

Article

# Big Data Blind Separation

Mujahid N. Syed 

Department of Systems Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; smujahid@kfupm.edu.sa or snumujahid@gmail.com; Tel.: +966-013-860-7908

Received: 8 December 2017; Accepted: 23 February 2018; Published: 27 February 2018

**Abstract:** Data or signal separation is one of the critical areas of data analysis. In this work, the problem of non-negative data separation is considered. The problem can be briefly described as follows: given  $\mathbf{X} \in \mathbb{R}^{m \times N}$ , find  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{S} \in \mathbb{R}_+^{n \times N}$  such that  $\mathbf{X} = \mathbf{AS}$ . Specifically, the problem with sparse locally dominant sources is addressed in this work. Although the problem is well studied in the literature, a test to validate the locally dominant assumption is not yet available. In addition to that, the typical approaches available in the literature sequentially extract the elements of the mixing matrix. In this work, a mathematical modeling-based approach is presented that can simultaneously validate the assumption, and separate the given mixture data. In addition to that, a correntropy-based measure is proposed to reduce the model size. The approach presented in this paper is suitable for big data separation. Numerical experiments are conducted to illustrate the performance and validity of the proposed approach.

**Keywords:** big data; blind signal separation; locally dominant sources; correntropy ranking

## 1. Introduction

Transforming data into information is a key research direction of the current scientific age. During the data collection phase, it is often the case that the data cannot be collected from the actual data generating locations (sources). Typically, a nearby physically connected location (station) that is accessible can be used for the data collection. If the station is influenced by more than one source, then the data collected at the station provides mixed information from the multiple sources (mixture data). This creates a challenging problem of identifying the source data from the given mixture data. Such a problem is typically known as a data (or signal) separation problem.

In this paper, a linear mixing type data separation problem is considered. The generative model of the problem in its standard form can be written as:

$$\mathbf{X} = \mathbf{AS}, \quad (1)$$

where  $\mathbf{X} \in \mathbb{R}^{m \times N}$  denotes the given mixture matrix,  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is the unknown mixing matrix, and  $\mathbf{S} \in \mathbb{R}_+^{n \times N}$  denotes the unknown source matrix. The problem can be further classified into overdetermined ( $m > n$ ), undetermined ( $m < n$ ), and square or determined ( $m = n$ ) cases. The overdetermined case can be transformed into the square case by using the Principal Component Analysis (PCA) method [1,2]. The undetermined cases often result in loss of information or redundancy in the representation. Usually, approximate recovery is done based on the prior probability assumptions. Typically, Gaussian or Laplacian priors are used to estimate the mixing matrix. The idea is to identify the directions of maximum density via clustering. These directions correspond to the identification of the mixing matrix [3,4]. Once the mixing matrix is estimated, the source matrix is obtained by solving series of least square problems. In this paper, the data separation problem for  $m \geq n$  cases will be considered.

In addition to that, the above problem given in Equation (1) can be solved under numerous assumptions based on the level of available information. In this work, both  $\mathbf{A}$  and  $\mathbf{S}$  are assumed to be unknown. In such a scenario, the above signal separation problem is known as the “Blind” Signal Separation (BSS) problem. Seminal ideas of the BSS problem can be found in [5], where the authors illustrated the idea of BSS via an example of two source signals ( $n = 2$ ) and two mixture signals ( $m = 2$ ). Their objective was to recover source signals from the mixture signals, without any further information. However, it has been known since then that the recovery of sources from a linear mixture is imperfect, and pragmatic recovery needs various identifiability conditions. In [6], various approaches and the identifiable conditions of the BSS problem are summarized.

One of the well known traditional solution approaches to the BSS problem is the Independent Component Analysis (ICA) [7]. The ICA-based approaches are built upon the key assumption of the statistical independence among the rows of  $\mathbf{S}$ . The basic objective of ICA is to separate the input data into statically independent components. During the last three decades, there have been a large amount of papers on signal analysis that are devoted to the usage of ICA. These approaches are designed not only for the linear but also for the nonlinear signal mixing scenarios. They have also been designed for the overdetermined, undetermined, as well as square cases of the BSS problem. Some of the well known approaches in ICA are based on mutual information [8]; negentropy [9]; projection pursuit [10]; infomax [11]. Recent applications of ICA include bio-medical data analysis [12–17]; power system analysis [18], and audio and speech [19]. It is out of the scope of this paper to summarize all the typical approaches of ICA. Therefore, interested readers are referred to [7,20] (and the references therein). One of the major critiques pertaining to ICA is the existence of independent sources. Some of the signal analysis areas (such as biomedical and hyperspectral image analysis) may not satisfy the independence criterion (see [21]). Thus, alternative approaches to ICA for BSS have also been a critical research direction in the area of signal analysis.

Other prominent approaches of BSS, apart from ICA, assume some sparsity or geometric structure in  $\mathbf{S}$  [22]. In this work, the focus is on the sparsity-based approaches of extracting non-negative sources. This area is widely known as Sparse Component Analysis (SCA) [23]. Specifically, following are the basic assumptions that are considered while solving the SCA problem.

**Basic SCA assumptions:**

1. Every column of the source matrix is non-negative.
2. Source matrix has a full row rank.
3. Mixing matrix has a full column rank, and  $m \geq n$ .
4. The rows of the source matrix, and columns of the mixing matrix have unit norm.
5. Source matrix is sparse.

