



# A geometric blind source separation method based on facet component analysis

Penghang Yin · Yuanchang Sun · Jack Xin

Received: 19 February 2014 / Revised: 31 August 2014 / Accepted: 1 September 2014 / Published online: 26 September 2014  
© Springer-Verlag London 2014

**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*. 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) 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, total variation technique may be employed. 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

## 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, and differential optical absorption spectroscopy. 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.

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

$$X = A S + N, \quad (1.1)$$

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 and  $S = [s(1), \dots, s(p)] \in \mathbb{R}^{n \times p}$  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, and nonnegativity [3, 6, 7, 11, 12, 14–16, 19, 22, 23]. As a matrix factorization problem, 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

P. Yin · J. Xin  
Department of Mathematics, University of California at Irvine,  
Irvine, CA 92697, USA  
e-mail: penghang@uci.edu

J. Xin  
e-mail: jxin@math.uci.edu

Y. Sun (✉)  
Department of Mathematics and Statistics, Florida International  
University, Miami, FL 33199, USA  
e-mail: yuasun@fiu.edu

$$X = A S + N = (A P \Lambda)(\Lambda^{-1} P^{-1} S) + N. \quad (1.2)$$

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, 13, 19–26]. The geometric observation [1, 19, 26] 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 [26]) is now a benchmark in HSI, see [5, 19–22] for its more recent variants. The method termed *vertex component analysis* (VCA) proposed in [21] is worth mentioning here being a fast unmixing algorithm for hyperspectral data. In NMR spectroscopy that motivates our work here, PPA was reformulated by Naanaa and Nuzillard in [19]. The source signals are only required to be nonoverlapping at some locations of acquisition variable (e.g., frequency).

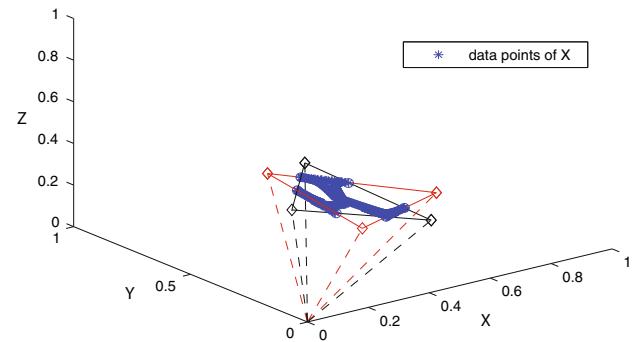
**Assumption (PPA):** For each  $i \in \{1, 2, \dots, n\}$ , there exists an  $j_i \in \{1, 2, \dots, p\}$  such that  $s_{i,j_i} > 0$  and  $s_{k,j_i} = 0$  ( $k = 1, \dots, i-1, i+1, \dots, n$ ), where  $s_{i,j_i}$  (resp.,  $s_{k,j_i}$ ) is the  $(i, j_i)$ -th (resp.,  $(k, j_i)$ -th) entry of  $S$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

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  $A$ . Moreover, every column of  $A$  is colinear to some column of  $X$ . To find the vertices (or edges) of the cone, the following optimization problem is solved for each column  $X^k$  of  $X$  [19]:

$$\min \text{ score} = \min_{\lambda_j \geq 0} \left\| \sum_{j=1, j \neq k}^p X^j \lambda_j - X^k \right\|_2^2. \quad (1.3)$$

A high score means that the corresponding column is far from being a nonnegative linear combination of other columns. So the  $n$  rows from  $X$  with highest scores are selected to form  $A$ , the mixing matrix.

If PPA is violated, the vertices are no longer in the data matrix  $X$  (up to a constant factor) 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. Recently, a dual cone-based approach to BSS problem was proposed in [20]. But this method appears indirect and computationally unwieldy, besides requiring the orthogonality of source



**Fig. 1** Nonuniqueness of cone  $\mathcal{A}$ . The data points of  $X$  (in blue) and two convex hulls  $\text{Convhull}(A)$  (black and red) are formed by selecting three facets of cone  $\mathcal{X}$  (color figure online)

signals. Here, we opt to solve the problem in a greedy manner directly by identifying the facets of the data cone under unique solvability conditions, which we call facet component analysis (FCA).

The rest of the paper is organized as follows. In Sect. 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 nonuniqueness. In Sect. 3, we present computational examples and results. For heavily noisy data, a denoising method based on total variation is discussed. Concluding remarks and future work are given in Sect. 4.

## 2 Proposed method

First of all, we emphasize 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 among their peaks.

### 2.1 Facet component analysis

In order to illustrate the basic idea behind FCA, we first consider nonnegative matrix factorization (NMF) problem in the noiseless case:

$$X = A S$$

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 overdetermined case.

