total variation

Penghang Yin · Jack Xin
Department of Mathematics, University of California at Irvine, Irvine, CA 92697, USA.
E-mail: penghany@uci.edu; jxin@math.uci.edu

Yuanchang Sun
Department of Mathematics and Statistics, Florida International University, Miami, FL 33199, USA.
E-mail: yuasun@fiu.edu

## 1 Introduction

Blind source separation (BSS) is a major area of research in signal and image processing [7]. It aims at recovering source signals from their mixtures with minimal knowledge of the mixing environment. The applications of BSS range from engineering to neuroscience. A recent emerging research direction of BSS is to identify chemical explosives and biological agents from their spectral sensing mixtures recorded by various spectroscopy such as Nuclear Magnetic Resonance (NMR), Raman spectroscopy, Ion-mobility spectroscopy (IMS), and differential optical absorption spectroscopy (DOAS), etc. Spectral sensing is a critical area in national security and a vibrant scientific area. The advances of modern imaging and spectroscopic technology have made it possible to classify pure chemicals by their spectral features. However, mixtures of chemicals subject to changing background and environmental noise pose additional challenges. The goal of this paper is to develop a BSS method to process data in the presence of noise based on geometric spectral properties.

To separate the spectral mixtures, one needs to solve the following matrix decomposition problem

$$X = A\,S + N\,, \tag{1.1}$$

where $A \in \mathbb{R}^{m\times n}$ is a full rank unknown basis (dictionary) matrix or the so called mixing matrix in some applications, $N \in \mathbb{R}^{m\times p}$ is an unknown noise matrix, $S = [s(1), \ldots, s(p)] \in \mathbb{R}^{n\times p}$ is the unknown source matrix containing signal spectra in its rows. Here $p$ is the number of data samples, $m$ is the number of observations, and $n$ is the number of sources. Various BSS methods have been proposed based on *a priori* knowledge of source signals such as statistical independence, sparseness, nonnegativity, [3,6,7,12,13,16,17,18,21,24,

25] among others. As a matrix factorization problem in the noise free case ($N = 0$), BSS has permutation and scaling ambiguities in its solutions similar to factorizing a large number into product of primes. For any permutation matrix $P$ and invertible diagonal matrix $\Lambda$, we have

$$X = A\,S = (A\,P\Lambda)(\Lambda^{-1}P^{-1}S). \tag{1.2}$$

then the solutions $(AP\Lambda,\ \Lambda^{-1}P^{-1}S)$ and $(A, S)$ are considered to be equivalent in the sense of BSS.

Recently there has been active research on BSS by exploiting data geometry [1,2,5,14,21,22,23,24,25,26, 27,28]. For simplicity, let $N = 0$ in (1.1). The geometric observation [1,21,28] is that if each row of $S$ has a dominant peak at some location (column number) where other rows have zero elements, then the problem of finding the columns of the mixing matrix $A$ reduces to the identification of the edges of a minimal cone containing the columns of mixture matrix $X$. In hyperspectral imaging (HSI), the condition is known as pixel purity assumption (PPA,[5]). In other words, each pure material of interest exists by itself somewhere on the ground. The PPA based convex cone method (known as N-findr [28]) is now a benchmark in HSI, see [5,21,22, 23,24] for its more recent variants. The method termed *vertex component analysis* (VCA) proposed in [23] is worth mentioning here being a fast unmixing algorithm for hyperspectral data. In Nuclear Magnetic Resonance (NMR) spectroscopy which motivates our work here, PPA was reformulated by Naanaa and Nuzillard (NN) in [21]. The source signals are only required to be non-overlapping at some locations of acquisition variable (e.g. frequency) which leads to a dramatic mathematical simplification of a general non-negative matrix factorization problem (1.1) which is non-convex [15]. In simple terms, PPA on NMR source signals is that each source has a stand alone peak at some location of the acquisition variable where the other sources are identically zero. More precisely, the source matrix $S \geq 0$ is assumed to satisfy the following

**Assumption** (PPA). *: For each $i \in \{1, 2, \ldots, n\}$ there exists an $j_i \in \{1, 2, \ldots, p\}$ such that $s_{i,j_i} > 0$ and $s_{k,j_i} = 0$ $(k = 1, \ldots, i-1, i+1, \ldots, n)$, where $s_{i,j_i}$ (resp., $s_{k,j_i}$) is the $(i, j_i)$-th (resp., $(k, j_i)$-th) entry of $S$.*

Rewriting the equation (1.1) in terms of columns, we obtain

$$X^j = \sum_{k=1}^{n} s_{k,j} A^k \quad j = 1, \ldots, p\,, \tag{1.3}$$

where $X^j$ denote the $j$th column of $X$, and $A^k$ the $k$th column of $A$. Assumption NNA implies that $X^{j_i} = s_{i,j_i}A^i \quad i = 1, \ldots, n$ or $A^i = \frac{1}{s_{i,j_i}} X^{j_i}$. Hence Eq. (1.3) is rewritten as

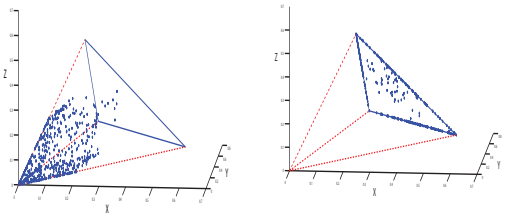

**Fig. 1** A cloud of data points (columns of $X$), left plot, is rescaled to lie on a plane determined by three vertices of the cone (right plot).

$$X^j = \sum_{i=1}^{n} \frac{s_{i,j}}{s_{i,j_i}} X^{j_i}\,, \tag{1.4}$$

which says that every column of $X$ is a nonnegative linear combination of the columns of $A$.

Simply put, the stand-alone peaks possessed by each source allow formation of a convex cone enclosing all the columns of $X$, and the edges (or vertices) of the cone are the columns of the mixing matrix. To illustrate the idea, let us consider an example of three sources and three mixtures ($A \in \mathbb{R}^{3\times 3}, X \in \mathbb{R}^{3\times p}, S \in \mathbb{R}^{3\times p}$, and $A$ is non-singular).

$$\begin{aligned}&\left[X^1, X^2, \cdots, X^p\right]\\ &= \left[A^1, A^2, A^3\right] \cdot \begin{pmatrix} * \cdots * & 1 & 0 & 0 & * \cdots * \\ * \cdots * & 0 & 1 & 0 & * \cdots * \\ * \cdots * & 0 & 0 & 1 & * \cdots * \end{pmatrix}\,,\end{aligned}$$

where $X^k$(resp., $A^k$) represents the $k$-th column of $X$ (resp., $A$), 1 represents nonzero entry, and the three $1's$ are three stand-alone peaks. It can be seen that columns of $A$ are actually among those of $X$ up to constants, and other columns of $X$ are nonnegative linear combinations of columns of $A$. Geometrically, the columns of $A$ span a convex cone enclosing all columns of $X$. The estimation of $A$ is equivalent to the identification of this cone. As a matter of fact, the vertices of the cone are the columns of $A$. Fig. 1 shows this vertex-represented cone containing all data points. To find the vertices (or edges) of the cone, the following optimization problem is solved for each column $k$ of $X$:

$$\begin{aligned} c \;=\; &\min \sum_{j=1, j\neq k}^{p} \lambda_j \\ &\text{s.t.} \quad \sum_{j=1, j\neq k}^{p} X^j \lambda_j = X^k \quad, \lambda_j \geq 0. \end{aligned}$$