The first assumption provides a mathematical advantage in designing the solution algorithms. Basically, the non-negativity assumption transforms the BSS problem into a convex programming problem [24–26]. In addition to that, non-negative source signals are very common in sound and image analysis. The next two assumptions ensure that the problem is recoverable (solvable). The fourth assumption is perhaps the limit of all BSS approaches, which is related to scalability and uniqueness (see [27]). The fifth assumption is the key sparsity assumption of the SCA-related approaches. Different scenarios of the SCA problem arise with different structures of the sparsity. The typical structures of sparsity discussed in the SCA literature can be classified as:

- **Locally Dominant Case:** In addition to the basic assumptions, for a given row  $r$  of  $\mathbf{S}$ , there exists at least one unique column  $c$  such that:

$$s_{i,c} \begin{cases} > 0 & \text{if } i = r \\ = 0 & \text{otherwise} \end{cases} \quad (2)$$

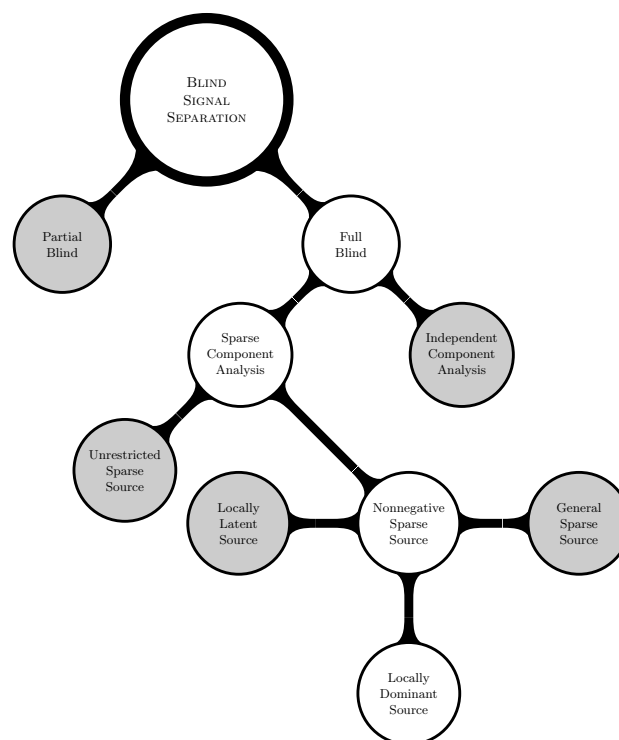
- **Locally Latent Case:** In addition to the basic assumptions, for a given row  $r$  of  $\mathbf{S}$ , there exists at least  $(n - 1)$  linearly independent and unique columns  $C_r$  such that:

$$s_{i,c} \begin{cases} = 0 & \text{if } i = r \text{ \& } c \in C_r \\ > 0 & \text{if } i \neq r \text{ \& } c \in C_r \end{cases} \quad (3)$$

- **General Sparse Case:** This is the default case.

The first case is one of the widely known cases in the SCA literature (see [28,29] for recent literature reviews). The second case is new to the of SCA literature, and few recent papers address this case [30]. The first two cases have identifiability conditions, which assure identification and recovery of the source signals. If the conditions of the first two cases do not apply to the given data, then the SCA problem belongs to the last general case. Typically, the general case may not have perfect identification (apart from the scalability and uniqueness issues). For the general case, minimum volume-based approaches [26,31,32], and extreme direction-based approaches [33] have been proposed to approximately recover the source matrix. Some special cases of sparse structure, apart from the above cases, that can recover original source data have also been studied in the literature, for example see [34].

In addition to the above, the conditions on  $\mathbf{X}$  that improve the separability of sources are studied in the literature [35,36]. Methods that exploit spectral variability can be seen in [37]. Time series (frequency and transformation analysis) based methods to identify sparse sources have been presented in [4,38]. Further methods developed on the assumption of SCA can be found in [39,40]. The prominent application areas of SCA include, but are not limited to, the following: Blind Hyperspectral Unmixing (BHU) [41], chemical analysis [42], Nuclear Magnetic Resonance (NMR) spectroscopy [43], etc. Figure 1 portrays various linear BSS methods available in the literature. The grayed areas in the figure represent the research areas that will not be considered further in this paper. Thus, the branches unrelated to the paper that emerge from the grayed areas are ignored in the figure.



**Figure 1.** Overview of the “Blind” Signal Separation (BSS) problem.

One of the critical gaps found in the SCA literature is the unavailability of a method to test the existence of a locally dominant assumption from the mixture data ( $\mathbf{X}$ ). In this paper, a novel mathematical programming and correntropy-based approach is presented for testing the locally dominant source assumption. The proposed approach also provides a solution for the locally dominant SCA problem.

Throughout this paper, the following notation styles are used. A capital letter bold face character, like  $\mathbf{B}$ , indicates a matrix. A small letter bold face character with or without a subscript, like  $\mathbf{b}_r$ , indicates the  $r$ th column vector of matrix  $\mathbf{B}$ . A small letter bold face character with a special subscript, like  $\mathbf{b}_{p\bullet}$ , indicates the transpose of the  $p$ th row vector of matrix  $\mathbf{B}$ . A non-bold small letter character, like  $b_{p,r}$ , represents the  $p$ th row  $r$ th column element of matrix  $\mathbf{B}$ .

The rest of the paper is organized as follows: Section 2 introduces the locally dominant case. Specifically, it displays the existing formulations from the literature, and presents the proposed novel formulation. A correntropy-based ranking method to eliminate the non-extreme data points is developed in Section 3. By incorporating the proposed model and the proposed ranking method, a tailored solution approach for the big data separation problem is developed in Section 4. A numerical study to assert the performance of the proposed approach is illustrated in Section 5. Finally, the paper is concluded with discussions in Section 6.

## 2. Locally Dominant Case

Consider the following determined or square version of the SCA model:

$$\mathbf{X} = \mathbf{A} \mathbf{S}. \quad (4)$$

Each column  $\mathbf{x}_i$  for  $i = 1, \dots, N$  of the mixture matrix ( $\mathbf{X}$ ) can be represented as follows:

$$\mathbf{x}_i = \sum_{j=1}^n s_{j,i} \mathbf{a}_j \quad \forall i = 1, \dots, N, \quad (5)$$

where  $\mathbf{a}_j$  is the  $j$ th column of the mixing matrix ( $\mathbf{A}$ ), and  $s_{j,i}$  is the  $j$ th row  $i$ th column element of the source matrix ( $\mathbf{S}$ ). Equation (5) highlights that every column vector of  $\mathbf{X}$  is a linear combination of the column vectors of  $\mathbf{A}$ . Since the source matrix ( $\mathbf{S}$ ) is non-negative (i.e.,  $s_{j,i} \geq 0$  for all  $i$  &  $j$ ), the combination is a conic combination. Thus, the columns of  $\mathbf{X}$  are spanned by the columns of  $\mathbf{A}$ . In other words, the extreme column vectors of  $\mathbf{X}$  are the columns of  $\mathbf{A}$ . Therefore, the locally dominant case boils down to the identification of the extreme vectors of  $\mathbf{X}$  [24].