The first source condition on the problem is as follows:

**Assumption 1** For each  $i = 1, \dots, m$ , 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

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

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  $\mathcal{X} = \text{Cone}(X)$  and  $\mathcal{A} = \text{Cone}(A)$ , then we have the following theorem (or combining Lemma 3 and Lemma 5 in [20]):

**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.

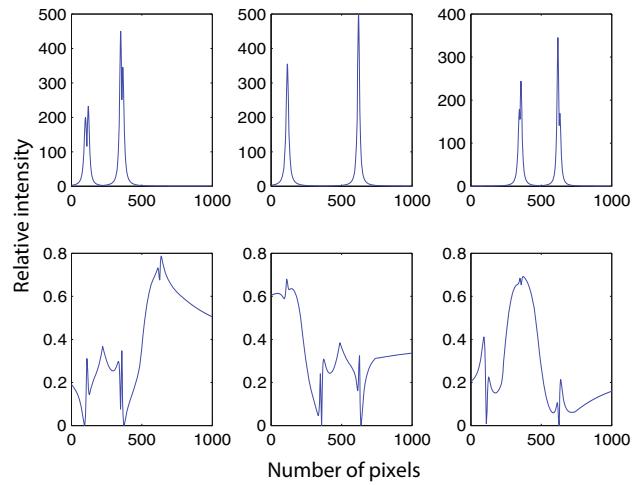
*Proof*  $\forall x \in \mathcal{X}$ , let  $x = X\alpha$ ,  $\alpha \geq 0$ . So  $x = A S\alpha = A(S\alpha)$ , where  $S\alpha \geq 0$ . So clearly  $x \in \mathcal{A}$ .

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)  $\mathcal{X} \subseteq \mathcal{A}$ .  $\square$

Since  $A$  is nonsingular,  $\mathcal{A}$  has  $m$  edges and thus has  $\binom{m}{m-1} = m$  facets. After projecting column vectors of  $X$  onto the hyperplane  $\mathbf{x}^\top \cdot \mathbf{1} = 1$ , where  $\mathbf{x} = (x_1, \dots, x_m)^\top$  and  $\mathbf{1} = (1, \dots, 1)^\top$ , the resulting data points together with the origin  $\mathbf{0} = (0, \dots, 0)^\top$  form a  $m$ -dimensional convex hull denoted by  $\text{Convhull}(X)$ . We then acquire all the facets and the associated vertices of  $\text{Convhull}(X)$ , 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.

However, Theorem 1 implies that  $\mathcal{X}$  generally has more facets than  $\mathcal{A}$ , making the selections of  $A$  clearly nonunique. So we propose an additional source assumption based on the density property of the data points to provide a selection criterion.

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



**Fig. 2** Top row sources recovered by selecting  $\mathcal{A}$  according to Assumption 2 (red convex hull in Fig. 1). Bottom row sources recovered by selecting black convex hull in Fig. 1

**Remark 1** We note in passing that PPA is a special (much more restrictive) case of Assumption 1. Moreover, PPA also implies Assumption 2 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.

**Remark 2** 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 FCA. In Fig. 1, the red  $\text{Convhull}(A)$  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 are shown in the top row of Fig. 2, while the bottom row of Fig. 2 are the recovered sources if we selected the black  $\text{Convhull}(A)$  in Fig. 1. This example is an illustration that selecting mixing matrix  $A$  under Assumption 2 finally yields sparse sources  $S$ .

We then count and sort the number of data points in each facet of  $\text{Convhull}(X)$  followed by 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  $\mathbf{x}^T \cdot \mathbf{1} = 1$ , we obtain all the columns of  $A$ . In the last step, we use nonnegative least squares to recover the source matrix  $S$ .

Let us include the additive noise and summarize the algorithm under Assumptions 1–2 as follows:

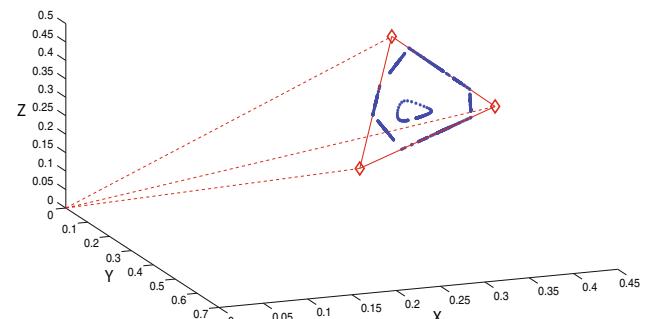
**Algorithm 1** (Facet component analysis)  $(A, S) = FCA(X, \rho, \epsilon, \sigma, \delta)$ ; parameters  $\rho > 0$ ;  $\epsilon, \sigma, \delta \in (0, 1)$ .