It is shown [9] that $X^k$ *is an edge of the convex cone if and only if the optimal objective function value* $c^*$ *is greater than 1.*

If the data are contaminated by noise, the following optimization problem is solved for each $k$ [21]:

$$\min \text{ score} = \min_{\lambda_j \geq 0} \| \sum_{j=1, j\neq k}^{p} X^j \lambda_j - X^k \|_2^2. \tag{1.5}$$

A score is associated with each column of $X$. Columns with low scores are unlikely to be a column of $A$ because this column is roughly a nonnegative linear combination of the other columns of $X$. On the other hand, a high score means that the corresponding column is far from being a nonnegative linear combination of other columns. The $n$ rows from $X$ with highest scores are selected to form $A$, the mixing matrix.

Though the vertex based convex cone method above is geometrically elegant, the working condition is still restrictive. The success of the cone method highly depends on the recognition of its vertices. If PPA is violated, the vertices are not in the data matrix $X$, and may not be the primary objects for identification. In this paper, we consider a more general scenario where data points (scaled columns of $X$) lie either inside or on the facets(possibly located at the vertices) of a convex cone, see Fig. 1 for an example. The cone structure will be reconstructed from its facets instead of the vertices. A facet component analysis (FCA) should be pursued for the data considered. Roughly speaking, the source condition for this case is that *a number of columns of the source matrix* $S$ *possess at least one zero entry.* We shall pursue a more precise study on the source condition for solvability in the next section. The problem above calls for an identification method of flat submanifolds from a point cloud, as the facets of a convex cone in general lie on hyperplanes. The vertices are obtained from the intersections of the hyperplanes expanded from the facets, thereby recovering all the columns of the mixing matrix $A$. We shall study the identification of these flat manifolds and the subsequent recovery of the source signals. The extraction of meaningful geometric data structures help the data matrix factorization and reduce the computational cost.

Recently, a dual cone based approach to BSS problem was proposed in [22]. The method first calculates the dual of the data cone, then selects a subset of the source signals from a set of source candidates by means of a greedy process. Specifically, the first step consists in computing a generating matrix of the dual of the data cone by the double description method [20]. The second step consists in extracting an estimate of the source matrix from a larger matrix which is the product of the generating matrix with its transpose. Multiple solutions can be obtained in the second step, so the author imposed an additional orthogonality constraint on the source signals to get a sub-optimal solution. Overall, the method proposed in [22] appears indirect and computationally unwieldy, besides requiring the orthogonality of source signals. Here we opt to solve the problem directly by identifying the facets of the data cone under a unique solvability condition.

The rest of the paper is organized as follows. In Section 2, we propose a new source condition for solving (1.2) based on the geometric structure of the data points, and develop an algorithm for facet identification and reconstruction of the associated convex cone. We also discuss template assisted FCA solutions in the regime of non-uniqueness. In Section 3, we present computational examples and results. For heavily noisy data, we show that a denoising method may be needed to help remove or reduce the noise. A denoising method based on total variation and distance function is discussed. Concluding remarks and future work are in Section 4.

## 2 Proposed Method

We point out that FCA is not only applicable to all the data separations for which VCA works, as will be seen later, but also designed for tasks where the sources are sparse with moderate overlaps amongst their peaks.

### 2.1 Facet Component Analysis

In order to illustrate the basic idea behind FCA, we first consider the noiseless case:

$$X = A\,S$$

We are concerned with non-negative matrix factorization (NMF) problem. We assume that $X, S \in \mathbb{R}^{m\times p}$ (i.e. the numbers of available mixtures and source signals are equal) where $p \gg m$, and that the mixing matrix $A \in \mathbb{R}^{m\times m}$ is of full rank. Since $A$ is a square matrix, what we consider here is just the determined case. However, the proposed method can be easily extended to the over-determined case.

The first source condition on the problem is as follows:

**Assumption 1.** *For each* $i = 1, ..., m$*, at least* $m-1$ *linear independent columns of the source matrix* $S$ *have a zero entry in the* $i$*-th element.*

Under Assumption 1, there are at least $m-1$ data points lie on each facet of the $m$-dimensional convex cone that encloses the columns of $X$. As we notice that

each facet is corresponding to a $(m-1)$-dimensional hyperplane, linear independence of the data points makes each of the facets identifiable. The columns of the mixing matrix $A$ are obtained from the intersections of the hyperplanes expanded from the facets. To state this more precisely, we first start with the definition of convex cones: let $M \in \mathbb{R}^{m\times p}$, the subset of $\mathbb{R}^m$ defined by

$$\mathcal{M} = \mathrm{Cone}(M) := \{M\,\alpha \in \mathbb{R}^m \,|\alpha \in \mathbb{R}^p \geq 0\}$$

is a convex cone, and $M$ is said to be a generating matrix of $\mathcal{M}$ since every element of $\mathcal{M}$ is a nonnegative linear combination of the columns of $M$. Let $\mathcal{X} = \mathrm{Cone}(X)$ and $\mathcal{A} = \mathrm{Cone}(A)$, then we have the following theorem (or combining Lemma 3 and Lemma 5 in [22]):

**Theorem 1** *If $X = A\,S$, and $A$, $S \geq 0$, then $\mathcal{X} \subseteq \mathcal{A}$. Moreover, each facet of $\mathcal{A}$ contains a facet of $\mathcal{X}$.*

For readers' convenience, a short proof is given below.

*Proof.* $\forall\, x \in \mathcal{X}$, let $x = X\,\alpha, \alpha \geq 0$. So $x = A\,S\,\alpha = A\,(S\,\alpha)$, where $S\,\alpha \geq 0$. So clearly $x \in \mathcal{A}$.