If all the columns of  $\mathbf{X}$  are non-negative, then normalizing every column of  $\mathbf{X}$  with respect to Norm-1 makes the columns of  $\mathbf{X}$  coplanar. That is, all the columns are contained in the following lower dimensional plane:

$$\sum_{j=1}^n x_{j,i} = 1 \quad \forall i = 1, \dots, N. \quad (6)$$

Now, the extreme points of  $\mathbf{X}$  on the lower dimensional plane correspond to the columns of  $\mathbf{A}$ . In addition to that, if some of the elements of  $\mathbf{X}$  are negative, then the columns of  $\mathbf{X}$  are projected onto a suitable lower dimensional plane. There are many approaches in the literature that are designed to work on this lower dimensional plane (affine hull) [25,44]. The advantage of working on this plane is that the extreme vector columns of  $\mathbf{X}$  will form the vertices of a lower dimensional simplex. Thus, identifying the extreme points will result in the identification of the mixing matrix. Next, a few well known mathematical formulations and solution approaches for SCA from the literature are presented.

### 2.1. Conventional Formulations

One of the earliest mathematical formulations that identifies the extreme vectors of  $\mathbf{X}$  is proposed in [24]. The idea is to pick one column of  $\mathbf{X}$ , say  $\mathbf{x}_c$ , and check the possibility of it being an extreme vector. The formulation corresponding to  $\mathbf{x}_c$  is given as follows:

$$\begin{aligned} \text{min. :} \\ \left\| \sum_{\substack{i=1 \\ i \neq c}}^N \alpha_i \mathbf{x}_i - \mathbf{x}_c \right\|_2 \end{aligned} \tag{7}$$

$$\begin{aligned} \text{s.t. :} \\ \alpha_i \geq 0 \quad \forall i = 1, \dots, N, \ \& \ i \neq c, \end{aligned} \tag{8}$$

where  $\alpha_i \geq 0, \in \mathbb{R}$  is the variable that corresponds to the weight of  $\mathbf{x}_i \in \mathbf{X}$  for  $i = 1, \dots, N$ . The key idea that is exploited in the formulation is that the extreme vectors cannot be represented by a non-negative weighted combination of the other data vectors. The above formulation is a least square minimization problem. In the worse case, the formulation has to be executed  $N$  times for each  $\mathbf{x}_c$ , i.e., for  $c = 1, \dots, N$ . In the best case, the formulation has to be executed  $n$  times. A nonzero value of the objective function indicates that the vector  $\mathbf{x}_c$  is an extreme vector of  $\mathbf{X}$ .

Another approach, called Convex Analysis of Mixtures of Non-negative Sources (CAMNS), that works on the affine hull is presented in [25]. The solution approach of CAMNS involves two major steps. In the first step, the parameters  $\mathbf{C} \in \mathbb{R}^{N \times (n-1)}$  and  $\mathbf{d} \in \mathbb{R}^N$  of the affine hull are estimated as follows:

$$\mathbf{d} = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_{j\bullet} \tag{9}$$

$$\mathbf{U} = \mathbf{X}^T - \mathbf{d} \tag{10}$$

$$\mathbf{C} = [\text{eigv}(\mathbf{U}\mathbf{U}^T)]_{1, \dots, (n-1)} \tag{11}$$

where  $\mathbf{x}_{j\bullet}$  is the  $j$ th row of  $\mathbf{X}$ . In Equation (10), vector  $\mathbf{d}$  is subtracted from all the rows of  $\mathbf{X}$ , and in Equation (11), matrix  $\mathbf{C}$  contains the columns corresponding to the eigenvectors associated with the largest  $(n - 1)$  eigenvalues of  $\mathbf{U}\mathbf{U}^T$ . Basically, this first step is similar to the dimensionality reduction process executed in the PCA. In the second step, the following mathematical model is repeatedly solved:

$$\begin{aligned} \text{max/min :} \\ \mathbf{r}^T (\mathbf{C}\boldsymbol{\alpha} + \mathbf{d}) \end{aligned} \tag{12}$$

$$\begin{aligned} \text{s.t. :} \\ \mathbf{c}_{i\bullet}^T \boldsymbol{\alpha} + d_i \geq 0 \quad \forall i = 1, \dots, N \end{aligned} \tag{13}$$

where  $\mathbf{r} \in \mathbb{R}^N$  is a generated vector,  $\mathbf{c}_{i\bullet}$  is an  $i$ th row of  $\mathbf{C}$ , and  $\boldsymbol{\alpha} \in \mathbb{R}^{(n-1)}$  is the unknown variable. The above formulation exploits the notion that the optimal solution of a linear program exists at the extreme points. For a given  $\mathbf{r}$ , the formulation is solved twice; one time as the maximization problem and the other time as the minimization problem. This is done in order to get one (or maybe two) new extreme point(s) in every iteration. The  $(n - 1)$  extreme points are identified by resolving the above formulation repeatedly with respect to different  $\mathbf{r}$  vectors. The crux of this method is hidden in the generation of  $\mathbf{r}$  and convergence of the approach, which are the main concepts presented in [25].

Notice that the above two ideologies extract the columns of  $\mathbf{A}$  sequentially. In addition to the above approaches, typical approaches in the literature identify the columns of  $\mathbf{A}$  sequentially [28,29,45]. Therefore, there is no mechanism to validate the locally dominant assumption from  $\mathbf{X}$ . One of the main

objectives of this paper is to build such a mechanism using a mathematical formulation-based approach. In the following subsection, a novel formulation is presented that can provide the above mechanism.