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  $\mathbf{x}^T \cdot \mathbf{1} = 1$ .
2. (Convex hull) Add the origin  $\mathbf{0}$  as the first column to  $X$ . Find all the facets and vertices of  $\text{Convhull}(X)$  using the MATLAB function `convhulln`, keep only the facets with the vertex  $\mathbf{0}$ . 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  $\mathbf{x}^T \cdot \mathbf{b}_i = 0$ , where  $\mathbf{b}_i$  is the normal vector with length 1. Select  $m$  planes from those of  $G_i$  with the largest cardinalities such that  $\mathbf{b}_i^T \cdot \mathbf{b}_j < \delta \approx 1$ , 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  $\mathbf{x}^T \cdot \mathbf{1} = 1$  to form  $A$ .
6. (Source recovering) For each  $X^j$ , solve the nonnegative least squares problem to find  $S^j$ :

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

**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  $\text{SNR} \approx 30$  dB,  $\epsilon$  and  $\sigma$  should be on the order of  $10^{-2}$ ; if high-level noise is present, they increase up to the order of 0.1.
2. The value of  $\rho$  is also positively correlated with the amount of noise. We suggest  $\rho$  not exceed  $\frac{1}{2} \|X\|_1$ , where  $\|X\|_1$  is maximum absolute column sum of  $X$ .
3. In step 4, we need the parameter  $\delta$  to guarantee the selected facets of  $\text{Convhull}(X)$  to be sufficiently distinct. We can just fix  $\delta = 0.99$ , which works well in all of our experiments.



**Fig. 3** 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 (color figure online)

## 2.2 Template-assisted FCA

We recognize another scenario of nonuniqueness: Among the  $n$  facets  $\mathcal{X}$ ,  $q$  of them contain the same number of data points and  $q > m$ ; see Fig. 3, for example. In this case, Assumption 2 would fail to guarantee the unique selection of cone  $\mathcal{A}$ . There are  $\binom{q}{m} = \frac{q!}{m!(q-m)!}$  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  $q$  is large. To overcome this nonuniqueness 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  $T \in \mathbb{R}^{N \times q}$  ( $N \gg m$ ) 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  $S = C T$ , 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,

$$\min_{C_j \geq 0} \mu \|C_j\|_1 + \frac{1}{2} \|S_j - C_j T\|_2^2, \quad (2.1)$$

Define  $u = (C_j)^T$ ,  $f = (S_j)^T$ ,  $B = T^T$  and rewrite (2.1) as

$$\min_{u \geq 0} \mu \|u\|_1 + \frac{1}{2} \|f - B u\|_2^2. \quad (2.2)$$

This problem could be solved by the linearized Bregman iterative method [10] by introducing an auxiliary variable  $v^l$ :

$$\begin{cases} v^{l+1} = v^l - B^T (B u^l - f), \\ u^{l+1} = \delta \cdot \text{shrink}_+(v^{l+1}, \mu), \end{cases} \quad (2.3)$$

where  $u^0 = v^0 = \mathbf{0}$ ,  $\delta > 0$  is the step size, and  $shrink_+$  is given by

$$shrink_+(v, \mu) = \begin{cases} v - \mu, & \text{if } v > \mu, \\ 0, & \text{if } v < \mu. \end{cases} \quad (2.4)$$

**Remark 4** 1. If  $k = m$ , all the source signals being recovered have spectral templates in the database. Solving  $S = C T$  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 nonuniqueness 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.