The second claim follows as we notice the facts: (1) $\mathcal{A}$ has $m$ facets and each one is spanned by $m-1$ column vectors of $A$; (2) Using Assumption 1, $X = A\,S$ has at least $m-1$ linearly independent column vectors located in each facet of $\mathcal{A}$; (3) $\mathcal{X} \subseteq \mathcal{A}$. □

Since $A$ is nonsingular, $\mathcal{A}$ has $m$ edges, and thus has $\binom{m}{m-1} = m$ facets. Based on Theorem 1, our method aims to identify the $m$ facets of $\mathcal{X}$ contained in the facets of $\mathcal{A}$. If we project column vectors of $X$ onto the hyperplane $\mathbf{x}^{\mathrm{T}} \cdot \mathbf{1} = 1$, where $\mathbf{x} = (x_1, \ldots, x_m)^{\mathrm{T}}$ and $\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}}$, the resulting data points together with the origin $\mathbf{0} = (0, \ldots, 0)^{\mathrm{T}}$ form a $m$-dimensional convex hull denoted by Convhull$(X)$. In fact, Convhull$(X)$ results from the cone $\mathcal{X}$ being truncated by the hyperplane $\mathbf{x}^{\mathrm{T}} \cdot \mathbf{1} = 1$. We then acquire all the facets and the associated vertices of Convhull$(X)$, which can be done by means of the MATLAB function **convhulln**. It is an implementation of the *Quickhull algorithm* computing the convex hull of a point cloud.

By Theorem 1, $\mathcal{X}$ has at least the same number of facets as $\mathcal{A}$. In fact, it is very likely that $\mathcal{X}$ has more facets than $\mathcal{A}$, thus making the selections of $A$ clearly nonunique. For example, in Fig. 2, the columns of $X$ are enclosed in a hexagonal cone $\mathcal{X}$ with six facets, if we select every other facet from the six facets of $\mathcal{X}$, we can form two cones and both could be choices of Convhull$(A)$. Therefore we need an additional source assumption to provide a selection criterion:

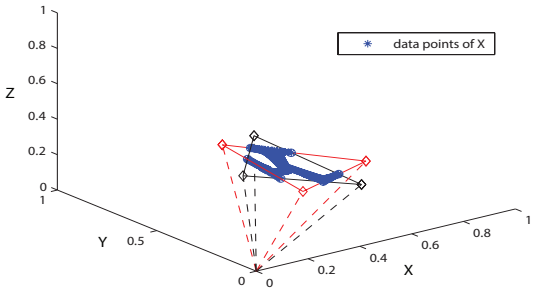


**Fig. 2** Non-uniqueness of cone $\mathcal{A}$. The data points of $X$ (in blue), two convex hulls Convhull$(A)$ (black and red) are formed by selecting three facets of cone $\mathcal{X}$.

**Assumption 2.** *The $m$ facets of $\mathcal{X}$ containing the largest numbers of data points are contained in the $m$ facets of $\mathcal{A}$.*

**Remark 1.** *The intuition behind Assumption 2 can be that it is essentially a geometric interpretation of the source matrix $S$ being sparse. To see this, let $X^j$ (nonzero) be the $j$-th data point of $X$, we have*

$$X^j = \sum_{i=1}^{m} S_{ij} A^i.$$

*If $S_{ij} = 0$ for some index $i$, then $X^j$ is a nonnegative linear combination of all columns of $A$ but $A^i$. Geometrically speaking, the data point $X^j$ is located in the facet of $\mathcal{A}$ corresponding to the edge (vertex) $A^i$. On the other hand, if $S_{ij} \neq 0$ for all $i$, then $X^j$ is an interior point of $\mathcal{A}$, and vice versa. Assumption 2 basically says that we want as many data points as possible to be contained in the facets of $\mathcal{A}$, which is equivalent to looking for the sparsest $S$ via NMF. In Fig. 2, the red Convhull$(A)$ should be selected according to Assumption 2 as one can see that there are many data points lie in its facets. The corresponding recovered sources is shown in the top row of Fig. 3, while the bottom row of Fig. 3 are the recovered sources if we selected the black Convhull$(A)$ in Fig. 2. This example is an illustration that selecting mixing matrix $A$ under Assumption 2 finally yields sparse sources $S$.*

**Remark 2.** *We note in passing that PPA (stand-alone peak assumption) is a special (much more restrictive) case of Assumption 1. Moreover, given that PPA is true, Assumption 2 is also satisfied since $\mathcal{X}$ and $\mathcal{A}$ would have the same number of facets. These observations actually validate our claim at the very beginning of this section.*

By this assumption, we count and sort the number of data points in each facet of Convhull$(X)$ followed by

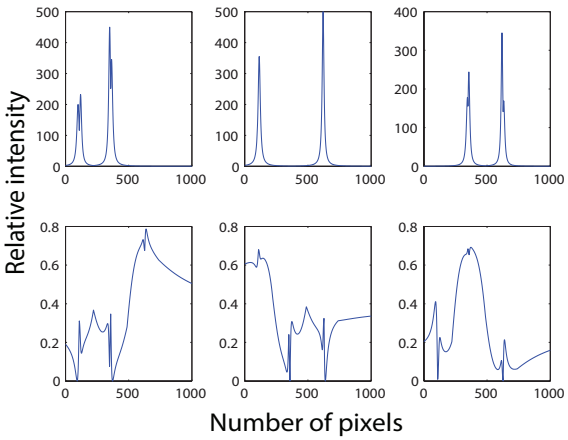


**Fig. 3** Top row: sources recovered by selecting $\mathcal{A}$ according to Assumption 2 (red convex hull in Fig. 2). Bottom row: sources recovered by selecting black convex hull in Fig. 2.

selecting the $m$ facets with the largest numbers of data points. Each one is contained in some facet of $\mathcal{A}$. So the intersection of any $m-1$ facets out of the obtained $m$ facets is an edge of $\mathcal{A}$. By intersecting all $m$ edges with the hyperplane $\mathbf{x}^{\mathrm{T}} \cdot \mathbf{1} = 1$, we obtain the $m$ column vectors of the mixing matrix $A$. In the last step, we use nonnegative least squares method to recover the columns of the source matrix $S$.

Based on the ideas above, we now include the additive noise and put forward the algorithm under Assumptions 1-2 as follows:

**Algorithm 1** (Facet Component Analysis)**.** $(A, S) = FCA(X, \rho, \epsilon, \sigma, \delta)$*; parameters* $\rho > 0$*;* $\epsilon, \sigma, \delta \in (0, 1)$*.*

1. *(Preprocessing) Set the negative entries in* $X$ *to 0, then discard those column vectors with norm less than* $\rho$*. Let us still denote by* $X$ *the resulting matrix. Project column vectors of* $X$ *onto the hyperplane* $\boldsymbol{x}^{\mathrm{T}} \cdot \boldsymbol{1} = 1$*.*
2. *(Convex hull) Add the origin* $\boldsymbol{0}$ *as the first column to* $X$*. Find all the facets and vertices of Convhull*$(X)$ *using the MATLAB function* `convhulln`*, keep only the facets with the vertex* $\boldsymbol{0}$*. Denote by* $F_i$ *the* $i$*th facet and* $V_i$ *the set of its vertices.*
3. *(Grouping) Initialize a group* $G_i = V_i$*. For the* $j$*-th column* $X^j \notin G_i$*, if the distances* $D(X^j, F_i) < \epsilon$ *and* $D(X^j, V_i) > \sigma$*, then add* $X^j$ *to* $G_i$*.*
4. *(Plane fitting) For each* $G_i$*, obtain the fitting plane denoted as* $\boldsymbol{x}^{\mathrm{T}} \cdot \boldsymbol{b}_i = 0$*, where* $\boldsymbol{b}_i$ *is the normal vector with length* 1*. Select* $m$ *planes from those of* $G_i$ *with the largest cardinalities such that* $\boldsymbol{b}_i^{\mathrm{T}} \cdot \boldsymbol{b}_j < \delta \approx 1$*, where* $i \neq j$*.*
5. *(Intersecting) Obtain the* $m$ *intersections of any* $m-1$ *planes out of* $m$ *planes from step 4 and the hyperplane* $\boldsymbol{x}^{\mathrm{T}} \cdot \boldsymbol{1} = 1$ *to form* $A$*.*
6. *(Source recovering) For each* $X^j$*, solve the nonnegative least squares problem to find the corresponding column* $S^j$ *in* $S$ *as:*