### 2.2. Envelope Formulation

In this section, a mathematical model that can simultaneously identify all the extreme vectors of  $\mathbf{X}$  under the locally dominant assumption is proposed. To the best of our knowledge, an exact method that identifies all the extreme vectors in a single shot is unavailable.

Let  $\mathbf{X}_2$  be the data obtained after normalizing each column of  $\mathbf{X}$  with respect to Norm-2. Let  $\mathbf{y}_i \in \mathbf{X}_2$  be the  $i$ th column of  $\mathbf{X}_2$ . Let  $\mathbf{a}^T \mathbf{y} = b$  be a plane that is inclined in such a way that all the columns of  $\mathbf{X}_2$  are contained in one halfspace of the hyperplane, and the origin is in the other halfspace. Such a hyperplane is referred to as linear envelope in this work. The envelope can be written as:

$$\mathbf{a}^T \mathbf{y}_i \geq b \quad \forall i = 1, \dots, N, \tag{14}$$

where  $\mathbf{a} \in \mathbb{R}^m$  corresponds to the normal vector of the envelope, and  $b \geq 0$  is a constant. Out of infinite possible representations of the above envelope, an envelope with  $b = 1$  is selected for further analysis. Now, the distance between any vector  $\mathbf{y}_i$  and the envelope can be written as:

$$d(\mathbf{y}_i) = \frac{|\mathbf{a}^T \mathbf{y}_i - 1|}{\|\mathbf{a}\|_2} \tag{15}$$

$$= \frac{\mathbf{a}^T \mathbf{y}_i - 1}{\|\mathbf{a}\|_2} \tag{16}$$

$$\propto \mathbf{a}^T \mathbf{y}_i - 1 = p(\mathbf{y}_i) \quad \forall i = 1, \dots, N. \tag{17}$$

Equation (16) follows from Equation (14). Ignoring the denominator ( $\|\mathbf{a}\|_2$ ) in Equation (16) results in a proportional or scaled distance ( $p(\mathbf{y}_i)$ ). Among infinite possible linear envelopes, the aim is to find the tightest or supporting envelope. A formulation that can identify the tightest envelope is given as follows:

$$\begin{aligned} \text{min. :} \\ \sum_{i=1}^N (\mathbf{a}^T \mathbf{y}_i - 1) \end{aligned} \tag{18}$$

$$\begin{aligned} \text{s.t. :} \\ \mathbf{a}^T \mathbf{y}_i \geq 1 \quad \forall i = 1, \dots, N, \end{aligned} \tag{19}$$

where  $\mathbf{a} \in \mathbb{R}^m$  is the unknown variable. The above formulation can be equivalently written as:

Formulation: (20)

$$\begin{aligned} \text{min. :} \\ \boldsymbol{\mu}^T \mathbf{a} \end{aligned} \tag{21}$$

$$\begin{aligned} \text{s.t. :} \\ \mathbf{y}_i^T \mathbf{a} \geq 1 \quad \forall i = 1, \dots, N, \end{aligned} \tag{22}$$

where  $\boldsymbol{\mu} \in \mathbb{R}^m$  is defined as  $\boldsymbol{\mu} = \sum_{i=1}^N \mathbf{y}_i$ . It is assumed that the duplicate and/or all zero columns of  $\mathbf{X}_2$  are removed before the execution of the above model. The aim of the above model is to find the envelope that has the minimum distance with respect to all the columns of  $\mathbf{X}_2$ . The above formulation is linear, and needs to be executed only once to identify all the extreme vectors. That is, at the optimal solution, the data points ( $\mathbf{y}_i$ 's) corresponding to the active constraints in Equation (19) will correspond to the extreme vectors of  $\mathbf{X}_2$ . It can be seen that the above formulation is always feasible. Furthermore,

due to the design of the constraint given in Equation (19), the problem is linear. That is, the design allowed the usage of proportional distance (Equation (17)) instead of the nonlinear actual distance (Equation (15)).

In addition to that, from the LP theory, only  $m$  constraints will be active at the optimal solution. Let  $\Theta$  be the matrix containing the columns of  $\mathbf{X}$  corresponding to the  $m$  active constraints. Identifying the columns of  $\mathbf{A}$  requires the following additional steps: Calculate  $\mathbf{q}_i = \Theta^{-1}\mathbf{x}_i$  for  $i = 1, \dots, N$ . If  $\mathbf{q}_i \geq \mathbf{0}$  for  $i = 1, \dots, N$ , then  $\Theta$  corresponds to  $\mathbf{A}$ . However, if any element of  $\mathbf{q}_i$  is strictly less than 0 for any  $i = 1, \dots, N$ , then it indicates that the locally dominant assumption is invalid for the given  $\mathbf{X}$ . Thus, this serves as a test for the existence of the locally dominant assumption.

The above test works for typical data mixing without noise. However, the image data is usually integer data. There is always rounding, taking place at the source or mixture level. Therefore, the rounding effect needs to be incorporated into the above condition. It is proposed in this paper to follow the heuristic method to incorporate the rounding effect. Let  $\nu = 0.5|\Theta^{-1}\mathbf{e}|$ , where  $\mathbf{e} \in \mathbb{R}^n$  is a vector of all ones. The notion is that the rounding will create a maximum error of  $\pm 0.5$  in each pixel element. Thus, the maximum error in any pixel element will be strictly less than 0.5. Therefore, the check for image data will be as follows: If  $\mathbf{q}_i + \nu \geq \mathbf{0}$  for  $i = 1, \dots, N$ , then  $\Theta$  corresponds to  $\mathbf{A}$ .