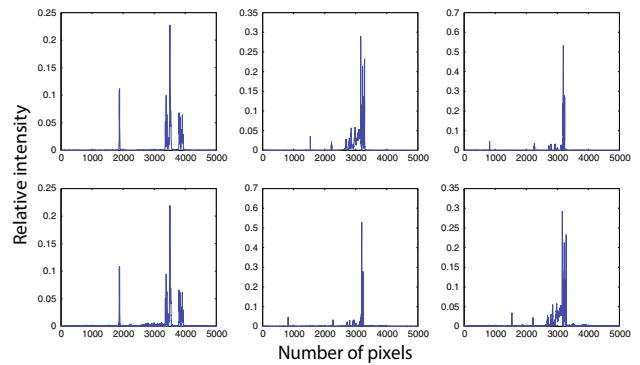
### 3 Numerical experiments

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

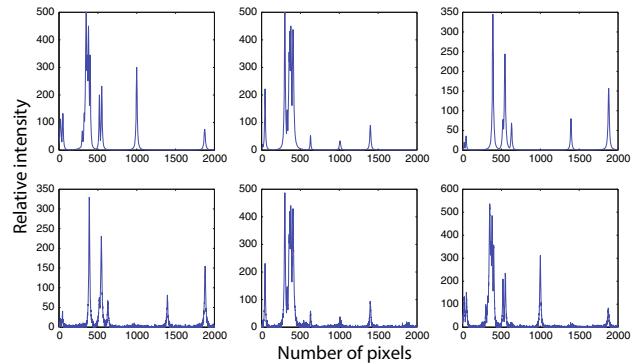
The first example involves source signals with stand-alone peaks. For the data, we use true NMR spectra of compounds  $\beta$ -cyclodextrine,  $\beta$ -sitosterol, and menthol as source signals. The NMR spectrum of a chemical compound is produced by the Fourier transformation of a time-domain signal which is a sum of sine functions with exponentially decaying envelopes [9]. 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. 4 from left to right. The mixtures were obtained by adding white Gaussian noise with SNR = 30 dB. For the parameters, we set  $\rho = 5 \times 10^{-2}$ ,  $\epsilon = \sigma = 10^{-2}$ ,  $\delta = 0.99$  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.

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

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



**Fig. 4** 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. 5** Top row true sources in Example 2. Bottom row computed sources by our method

In the second example, we use three Lorentzian-shaped synthetic sources that violate the PPA. We add white Gaussian noise with SNR = 30 dB, and the recovery result is plotted in Fig. 5. True mixing matrix  $A_2$  and computed  $\hat{A}_2$  are listed below. The result is achieved by setting  $\rho = 80$ ,  $\epsilon = \sigma = 2 \times 10^{-2}$ ,  $\delta = 0.99$ .

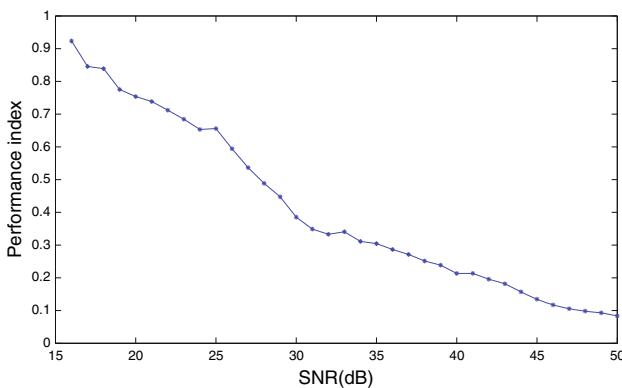
$$A_2 = \begin{pmatrix} 0.3000 & 0.5000 & 0.4545 \\ 0.5000 & 0.4286 & 0.1818 \\ 0.2000 & 0.0714 & 0.3636 \end{pmatrix}$$

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

In [8], the Comon's index is introduced to measure the performance of source separations. Let  $A$  and  $\hat{A}$  be two nonsingular matrices with  $\ell_2$ -normalized columns. Then, the Comon's index  $\epsilon(A, \hat{A})$  between  $A$  and  $\hat{A}$  is defined as

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

$$+ \sum_i \left| \sum_j |d_{ij}|^2 - 1 \right| + \sum_j \left| \sum_i |d_{ij}|^2 - 1 \right|,$$



**Fig. 6** Robust performance of FCA in the presence of noise

where  $d_{ij}$  is the entry of  $A^{-1}\hat{A}$ .  $A$  and  $\hat{A}$  are considered nearly equivalent in the sense of BSS (i.e.,  $\hat{A} = A P \Lambda$ ) if  $\epsilon(A, \hat{A}) \approx 0$ . With this concept, we then show the robust performances of FCA in the presence of noises. The three sources in Example 2 were combined to generate three mixtures corrupted by Gaussian noises with SNR varying from 16 to 50 dB. At each noise level, Comon's indices were averaged over 50 independent trials. Figure 6 indicates the robustness of our method with small indices even in the low SNR zone.

In the next example, we show the computational results of template-assisted FCA. We are to separate three source signals from three mixtures, where the Convhull( $X$ ) is a hexagon in Fig. 3. The data are synthetically generated so that each side contains the same number of data points. In Fig. 7, 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. 7, 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. 7. The corresponding source recovery is shown in Fig. 8. Suppose the three source signals have spectral references from a database of size ten. For each three recovered sources from Fig. 8, we compare the recovered signals against the database by solving (2.1). 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,