$$\min \|X^j - A\,S^j\|_2^2 \quad \text{s.t.} \quad S^j \geq 0.$$

**Remark 3.** *The choices of the four parameters are basically heuristic, so we provide the following rules of thumb for selection:*

1. *The values of* $\epsilon$ *and* $\sigma$ *should rely on the level of noise. Normally we can set them equal. A higher level of noise demands larger* $\epsilon$ *and* $\sigma$*. For instances, in noise-free cases, they can be as small as* $10^{-6}$*; when noise level is SNR* $\approx$ *30dB,* $\epsilon$ *and* $\sigma$ *should be on the order of* $10^{-2}$*; if high level noise are present, they increase up to the order of 0.1.*
2. *The value of* $\rho$ *is also positively correlated to the amount of noise. We suggest* $\rho$ *not exceed* $\frac{1}{2}\|X\|_1$*, where* $\|X\|_1$ *is maximum absolute column sum of the mixture matrix.*
3. *In step 4, we need the parameter* $\delta$ *to guarantee the selected facets of Convhull*$(X)$ *to be sufficiently distinct. We can just fix* $\delta = 0.99$*, which works well in all of our experiments.*

## 2.2 Template assisted FCA

Assumption 2 provides a selection criterion based on the density property of the data points, under this assumption $m$ facets of $\mathcal{X}$ with largest numbers of data points are selected as the facets of $\mathcal{A}$. This source assumption proves to be a suitable condition for many data sets we have tested. Here we recognize another scenario of non-uniqueness: among the $n$ facets $\mathcal{X}$, $p$ of them contain the same number of data points and $p > m$, in this case Assumption 2 would fail to guarantee the unique selection of cone $\mathcal{A}$. Technically there are $\binom{p}{m} = \frac{p!}{m!(p-m)!}$ possible choices for $\mathcal{A}$. Although the nonnegativity may rule out some choices of the convex cones when their vertices fall outside the positive sector of the space, we are still left with a number of solutions especially when $p$ is large. In Fig. 4, cone $\mathcal{X}$ contains six facets of same number of data points, ideally selection of any three facets would yield a cone $\mathcal{A}$ ( there are $\binom{6}{3} = 20$ possibilities ). To overcome this non-uniqueness issue or reduce the number of possible solutions, a template of source signals may be used. Suppose that the $k$ ($k \leq m$) sources being recovered have spectral template in a database of size $N$. Let $T \in \mathbf{R}^{N \times p}$ ($N \gg m$ ) be the database matrix. For each

possible cone $\mathcal{A}$, we compare the recovered source matrix $S$ against the database $T$ by solving the equation below

$$S = C\,T\,, \tag{2.1}$$

where $C$ serves as an indicator matrix and is sparse. The nonzero entries of $C_j$ ($j$-th row of $C$) imply that $S_j$ ($j$-th row of $S$) is a linear combination of the corresponding spectral references from $T$. The sparsity of $C$ suggests that we solve the following $\ell_1$ optimization,

$$\min_{C_j \geq 0} \mu\|C_j\|_1 + \frac{1}{2}\|S_j - C_j\,T\|_2^2\,, \tag{2.2}$$

This problem could be solved by the Linearized Bregman method [11] described below. We introduce $u = (C_j)^{\mathrm{T}}, f = (S_j)^{\mathrm{T}}, B = T^{\mathrm{T}}$ and rewrite (2.2) as

$$\min_{u \geq 0} \mu\|u\|_1 + \frac{1}{2}\|f - B\,u\|_2^2\,. \tag{2.3}$$

When there is minimal measurement error, one must assign a tiny value to $\mu$ to heavily weigh the fidelity term $\|f - B\,u\|_2^2$ in order for $B\,u = f$ to be nearly satisfied. The linearized Bregman method can be written iteratively by introducing an auxiliary variable $v^j$:

$$\begin{cases} v^{j+1} = v^j - B^{\mathrm{T}}(B\,u^j - f), \\ u^{j+1} = \delta \cdot shrink_+(v^{j+1}, \mu), \end{cases} \tag{2.4}$$

where $u^0 = v^0 = \mathbf{0}$, $\delta > 0$ is the step size, and $shrink_+$ is for computing nonnegative solutions,

$$shrink_+(v, \mu) = \begin{cases} v - \mu, & \text{if } v > \mu\,, \\ 0, & \text{if } v < \mu\,. \end{cases} \tag{2.5}$$

**Remark 4.** *1. If $k = m$, all the source signals being recovered have spectral templates in the database. Solving Eq .(2.1) for all possible recovered source signals $S$ from FCA, we obtain sparse indicator matrices. The desired source matrix $S$ is the one corresponding to a matrix $C$ of row sparsity one (only one nonzero entry). In this case the non-uniqueness issue can be fully resolved.*

*2. If $k < m$, not all source signals being recovered have spectral templates in the database. Then FCA with template would not be able to pinpoint the true solutions. However, it can reduce the number of possibilities. In fact, the method can identify those recovered source matrices containing these $k$ signals, and as a result the searching for true solution will be narrowed down to a much smaller number of solutions which will be handed to practitioners for further analysis with their knowledge and experience.*

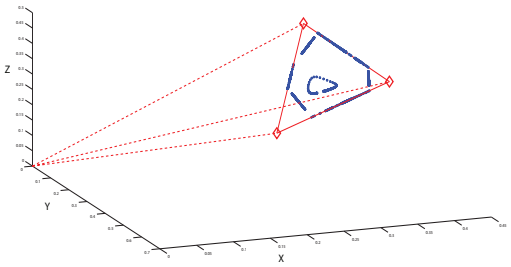


**Fig. 4** Scatter plot of columns of $X$ where the data points are rescaled to lie on the plane $x + y + z = 1$. Six facets of cone $\mathcal{X}$ contain same number of data points. Red cone is the true $A$ used to generate data.

## 3 Numerical Experiments

We report the numerical results of our algorithm in this section. The data we tested includes real-world NMR spectra as well as synthetic mixtures.