Furthermore, the above idea can be extended for mixing scenarios containing noise. For instance, a level of tolerance can be used to analyze the noisy mixture data. Let  $\psi \geq \mathbf{0}, \in \mathbb{R}^m$  be a tolerance parameter selected by the user. Then, based on the earlier discussion, the check for noisy data will be as follows: If  $\mathbf{q}_i + \psi \geq \mathbf{0}$  for  $i = 1, \dots, N$ , then  $\Theta$  corresponds to  $\mathbf{A}$ . The precise value of  $\psi$  may not be available for a given scenario. Hence, the parameter will be empirically selected based on trial experiments. In Section 5, an experiment that highlights the usage of parameter  $\psi$  for noisy data is illustrated.

### 3. Point Correntropy

Correntropy is a generalized correlation measure based on the concept of entropy. It is typically used in detecting local similarity between two random variables. Roughly speaking, it is a kernelized version of the conventional correlation measure. The measure first appeared in [46,47], and its usage as a cost function was illustrated in [48–52]. The optimization properties of the cost function are presented in [53]. The correntropy cost function (or the correntropic loss) for  $N$  errors is defined as:

$$\mathcal{F}(\boldsymbol{\varepsilon}) = \beta \left( 1 - \frac{1}{N} \sum_{i=1}^N k(\varepsilon_i, \sigma) \right), \tag{23}$$

where  $\beta = \left[ 1 - e^{\left(\frac{-1}{2\sigma^2}\right)} \right]^{-1}$  is a scaling parameter,  $\boldsymbol{\varepsilon} \in \mathbb{R}^N$  is an array of errors, and  $k(\cdot)$  is the transformation kernel function with parameter  $\sigma$ . In this work, a Gaussian kernel is selected, i.e.,  $k(\varepsilon, \sigma) = e^{\left(\frac{-\varepsilon^2}{2\sigma^2}\right)}$ . Equation (23) is readily separable with respect to sample errors, and can be rewritten as:

$$\mathcal{F}(\boldsymbol{\varepsilon}) = \sum_{i=1}^N f(\varepsilon_i) \tag{24}$$

where  $f(\varepsilon_i) = \frac{\beta}{N} (1 - k(\varepsilon_i, \sigma))$ , and it will be referred to as point estimate of the correntropic loss.

Let  $\varepsilon_i$  be an error corresponding to the  $i$ th column vector of  $\mathbf{X}$ . For a given kernel parameter  $\sigma$ , the point estimate provides similarity information of the  $i$ th vector with respect to the other data vectors of  $\mathbf{X}$ . Based on the geometry of the vectors, the extreme vectors' similarity with respect to the central vectors should be typically less than the other non-extreme vectors of  $\mathbf{X}$ . Thus, the point estimate of the correntropic loss function can be used as a measure to differentiate extreme and non-extreme vectors of  $\mathbf{X}$ .

### 4. Solution Methodology

Our goal is to develop a geometric separation method for the non-negative data mixing problem, that can be applied to the “Big Data” scenarios. The concepts developed in Sections 2 and 3 are tailored with respect to big data, and the following solution approach is proposed. The summary of the proposed approach is illustrated in Figure 2 and Algorithm 1.

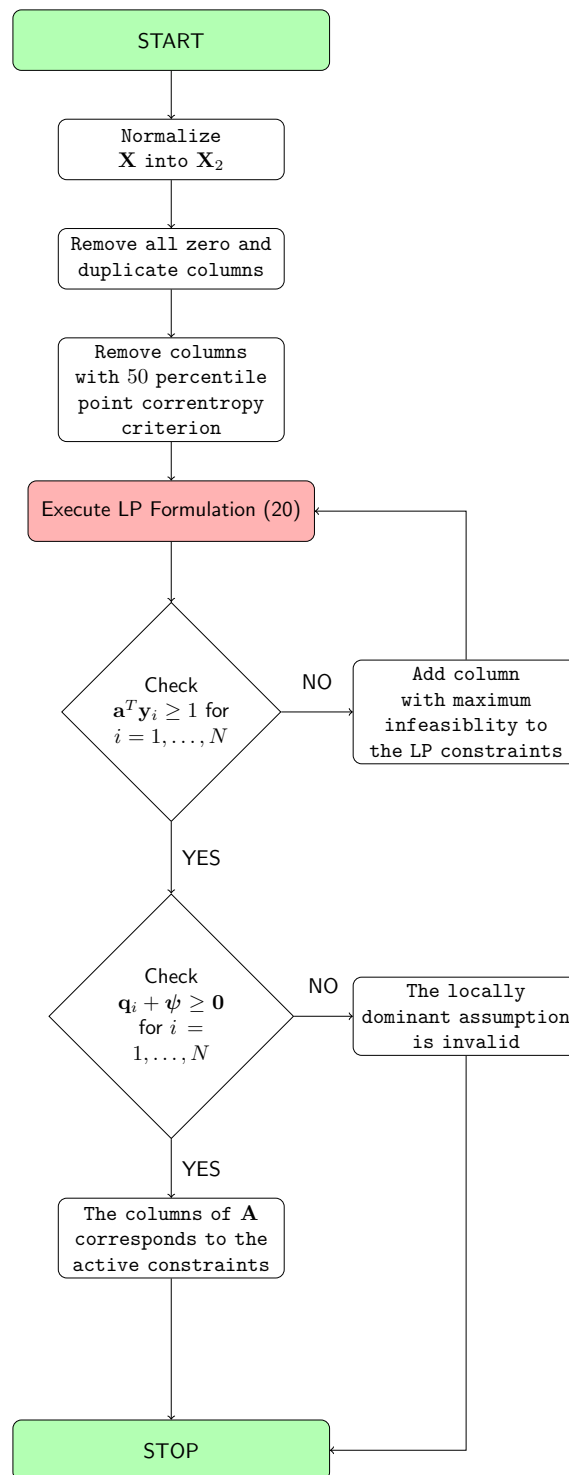


Figure 2. The proposed approach.