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

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

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

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

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

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

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

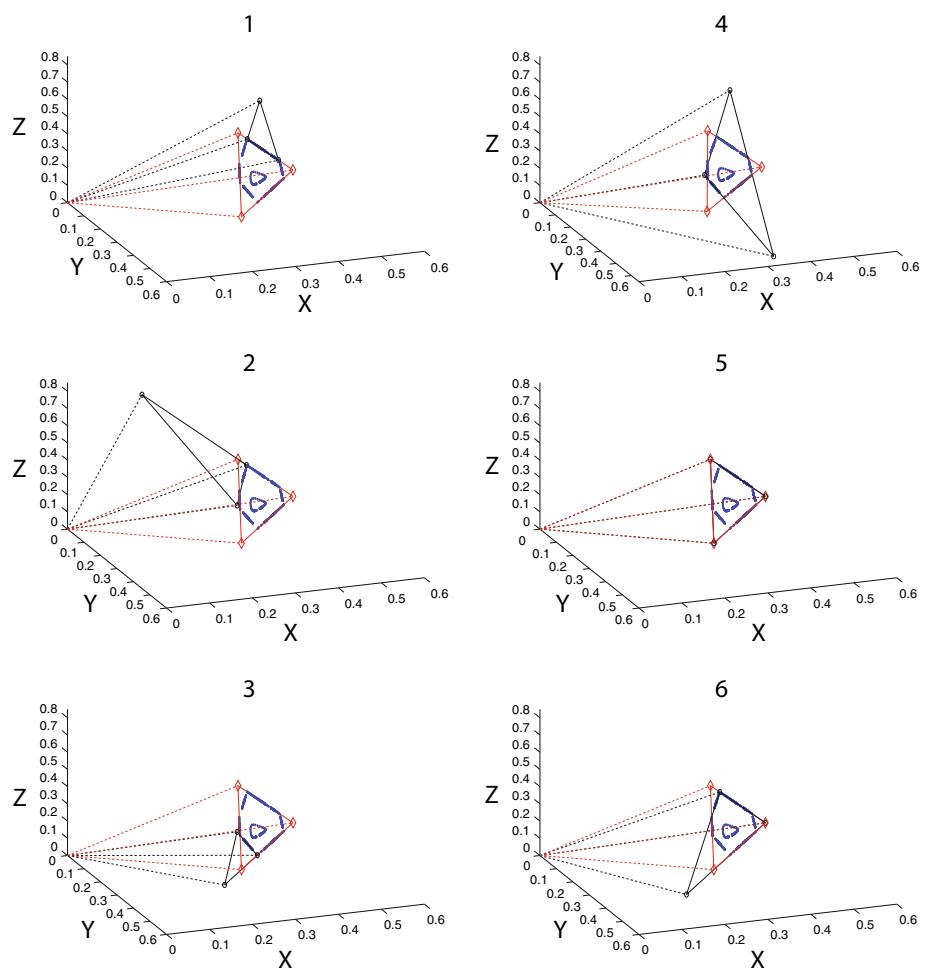
Both  $C^5$  and  $C^6$  contain a row of sparsity one, which indicates that the recoveries  $S^5$  and  $S^6$  contain a source that has a spectral reference from the database. Therefore, we will eliminate cone 4 but keep cones 5 and 6 as two possible solutions.

In the last example, we present the comparisons of FCA with some existing BSS methods. The first comparison is between FCA and VCA. VCA focuses on identifying the cone by locating its vertices based on PPA. FCA also works for the separation of data of PPA, which has been shown previously. Here, we present in Fig. 9 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. Computation was performed on a PC with 4G RAM and 3.0 GHz Intel Pentium CPU. The CPU time for FCA was 0.3750 seconds, and it was 0.6094 seconds for VCA. FCA sees a cpu time saving 38 % comparing to VCA, and this improvement is consistent across other data sets under the same assumption. We also compare FCA and NMF and show the result in Fig. 10. NMF decomposes the mixture  $X$  into  $A$  and  $S$  by solving

$$\min_{A, S} \|X - AS\|_2^2 \quad \text{s.t. } S \geq 0, \quad A \geq 0. \quad (3.1)$$

Optimization approaches for solving (3.1) include gradient descent and alternating least squares. FCA is clearly better than NMF for the stand-alone peak source signals. But we point out that if no knowledge of the source spectra other than the positivity is available, NMF should be used.

**Fig. 7** Nonuniqueness of cone  $\mathcal{A}$  when Assumption 2 fails to be satisfied. Data points are in blue, the ground truth of convex hull  $\text{Convhull}(A)$  is red, while the black color shows the identified  $\text{Convhull}(A)$  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) (color figure online)



### 3.1 Total variation 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 propose to apply the total variation (TV) idea of image denoising by Rudin et al. [27] to noise removal, combining with the FCA after step 3. In the following, we shall use the example of a point cloud in  $xyz$  plane to illustrate the idea of TV denoising. We first preprocess the data by rescaling them onto a plane  $x + y + z = 1$ , and the projected data are two dimensional. For each point  $(x, y)$ , a distance function to the point cloud is  $d(x, y) = \min_{x_i, y_i} \sqrt{(x - x_i)^2 + (y - y_i)^2}$ , where  $(x_i, y_i)^T$  corresponds to the  $i$ -th column of  $X$ , note that  $z_i$  is not included since the  $z_i = 1 - x_i - y_i$ . For computation, the distance function will be restricted on a rectangular region which contains the point cloud.