As aforementioned, our algorithm works for the separations of data where the source signals have standalone peaks. For the data, we used true NMR spectra of compounds $\beta$ -cyclodextrine, $\beta$-sitosterol, and menthol as source signals. The NMR spectrum of a chemical compound is produced by the Fourier transformation of a time-domain signal which is a sum of sine functions with exponentially decaying envelopes [10]. The real part of the spectrum can be presented as the sum of symmetrical, positive valued, Lorentzian-shaped peaks. The NMR reference spectra of $\beta$-cyclodextrine, $\beta$-sitosterol, and menthol are shown in the top panel of Fig. 5 from left to right. The mixtures were obtained by adding white Gaussian noise with SNR = 30 dB. For the parameters, we set $\rho = 5\times10^{-2}, \epsilon = \sigma = 10^{-2}, \delta = 0.99$ and list the recovery result below. $A_1$ is the column-wise rescaled true mixing matrix, while $\hat{A}_1$ is the computed mixing matrix via our method. By comparing the two matrices, we can see that the recovery is almost perfect.

$$A_1 = \begin{pmatrix} 0.4000 & 0.2778 & 0.4118 \\ 0.2667 & 0.2778 & 0.1765 \\ 0.3333 & 0.4444 & 0.4118 \end{pmatrix}$$

$$\hat{A}_1 = \begin{pmatrix} 0.4007 & 0.2788 & 0.4109 \\ 0.2667 & 0.2772 & 0.1778 \\ 0.3326 & 0.4440 & 0.4113 \end{pmatrix}$$

In the second example, we used three Lorentzian-shaped synthetic sources, where PPA is violated. We created three noisy mixtures $X \in \mathbb{R}^{3\times2000}$ by adding white Gaussian noise with SNR = 30 dB. The geometry of data points and recovery results are shown in Fig. 6 and Fig. 7. True mixing matrix $A_2$ and computed

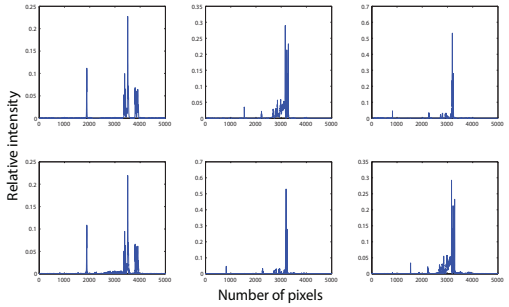


**Fig. 5** Top row: from left to right, the three reference spectra of $\beta$-cyclodextrine, $\beta$-sitosterol, and menthol in Example 1. Bottom row: recovery results by our method.

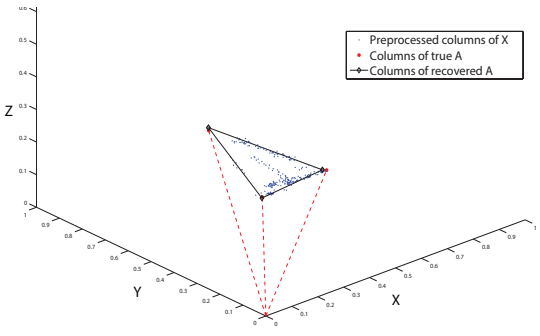


**Fig. 6** Scatter plot of rescaled columns of mixture matrix $X$(blue dots), true mixing matrix $A_2$(red asterisks) and recovered mixing matrix $\hat{A}_2$(black diamond) in Example 2.

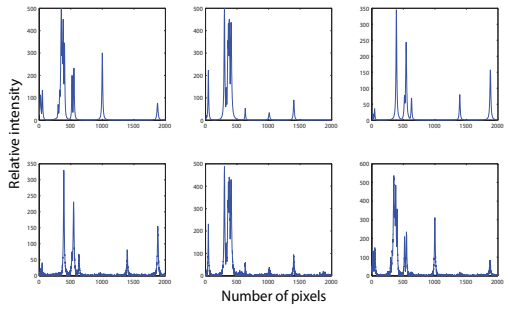


**Fig. 7** Top row: true sources in Example 2. Bottom row: computed sources by our method.

$\hat{A}_2$ are listed below. The result is achieved by setting $\rho = 80, \epsilon = \sigma = 2 \times 10^{-2}, \delta = 0.99$. The test was run on a laptop with 4GB RAM and 2.67GHz Intel i5 CPU, and the consuming time was 0.762s.

$$A_2 = \begin{pmatrix} 0.3000 & 0.5000 & 0.4545 \\ 0.5000 & 0.4286 & 0.1818 \\ 0.2000 & 0.0714 & 0.3636 \end{pmatrix}$$

$$\hat{A}_2 = \begin{pmatrix} 0.3083 & 0.5030 & 0.4483 \\ 0.4911 & 0.4298 & 0.1751 \\ 0.2005 & 0.0672 & 0.3766 \end{pmatrix}$$

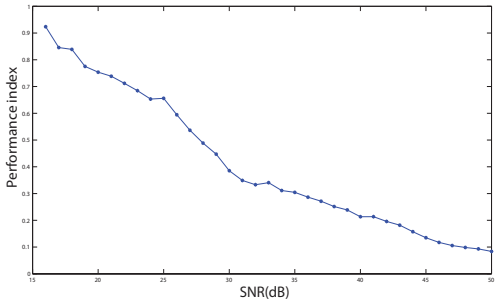


**Fig. 8** Test the performance of our method in the presence of noise.

In [8] the Comon's index was introduced to test the performance of source separations. The index is defined as follows: let $A$ and $\hat{A}$ be two nonsingular matrices with $\ell_2$-normalized columns. Then the distance between $A$ and $\hat{A}$ denoted by $\epsilon(A, \hat{A})$ which reads

$$\epsilon(A,\hat{A}) = \sum_i \left| \sum_j |d_{ij}| - 1 \right|^2 + \sum_j \left| \sum_i |d_{ij}| - 1 \right|^2 + \sum_i \left| \sum_j |d_{ij}|^2 - 1 \right| + \sum_j \left| \sum_i |d_{ij}|^2 - 1 \right|,$$

where $D = A^{-1}\hat{A}$, and $d_{ij}$ is the entry of $D$. In [8] Comon proved that $A$ and $\hat{A}$ are considered nearly equivalent in the sense of BSS (i.e., $\hat{A} = A\,P\,\Lambda$) if $\epsilon(A, \hat{A}) \approx 0$.

We then show the performance of FCA in the presence of different levels of noises. The three sources in Example 2 were combined to generate three mixtures corrupted by Gaussian noises with SNR varying from 16 dB to 50 dB. At each noise level, 50 independent trials were performed using Algorithm 1 and the parameters were tuned for each trial. The average of Comon's indices were then computed. As shown in Fig. 8, the performance indices stayed below 1 even in low SNR zone, which were satisfactory results.