**Algorithm 1:** The Proposed Algorithm.

---

**Data:** Given  $\mathbf{X} \in \mathbb{R}^{m \times N}$   
**Result:** Find  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{S} \in \mathbb{R}_+^{n \times N}$  such that  $\mathbf{X} = \mathbf{AS}$   
 $\mathbf{X}_2 = \text{normalize}(\mathbf{X})$ ;  
Remove all zero columns and duplicate columns from  $\mathbf{X}_2$ , and say  $\mathbf{X}_2 \in \mathbb{R}^{m \times \hat{N}}$ ;  
Estimate  $\sigma$  from  $\mathbb{S}$ ;  
Obtain  $\mathbf{X}_R$  by removing all columns with the 50 percentile point correntropy criterion from  $\mathbf{X}_2$ ;  
Let  $\mathbf{X}_E = \mathbf{X}_2 \setminus \mathbf{X}_R$ ;  
Let  $\mathbf{y}_i$  be the  $i$ th column of  $\mathbf{X}_2$ ;  
 $\mathbf{a} =$  Solution of LP Formulation (20) with respect to data  $\mathbf{X}_R$ ;  
**while**  $\mathbf{a}^T \mathbf{y}_i < 1$  for  $i = 1, \dots, N$  **do**  
     $\mathbf{X}_R = \mathbf{X}_R \cup \mathbf{w}$ ;  
    where  $\mathbf{w} \in \mathbf{X}_E$  is the column with maximum infeasibility to the LP constraints;  
    Get  $\mathbf{a} =$  Solution LP Formulation (20) with respect to data  $\mathbf{X}_R$ ;  
**end**  
Let  $\Theta$  be the matrix containing the columns of  $\mathbf{X}$  corresponding to the active constraints at the optimal solution of Formulation (20).;  
Calculate  $\mathbf{q}_i = \Theta^{-1} \mathbf{x}_i$  for  $i = 1, \dots, N$ ;  
Set  $\psi$  equal to 0 for non-noisy non-image data mixing, equal to  $\nu$  for non-noisy image data mixing, or equal to the user-specified value for noisy mixing.;  
**if**  $\mathbf{q}_i + \psi \geq \mathbf{0} \quad \forall i = 1, \dots, N$  **then**  
     $\mathbf{A} = \Theta$ ;  
    Stop;  
**else**  
    Locally dominant assumption invalid;  
    Stop;  
**end**

---

**Data Ranking:** As seen earlier, the extreme vectors of  $\mathbf{X}_2$  contain all the relevant information that is needed for separation (i.e., identifying  $\mathbf{A}$  and  $\mathbf{S}$ ). Other data vectors are redundant in identifying the mixing matrix. Thus, the point estimate of the correntropic loss can be evaluated at all the data points with respect to the central columns of  $\mathbf{X}_2$ . Those data points that have low value of the correntropic loss can be removed from the data set  $\mathbf{X}_2$  without losing any information. The major issues in implementing the above idea are as follows: how to select the right value for  $\sigma$ , and how to define  $\varepsilon_i$  corresponding to  $\mathbf{y}_i$  for  $i = 1, \dots, N$ .

The value of  $\sigma$  represents the kernel width, and should be large enough to contain the central vectors. However, it should be small enough to exclude the extreme vectors. In the following, we propose a practical method to estimate the value of  $\sigma$ . Let  $\mathbb{S}$  be a sample of columns randomly selected from  $\mathbf{X}_2$ . Let  $\delta_j = \max_{\substack{\forall i1, i2 \in \mathbb{S} \\ i1 < i2}} \{y_{j,i1} - y_{j,i2}\}$ , where  $y_{j,i1}$  is the  $j$ th element of  $\mathbf{y}_{i1}$ . Based on the trial experiments, we found that  $\sigma = \frac{\sum_{j=1}^n \delta_j}{\sqrt{2n}}$  is a good choice for the kernel width. Furthermore, the value of  $\mathbf{c}$  should correspond to the center of the columns of  $\mathbf{X}_2$ . An approximate estimate for  $\mathbf{c}$  can be  $\mathbf{c} = \frac{\mathbf{g} + \mathbf{h}}{2}$ , where  $g_j = \max_j \{y_{j,i}\}$ ,  $h_j = \min_j \{y_{j,i}\}$  for  $i \in \mathbb{S}$ . Thus, simple (and practical for big data) estimate of error for  $\mathbf{y}_i$  will be  $\varepsilon_i = \mathbf{y}_i - \mathbf{c}$  for  $i = 1, \dots, N$ . Based on the trial experiments, it can be concluded that the larger the size of  $\mathbb{S}$ , the better the estimation. Furthermore, the strategy to eliminate the columns from the trial experiments is as follows. All the columns of  $\mathbf{X}_2$  that have the point estimate value lower than 50 percentile are removed from further consideration.

**Handling a Large Number of Constraints:** From Formulation (20), it can be seen that the big data corresponds to a large number of constraints. However, only  $m$  constraints are active, and the rest of the constraints are redundant (i.e., the rest of the constraints will never be active). The proposed data ranking method eliminates a good amount of the redundant constraints, depending upon the distribution of columns in the data set.

Let  $\mathbf{X}_R \subseteq \mathbf{X}_2$  be the data matrix obtained after eliminating possible central vectors, and let  $\mathbf{X}_E \subset \mathbf{X}_2$  be the eliminated central vectors. If the columns of  $\mathbf{X}_2$  are reorganized, then there exists a partition such that  $\mathbf{X}_2 = [\mathbf{X}_R | \mathbf{X}_E]$ . Once the tightest envelope is obtained for  $\mathbf{X}_R$  by solving Formulation (20), the envelope is validated with respect to  $\mathbf{X}_E$ . If all the columns of  $\mathbf{X}_E$  fall in the same half space (i.e., Equation (14) is feasible with respect to  $\mathbf{X}_2$ ), then it can be guaranteed that none of the extreme vectors of  $\mathbf{X}_2$  were eliminated. However, if there is any infeasibility detected, then the column of  $\mathbf{X}_E$  with maximum infeasibility is added to the LP, and the LP is resolved.