We use the recent *Chambolle's Algorithm* [4] to solve the Rudin–Osher–Fatemi model to obtain a denoised distance function  $u(x, y)$ :

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

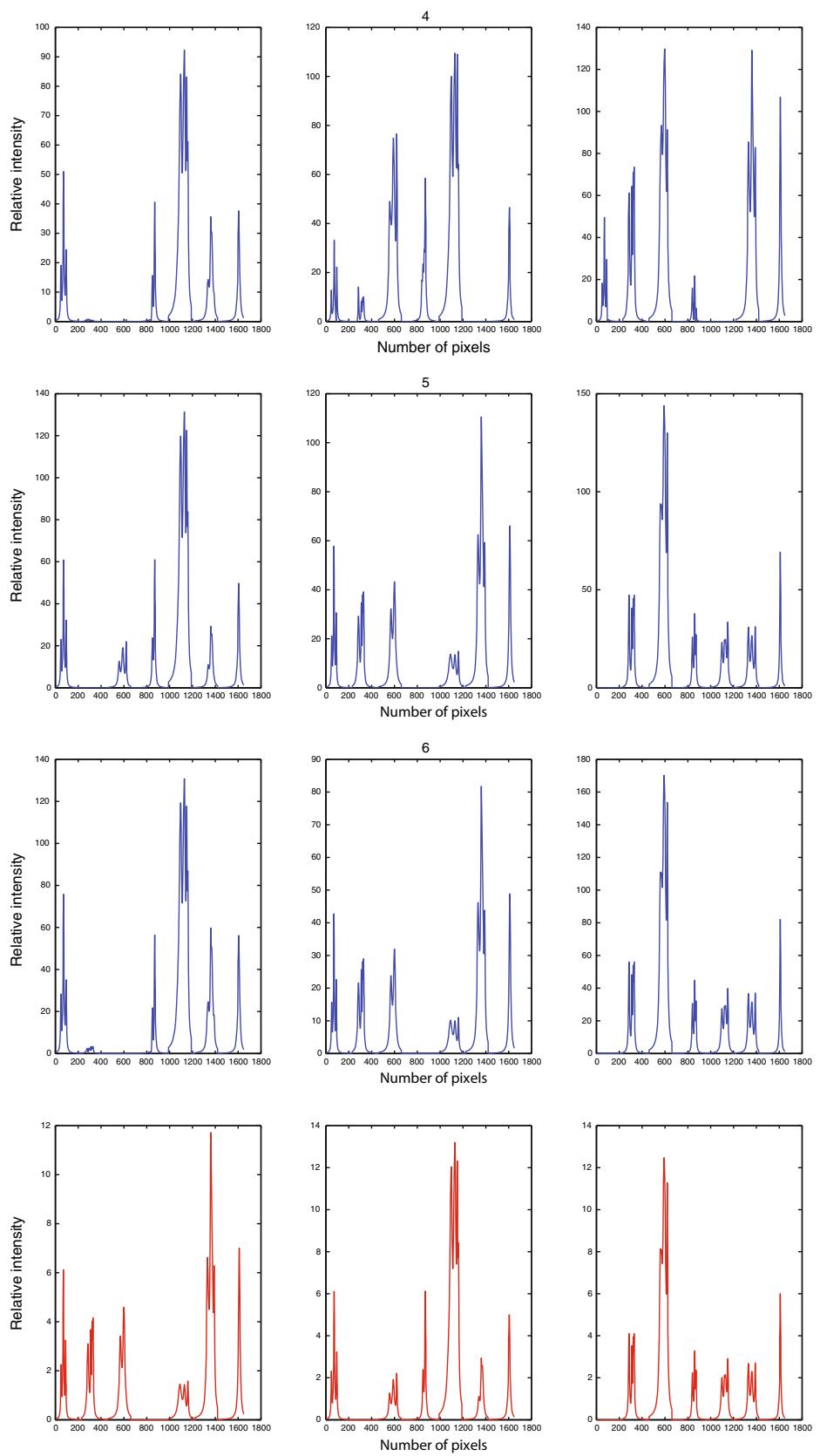
where  $\text{TV}(u)$  is the anisotropic TV of  $u$  defined as