In next example, we show the computational results of template assisted FCA. In the example, we are to separate three source signals from three mixtures. The geometry of the data points is shown in Fig. 4, where the Convhull($X$) is a hexagon. The data is synthetically generated so that each side contains the same number of data points. In Fig. 9, we show an example that cone $\mathcal{X}$ has 6 facets containing same number of data points. Technically there are $\binom{6}{3} = 20$ choices for cone $\mathcal{A}$. The cones that lie in the positive sector of the space will be kept for further analysis due to the nonnegativity of $A$. We are left with six cones showing in Fig. 9, and the three cones in the first column lie outside the data points and are not meaningful solutions. As

a result, we have three choices of cone $\mathcal{A}$ shown in the right panel of Fig. 9. The corresponding source recovery are in Fig. 10. Suppose the three source signals have spectral references from a database of size ten. For each three recovered sources from Fig. 10, we compare the recovered signals against the database by solving (2.2). The correct selection of the source signals should associate with a sparse indicator matrix $C$ of row sparsity one. We show the results of the coefficient matrix $C$ for the cone 4, 5, 6, respectively in the following,

$$C^4 = \begin{pmatrix} 0.001 & 0 & 0.02 & 0.015 & 0.106 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.03 & 0.01 & 0 & 0 & 1 & 0.59 & 0 & 0.001 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0.76 & 0 & 0.11 & 0 \end{pmatrix},$$

$$C^5 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix},$$

$$C^6 = \begin{pmatrix} 0.002 & 0 & 0.02 & 0.015 & 0.178 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Apparently $C^5$ is the sparsest and its row sparsity is one, hence the corresponding recovered source signals and cone will be selected as the desired solutions. The parameter used in the $\ell_1$ optimization $\mu = 0.03$.

If only one source signal has spectral reference from the database, the solutions of $C^4 - C^6$ are listed below

$$C^4 = \begin{pmatrix} 0.1 & 0.98 & 1 & 0.5 & 0.26 & 0 & 0.16 & 0.1 & 0 & 0 \\ 0.08 & 1 & 0.8 & 0.4 & 0.2 & 0 & 0 & 0.11 & 0.42 & 0.064 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0.93 & 0.11 \end{pmatrix},$$

$$C^5 = \begin{pmatrix} 0.097 & 1 & 0.98 & 0.48 & 0.22 & 0 & 0.12 & 0.11 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 & 0.37 & 0 & 0 & 0.03 & 1 & 0.13 \end{pmatrix},$$

$$C^6 = \begin{pmatrix} 0.1 & 0.98 & 1 & 0.5 & 0.3 & 0 & 0.17 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.295 & 0 & 0 & 0.37 & 0 & 0 & 0.028 & 1 & 0.13 \end{pmatrix}.$$

It is seen that both $C^5$ and $C^6$ contain a row of sparsity one, and this indicates the source recovery $S^5$ and $S^6$ contain a source which has a spectral reference from the database. Therefore we will eliminate cone 4 but keep cones 5 and 6 as two possible solutions.

Next, we present the comparison results of FCA with some existing BSS methods. The first comparison is between FCA and VCA (vertex component analysis). VCA focus on identifying the cone by locating its vertices based on a mutual sparsity condition which requires each source signal to possess a stand-alone peak

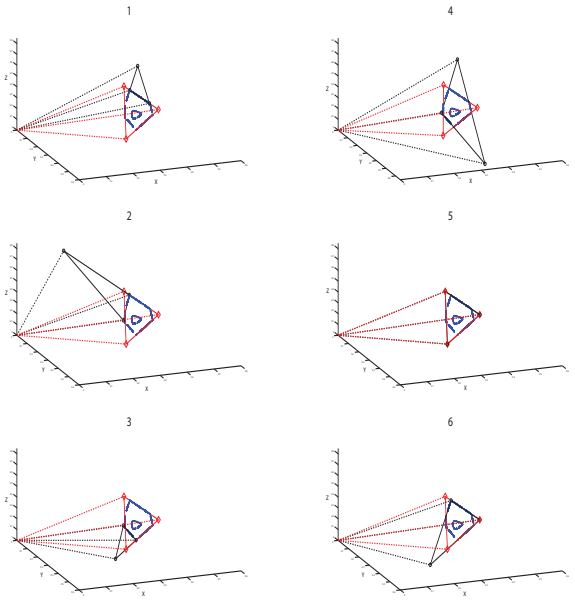


**Fig. 9** Non-uniqueness of cone $\mathcal{A}$ when Assumption 2 fails to be satisfied. Data points are in blue, the ground truth of convex hull Convhull($A$) is red; while the black color show the identified Convhull($A$) by our method. The black triangles (1,2,3) formed in the left panel lie outside the data points, and will be dropped. Then only reasonable solutions are the triangles in the right panel (4,5,6).

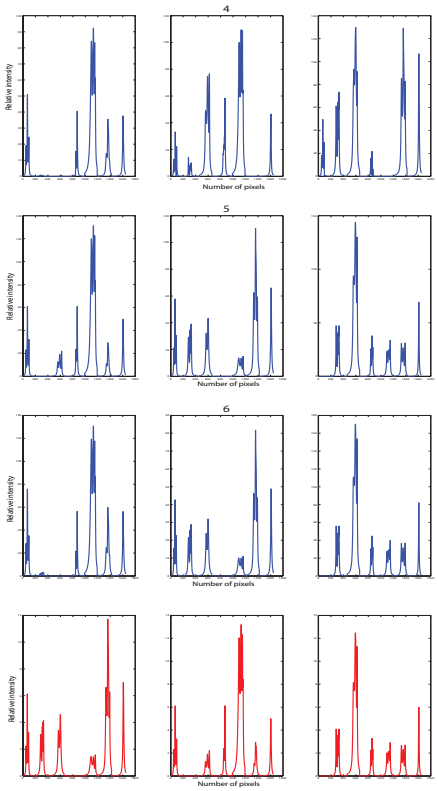


**Fig. 10** Rows one to three are the source signals recovered by selecting black convex hulls in the right panel in Fig. 9. Bottom row is the three true source signals.

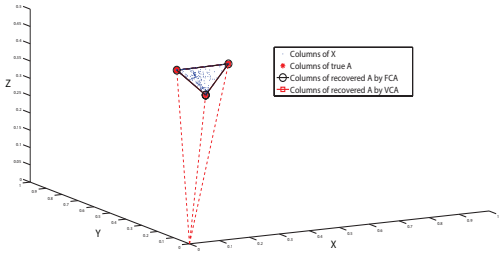


**Fig. 11** Comparison between FCA and VCA for data satisfying stand alone peak assumption.

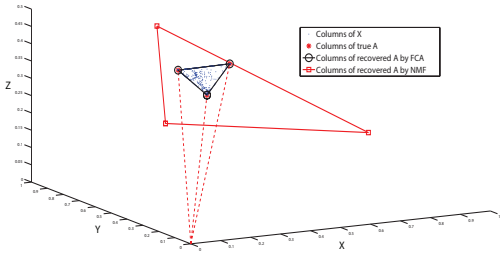


**Fig. 12** Comparison between FCA and NMF for data satisfying stand alone peak assumption.

(PPA) in its spectrum. Since PPA is a special case of Assumption 1, FCA works for the separation of data of PPA, and this has been shown in example one. Here we present in Fig. 11 the scatter plots along with the recovery results of mixing matrix $A$ via FCA and VCA respectively. It can be found that the results agree well in case of the stand alone peak condition which is the working assumption for VCA. Though both FCA and VCA are able to locate the vertexes of data cone, it should be pointed that FCA is more suitable for identifying cones from their facets while VCA from their vertices. In this sense, FCA provides an alternative to VCA. Computation is performed on a PC with 4G RAM and 3.0 GHz Intel Pentium CPU. The cpu time for FCA is 0.3750 seconds, and it is 0.6094 seconds for VCA (not including the cpu time of plotting the figures). FCA sees a cpu time saving 38% comparing to VCA, and this improvement is consistent across other data sets under the same assumption.