**Performance Index:** In order to evaluate the performance of the proposed approach, distance-based metrics will be used. Specifically, two metrics (one for the mixing matrix, and the other for the source matrix) will be used in this work. The following error measure is for the mixing matrix:

$$e_A = \frac{1}{n^2} \sum_{j=1}^n \left\| \frac{\mathbf{a}_j}{\|\mathbf{a}_j\|} - \frac{\hat{\mathbf{a}}_{[j]}}{\|\hat{\mathbf{a}}_{[j]}\|} \right\|_2 \quad (25)$$

where  $\mathbf{a}_j$  is the  $j$ th column of the original mixing matrix  $\mathbf{A}$ , and  $\hat{\mathbf{a}}_{[j]}$  is the corresponding column to  $\mathbf{a}_j$ , obtained from the recovered mixing matrix  $\hat{\mathbf{A}}$ . The corresponding columns are identified by the Hungarian algorithm [54]. The source matrix is obtained as follows:

$$\hat{\mathbf{S}} = \underset{\mathbf{S} \in \mathbb{R}_+^{n \times N}}{\operatorname{argmin}} \{ \|\mathbf{X} - \hat{\mathbf{A}}\mathbf{S}\|_2 \} \quad (26)$$

The above approach can be replaced by  $\hat{\mathbf{S}} = \hat{\mathbf{A}}^{-1}\mathbf{X}$ , whenever  $\hat{\mathbf{A}}^{-1}$  exists. Similar to  $e_A$ , the following error measure is for the source matrix:

$$e_S = \frac{1}{nN} \sum_{j=1}^n \left\| \frac{\mathbf{s}_{j\bullet}}{\|\mathbf{s}_{j\bullet}\|} - \frac{\hat{\mathbf{s}}_{[j\bullet]}}{\|\hat{\mathbf{s}}_{[j\bullet]}\|} \right\|_2 \quad (27)$$

where  $\mathbf{s}_{j\bullet}$  is the  $j$ th row of the original source matrix  $\mathbf{S}$ , and  $\hat{\mathbf{s}}_{[j\bullet]}$  is the corresponding row to  $\mathbf{s}_{j\bullet}$ , obtained from the recovered source matrix  $\hat{\mathbf{S}}$ .

## 5. Numerical Experiments

In order to illustrate the performance of the proposed approach, numerical experiments are presented. The experiments are divided into four groups. In the first group of experiments, simulated non-noisy data is used to test the performance and sensitivity of the proposed approach. In the second group of experiments, image data mixtures are used to test the applicability of the proposed approach on real image data. The third and fourth group of experiments compare the proposed approach with the well known SCA methods in the literature. In all the instances of this section, the following specifications were used:  $\mathbb{S} = \{1, \dots, N\}$ . All the random mixing matrices contain columns with unit norm. The LP resolving step is skipped in order to identify the number of instances in which the extreme vectors were eliminated. The LP was solved via the dual simplex method, using the state of the art Cplex 12.0 solver [55]. All the instances were solved on an Intel Xeon 2.4 GHz workstation, with 16 logical processors and 32 GB of RAM.

### 5.1. Simulated Data Separation

**Setup:** Given  $n$  and  $N$ , a random source matrix  $\mathbf{S}$  that satisfies the locally dominant assumption is generated. For the generated  $\mathbf{S}$  matrix, 100 random mixing matrices  $\mathbf{A}$  are generated. Then, using the  $\mathbf{X} = \mathbf{AS}$  equation, 100 random mixture matrices are generated. On each mixture matrix, the proposed approach is implemented. This study is executed for all the following combinations of  $n = 5, 7, 9$  and  $11$ , and  $N = 1000, 5000, 10,000, 50,000, 100,000, 500,000$  and  $1,000,000$ . In addition to that, this study is also executed for  $n = 10, 20, 40, 60, 80$  and  $100$  when  $N = 1,000,000$ . The value of  $\psi$  is set to zero in this experiment.

**Results:** Using one mixture matrix as input, and using the proposed approach, matrices  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{S}}$  are recovered. This experiment is repeated 100 times for a given combination of  $n$  and  $N$ . The performance of the proposed approach is displayed in Table 1. The column corresponding to **mErrA** (**vErrA**) indicates the mean (variance) of error  $e_A$  over the 100 instances. Similarly, columns **mErrS** and **vErrS** correspond to the mean and variance of error  $e_S$  respectively. The column corresponding to **mTime** (**vTime**) indicates the mean (variance) of the solution time per instance in seconds (milliseconds) over the 100 instances. In addition to that, the column corresponding to **mRed** (**vRed**) indicates the mean (variance) of the percentage of columns eliminated over the 100 instances. Finally, the column corresponding to **nMiss** indicates the number of times the extreme vectors were eliminated based on the 50 percentile criterion. Since the mixtures are clean (i.e., no noise is added to the mixture data), the recovery is perfect. This can be seen from the very low average error (**mErrA** and **mErrS**) over the 100 iterations. Furthermore, the method is consistent in the recovery of the matrices, and it can be justified from the low variance in the error (**vErrA** and **vErrS**). The suitability and applicability of the proposed approach to big data can be seen from the solution time. For instance, Figures 3 and 4 illustrate the average time in seconds required to solve one instance of the proposed approach for  $N$  data points and  $n$  data sources. The behavior of the solution time with respect to  $\log_{10}(N)$  is exponential. In other words, the solution time increases linearly with respect to  $N$ . Furthermore, from Figure 4, it can be observed that the solution time is linear with respect to  $n$ . Thus, the algorithm is suitable for big data scenarios. Moreover, the 50 percentile criterion removes exactly 50% of the data points in all the cases, with zero variance. This is due to the fact that the source matrices are uniformly randomly generated. Due to the uniform generation of the source matrices, none of the extreme vectors were eliminated.