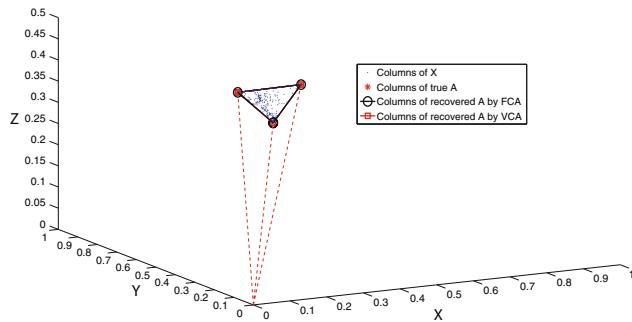
$$\text{TV}(u) = \sum_{i,j} |u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|.$$

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. 11. The detected planes from both of them are shown in Fig. 12 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) while smoothing away noise in flat regions. The idea of denoising distance function by total variation extends to point cloud of any dimension.

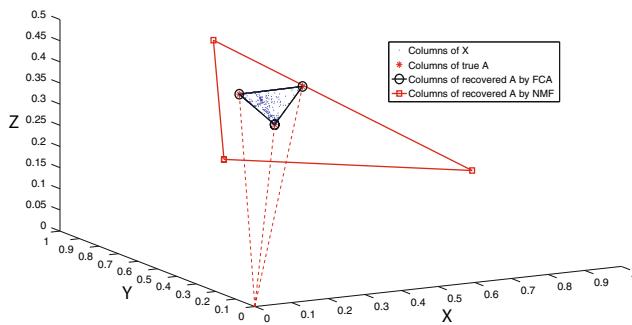
We show a comparison of FCA with and without TV denoising in Fig. 13. The mixtures were corrupted by Gaussian noises varying from 8 to 25 dB. With TV denoising the Comon's indices are substantially lower than without it, resulting in better separation performance.

**Fig. 8** Rows one to three are the source signals recovered by selecting black convex hulls in the right panel in Fig. 7. Bottom row is the three true source signals

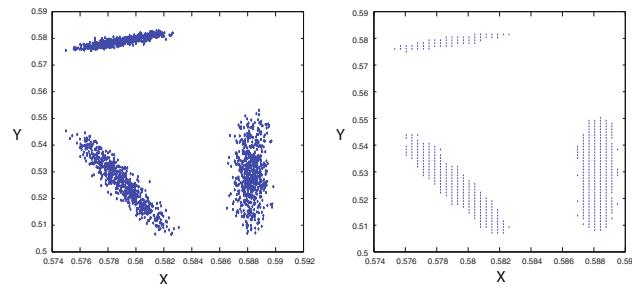




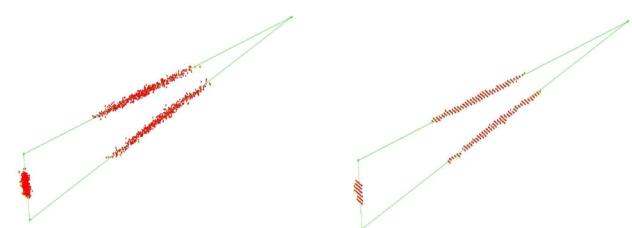
**Fig. 9** Comparison between FCA and VCA for data satisfying stand-alone peak assumption



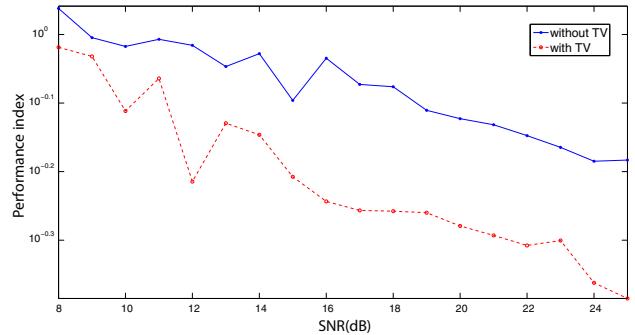
**Fig. 10** Comparison between FCA and NMF for data satisfying stand-alone peak assumption



**Fig. 11** 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  $x + y + z = 1$



**Fig. 12** Computational results of planes from noisy data (left) and denoised data (right) shown in Fig. 11. The Green lines are the detected planes using the method proposed in the paper (color figure online)



**Fig. 13** Comparison of Comon's indices (in log scale) of FCA with and without TV denoising at high noise levels

#### 4 Concluding remarks

We developed a novel FCA for nonnegative blind source separation problems. We presented facet-based unique solvability conditions 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, TV denoising 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 VCA (under PPA) and the NMF (under only nonnegativity).