Next, we show the comparison between FCA and NMF. NMF decomposes the mixture $X$ into $A$ and $S$ by solving a non-convex optimization problem

$$\min_{A,S} \|X - A\,S\|_2^2 \quad \text{s.t.} \quad s_{ij} \geq 0,\ a_{ij} \geq 0\,. \tag{3.1}$$

Optimization approaches for solving (3.1) include gradient descent, alternating least squares, and so on. The result is depicted in Fig. 12, FCA is clearly better than NMF for the stand-alone peak source signals. But we should point that if no knowledge of the source spectra other than the positivity is available, the separation problem becomes a non-convex one, for which NMF should be used. It should be noted that NMF type methods are non-convex optimizations which usually converge to a critical point or a local minimum of the problem, hence most likely producing sub-optimal solutions, though it does provide a general method when not much information is known about the signals.

### 3.1 Denoising

If there is considerable noise in the data, it would be desirable to reduce or remove the noise before feeding them to the proposed method. We shall propose to apply image denoising techniques to the FCA for noise reduction or removal. These denoising methods may be used when the noise level is high, and they can be combined with the FCA after step 3. In this paper, we propose to apply the total variation idea of image denoising for noise removal. The method is based on the principle that signals with excessive noise have high total variation, that is, the integral of the absolute gradient of the signal is high. An advantage of total variation denoising is that it can remove the unwanted detail while preserves important details such as edges. The concept was originated in Rudin, Osher, and Fatemi in 1992 [29], and it can be applied to the point cloud noise removal. In the following, we shall use the example of a point cloud in $xyz$ plane to illustrate the idea of total variation denoising. We first preprocess the data by rescaling them onto a plane $x + y + z = 1$, the projected data are two dimensional. For each point $(x, y)$, a distance function to the point cloud (data points) is defined as

$$d(x, y) = \min_{x_i, y_i} \sqrt{(x - x_i)^2 + (y - y_i)^2}\,,$$

where $(x_i, y_i)^{\mathrm{T}}$ corresponds to the $i$-th column of $X$, note that $z_i$ is not included since the $z_i = 1 - x_i - y_i$. Fig. 13 shows an example of distance function to a unit circle, the left plot is the distance function with no noise, while the right plot is the function with noise. For computation, the distance function will be restricted on a rectangular region which contains the point cloud. Note that $d(x, y) \geq 0$, the distance function actually defines intensities of an image. Clearly, the set of $\mathcal{S} = \{(x, y) : d(x, y) = 0\}$ is the point cloud for the noiseless case.

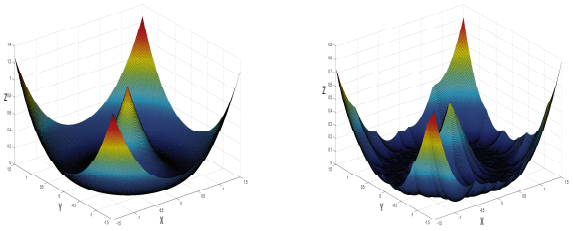

**Fig. 13** The distance function with (right) and without (left) noise.

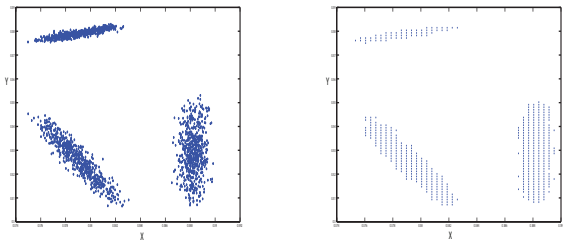

**Fig. 14** The point cloud before (left) and after (right) denoising. The points correspond to the columns of mixture matrix $X$, which are projected onto the plane $x + y + z = 1$.

We proceed to solve the Rudin-Osher-Fatemi (ROF) model to obtain a denoised distance function $u(x, y)$

$$\min_u \mathrm{TV}(u) + \lambda/2||d - u||_2^2,$$

where $TV(u)$ is the total variation of $u$ defined as

$$TV(u) = \sum_{i,j} \sqrt{|u_{i+1,j} - u_{i,j}|^2 + |u_{i,j+1} - u_{i,j}|^2}$$

for an image. A variation for ease of minimization is:

$$TV(u) = \sum_{i,j} |u_{i+i,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}| \ .$$

We shall use the recent *Chambolle's Algorithm* [4] to solve this minimization problem. Note that the smaller the parameter $\lambda$, the stronger the denoising. Then the zero level set of the resulting minimizer $u(x, y)$ will be taken as the denoised point cloud. In the real calculation, we will consider the following set with threshold $\mathcal{S} = \{(x, y) : 0 \leq u(x, y) \leq \tau\}$ where $\tau$ takes on a tiny value. The noisy point cloud and the result after the noise removal are depicted in Fig. 14. The detected planes from both of them are shown in Fig. 15, and their intersections, i.e., the vertexes of the cone. It can be noted that total variation denoising is very effective at preserving edges (thick lines in the figures) whilst smoothing away noise in flat regions. The idea of denoising distance function by total variation extends to point cloud of any dimension.

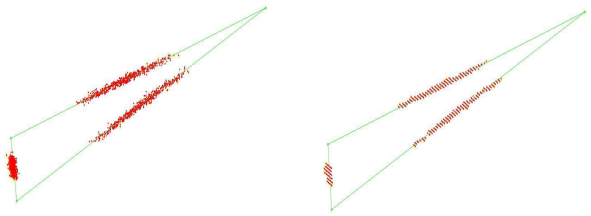

**Fig. 15** Computational results of planes from noisy data (left) and denoised data (right) shown in Fig. 14. The Green lines are the detected planes using the method proposed in the paper.

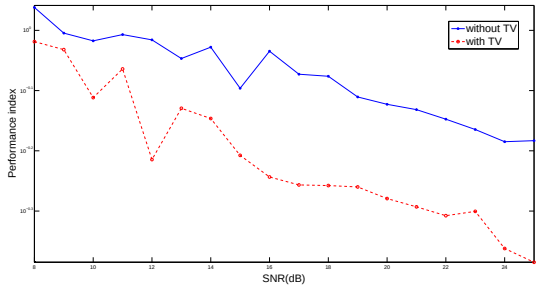


**Fig. 16** Comparison of Comon's indices (in log scale) of FCA with and without TV denosing at high noise levels.

We conducted experiments on performances of FCA with and without TV denoising. The mixtures were corrupted by Gaussian noises varying from 8 dB to 25 dB. A comparison of their performance is shown in Fig. 16. With TV denoising the Comon's indices are substantially lower than without it resulting in better separation performance.

## 4 Concluding Remarks

We developed a novel facet component analysis (FCA) for nonnegative blind source separation problems. We presented facet based solvability conditions for the unique solvability up to scaling and permutation by exploiting both the geometry of data matrix and the sparsity of the source signals. With the assistance from a template, the method proposed is able to remove or reduce the ambiguities of the solutions when the source signals fail to satisfy uniqueness. Numerical results on NMR signals validated the solvability conditions, and showed satisfactory performance of the proposed algorithm. For noisy data, total variation de-noising method serves as a viable preprocessing step. In terms of knowledge of the source signals, the FCA method works in a setting in between the vertex component analysis (under PPA)

and the nonnegative matrix factorization (under only non-negativity ).

A line of future work is to separate more source signals from their mixtures, known as an undetermined blind source separation, or uBSS. This problem presents more challenge than the determined or over-determined BSS in that the mixing matrix is non-invertible. Some recent study has been done by two of the authors based on a geometric approach to retrieve the mixing matrix under suitable solvability conditions [25].

## Acknowledgements

The work was partially supported by NSF-ATD grants DMS-0911277 and DMS-122507.

## References

1. J. Boardman, *Automated spectral unmixing of AVRIS data using convex geometry concepts*, in Summaries of the IV Annual JPL Airborne Geoscience Workshop, JPL Pub. 93-26, Vol. 1, 1993, pp 11-14.
2. J. Bobin, J-L. Starck, J. Fadili, and Y. Moudden, *Sparsity and Morphological Diversity in Blind Source Separation*, IEEE Trans. Image processing, Vol. 16 (2007), pp. 2662–2674.
3. P. Bofill and M. Zibulevsky, *Underdetermined blind source separation using sparse representations*, Signal Processing, 81 (2001), pp. 2353–2362.
4. A. Chambolle, *An Algorithm for Total Variation Minimization and Applications*, J of Math Imaging Vis, 20 (2004), pp. 89–97.
5. C-I Chang, ed., *Hyperspectral Data Exploitation: Theory and Applications*, Wiley-Interscience, 2007.
6. S. Choi, A. Cichocki, H. Park, and S. Lee, *Blind source separation and independent component analysis: A review*, Neural Inform. Process. Lett. Rev., 6 (2005), pp. 1–57.
7. A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, John Wiley and Sons, New York, 2005.
8. P. Comon, *Independent component analysis–a new concept?*, Signal Processing, 36 (1994), pp. 287–314.
9. J.H. Dulà and R.V. Helgason, *A new procedure for identifying the frame of the convex hull of a finite collection of points in multidimensional space*, European J. Oper. Res. 92 (1996), pp. 352–367.
10. R. Ernst, G. Bodenhausen, and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, 1987.
11. Z. Guo and S. Osher, *Template matching via $\ell_1$ minimization and its application to hyperspectral target detection*, Inverse Problem and Imaging, 5(2011), pp. 19–35.
12. P. Hoyer, *Non-negative Matrix Factorization with Sparseness Constraints*, Journal of Machine Learning Research, 5(2004), pp. 1457–1469.
13. A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley and Sons, New York, 2001.
14. B. Klingenberg., J. Curry and A. Dougherty, *Nonnegative matrix factorization:Ill-posedness and a geometric algorithm*, Pattern Recognition, 42 (2009), pp. 918–928.
15. D. D. Lee and H. S. Seung, *Learning of the parts of objects by non-negative matrix factorization*, Nature, 401 (1999), pp. 788–791.
16. J. Liu, J. Xin, Y-Y Qi, *A Dynamic Algorithm for Blind Separation of Convolutive Sound Mixtures*, Neurocomputing, 72(2008), pp. 521-532.
17. J. Liu, J. Xin, Y-Y Qi, *A Soft-Constrained Dynamic Iterative Method of Blind Source Separation*, SIAM J. Multiscale Modeling Simulations, 7(2009), pp. 1795-1810.
18. J. Liu, J. Xin, Y-Y Qi, F–G Zeng, *A Time Domain Algorithm for Blind Separation of Convolutive Sound Mixtures and $\ell_1$ Constrained Minimization of Cross Correlations*, Comm. Math Sci, Vol. 7, No. 1, 2009, pp. 109–128.
19. M. McDonnell, *Box-filtering Techniques*, Computer Graphics and Image Processing, 17 (1981), pp. 65–70.
20. T. Motzkin, H. Raiffa, G. Thompson, R. J. Thrall, *The Double Description Method*, Annals of Math Studies, Vol. 8, Princeton University Press, 1953, pp. 5173.
21. W. Naanaa and J.–M. Nuzillard, *Blind source separation of positive and partially correlated data*, Signal Processing, 85 (9) (2005), pp. 1711–1722.
22. W. Naanaa, *A Geometric Approach to Blind Separation of Nonnegative and Dependent Source Signals*, 18th European Signal Processing Conference (EUSIPCO-2010), Aalborg, Denmark, August 23-27, 2010, pp. 747–750
23. J.M.P. Nascimento and J.M. Bioucas-Diasm, *Vertex component analysis: a fast algorithm to unmix hyperspectral data*, IEEE Transactions on Geoscience and Remote Sensing, 43(4) (2005), pp. 898-910.
24. Y. Sun, C. Ridge, F. del Rio, A.J. Shaka and J. Xin, *Postprocessing and Sparse Blind Source Separation of Positive and Partially Overlapped Data*, Signal Processing, 91(8)(2011), pp. 1838-1851.
25. Y. Sun and J. Xin, *Under-determined Sparse Blind Source Separation of Nonnegative and Partially Overlapped Data*, SIAM J. Sci. Comput., 33 (4) (2011), pp. 2063-2094.
26. Y. Sun and J. Xin, *A Recursive Sparse Blind Source Separation Method and its Application to Correlated Data in NMR Spectroscopy of Bio-fluids*, J. of Sci Comput., 51 (2012), pp. 733-753.
27. Y. Sun and J. Xin, *Nonnegative Sparse Blind Source Separation for NMR Spectroscopy by Data Clustering, Model Reduction, and $\ell_1$ Minimization*, SIAM J. Imaging Sci., 5(3) (2012), pp. 886-911.
28. M.E. Winter, *N-findr: an algorithm for fast autonomous spectral endmember determination in hyperspectral data*, in Proc. of the SPIE, vol. 3753, 1999, pp. 266-275.
29. L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.