A line of future work is to separate more source signals from their mixtures, known as an undetermined blind source separation, or uBSS.

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

#### References

- Boardman, J.: Automated spectral unmixing of AVIRIS data using convex geometry concepts. In: Summaries of the IV Annual JPL Airborne Geoscience Workshop, JPL Pub. 93-26, **1**, 11–14 (1993)
- Bobin, J., Starck, J.-L., Fadili, J., Moudden, Y.: Sparsity and morphological diversity in blind source separation. *IEEE Trans. Image Process.* **16**, 2662–2674 (2007)
- Bofill, P., Zibulevsky, M.: Underdetermined blind source separation using sparse representations. *Signal Process.* **81**, 2353–2362 (2001)
- Chambolle, A.: An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **20**, 89–97 (2004)
- Chang, C.I. (ed.): *Hyperspectral Data Exploitation: Theory and Applications*. Wiley-Interscience, Hoboken (2007)
- Choi, S., Cichocki, A., Park, H., Lee, S.: Blind source separation and independent component analysis: a review. *Neural Inf. Process. Lett. Rev.* **6**, 1–57 (2005)

7. Cichocki, A., Amari, S.: Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications. John Wiley and Sons, New York (2005)
8. Comon, P.: Independent component analysis—a new concept? *Signal Process.* **36**, 287–314 (1994)
9. Ernst, R., Bodenhausen, G., Wokaun, A.: Principles of Nuclear Magnetic Resonance in One and Two Dimensions. Oxford University Press, Oxford (1987)
10. Guo, Z., Osher, S.: Template matching via  $\ell_1$  minimization and its application to hyperspectral target detection. *Inverse Probl. Imaging* **5**, 19–35 (2011)
11. Hoyer, P.: Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.* **5**, 1457–1469 (2004)
12. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley and Sons, New York (2001)
13. Klingenber, B., Curry, J., Dougherty, A.: Non-negative matrix factorization: Ill-posedness and a geometric algorithm. *Pattern Recognit.* **42**, 918–928 (2009)
14. Liu, J., Xin, J., Qi, Y.-Y.: A dynamic algorithm for blind separation of convolutive sound mixtures. *Neurocomputing* **72**, 521–532 (2008)
15. Liu, J., Xin, J., Qi, Y.-Y.: A soft-constrained dynamic iterative method of blind source separation. *SIAM J. Multiscale Model. Simul.* **7**, 1795–1810 (2009)
16. Liu, J., Xin, J., Qi, Y.-Y., Zeng, F.-G.: A time domain algorithm for blind separation of convolutive sound mixtures and  $\ell_1$  constrained minimization of cross correlations. *Comm. Math. Sci.* **7**(1), 109–128 (2009)
17. McDonnell, M.: Box-filtering techniques. *Comput. Graph. Image Process.* **17**, 65–70 (1981)
18. Motzkin, T., Raiffa, H., Thompson, G., Thrall, R.J.: The Double Description Method. *Annals of Math Studies*, Vol. 8, Princeton University Press, pp. 51–73 (1953)
19. Naanaa, W., Nuzillard, J.-M.: Blind source separation of positive and partially correlated data. *Signal Process.* **85**(9), 1711–1722 (2005)
20. Naanaa, W.: A Geometric Approach to Blind Separation of Non-negative and Dependent Source Signals. 18th European Signal Processing Conference (EUSIPCO-2010), Aalborg, Denmark, August 23–27, pp. 747–750 (2010)
21. Nascimento, J.M.P., Bioucas-Dias, J.M.: Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **43**(4), 898–910 (2005)
22. Sun, Y., Ridge, C., del Rio, F., Shaka, A.J., Xin, J.: Postprocessing and sparse blind source separation of positive and partially overlapped data. *Signal Process.* **91**(8), 1838–1851 (2011)
23. Sun, Y., Xin, J.: Under-determined sparse blind source separation of nonnegative and partially overlapped data. *SIAM J. Sci. Comput.* **33**(4), 2063–2094 (2011)
24. Sun, Y., Xin, J.: A recursive sparse blind source separation method and its application to correlated data in NMR spectroscopy of biofluids. *J. Sci. Comput.* **51**, 733–753 (2012)
25. Sun, Y., Xin, J.: Nonnegative sparse blind source separation for NMR spectroscopy by data clustering, model reduction, and  $\ell_1$  minimization. *SIAM J. Imaging Sci.* **5**(3), 886–911 (2012)
26. Winter, M.E.: N-findr: an algorithm for fast autonomous spectral endmember determination in hyperspectral data. *Proc. SPIE* **3753**, 266–275 (1999)
27. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* **60**, 259–268 (1992)