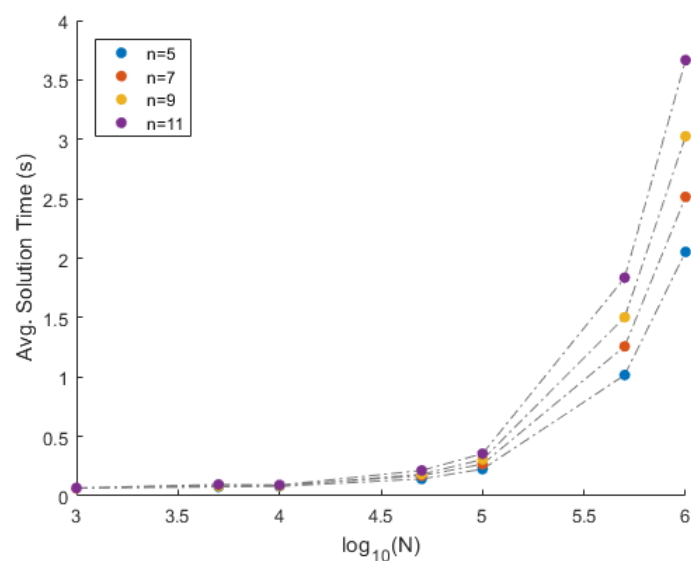
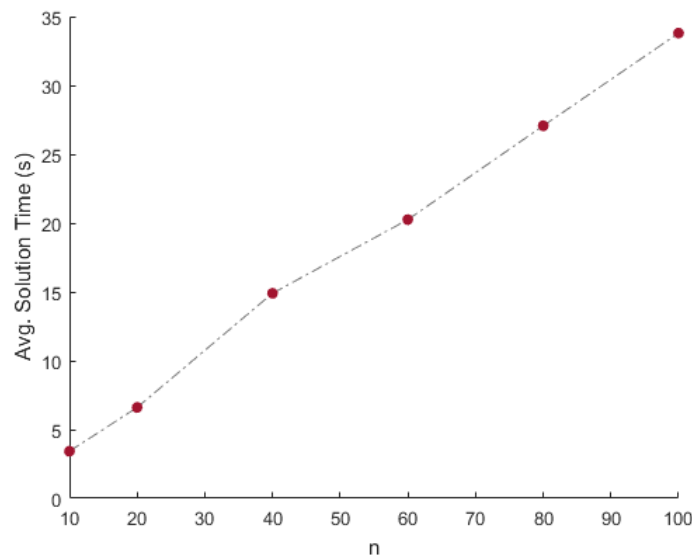


Figure 3. Time VS  $\log_{10}(N)$  comparison on the simulated data.

**Table 1.** Performance of the proposed approach on the simulated data.

$n \times N$	mErrA	vErrA	mErrS	vErrS	mTime	vTime	mRed	vRed	nMiss
5 × 1000	$1.04 \times 10^{-17}$	$9.78 \times 10^{-36}$	$1.17 \times 10^{-20}$	$4.76 \times 10^{-42}$	0.0647	0.07095	50	0	0
5 × 5000	$1.08 \times 10^{-17}$	$1.16 \times 10^{-35}$	$2.03 \times 10^{-21}$	$5.24 \times 10^{-43}$	0.0756	0.04114	50	0	0
5 × 10,000	$1.07 \times 10^{-17}$	$1.26 \times 10^{-35}$	$9.50 \times 10^{-22}$	$1.88 \times 10^{-43}$	0.08	0.16436	50	0	0
5 × 50,000	$1.10 \times 10^{-17}$	$1.13 \times 10^{-35}$	$5.10 \times 10^{-22}$	$5.39 \times 10^{-44}$	0.1427	0.09452	50	0	0
5 × 100,000	$1.02 \times 10^{-17}$	$8.22 \times 10^{-36}$	$4.85 \times 10^{-22}$	$3.21 \times 10^{-44}$	0.2254	0.12459	50	0	0
5 × 500,000	$1.01 \times 10^{-17}$	$9.28 \times 10^{-36}$	$1.81 \times 10^{-22}$	$6.50 \times 10^{-45}$	1.0166	2.2	50	0	0
5 × 1,000,000	$1.03 \times 10^{-17}$	$1.11 \times 10^{-35}$	$1.19 \times 10^{-24}$	$8.55 \times 10^{-50}$	2.0526	1.9	50	0	0
7 × 1000	$7.28 \times 10^{-18}$	$3.47 \times 10^{-36}$	$1.30 \times 10^{-20}$	$3.41 \times 10^{-42}$	0.0641	0.05835	50	0	0
7 × 5000	$6.85 \times 10^{-18}$	$2.72 \times 10^{-36}$	$1.71 \times 10^{-21}$	$1.08 \times 10^{-43}$	0.0816	0.06362	50	0	0
7 × 10,000	$6.86 \times 10^{-18}$	$2.28 \times 10^{-36}$	$1.01 \times 10^{-21}$	$7.70 \times 10^{-44}$	0.0891	0.11015	50	0	0
7 × 50,000	$7.08 \times 10^{-18}$	$1.89 \times 10^{-36}$	$7.05 \times 10^{-22}$	$4.27 \times 10^{-44}$	0.1689	0.14786	50	0	0
7 × 100,000	$6.78 \times 10^{-18}$	$2.46 \times 10^{-36}$	$1.14 \times 10^{-22}$	$5.67 \times 10^{-45}$	0.2675	0.14104	50	0	0
7 × 500,000	$7.26 \times 10^{-18}$	$2.61 \times 10^{-36}$	$6.11 \times 10^{-23}$	$1.80 \times 10^{-45}$	1.2584	1.3	50	0	0
7 × 1,000,000	$6.85 \times 10^{-18}$	$2.62 \times 10^{-36}$	$9.93 \times 10^{-23}$	$1.77 \times 10^{-45}$	2.5154	3.3	50	0	0
9 × 1000	$5.50 \times 10^{-18}$	$1.05 \times 10^{-36}$	$1.59 \times 10^{-20}$	$2.94 \times 10^{-42}$	0.067	0.10896	50	0	0
9 × 5000	$5.82 \times 10^{-18}$	$1.42 \times 10^{-36}$	$2.44 \times 10^{-21}$	$1.96 \times 10^{-43}$	0.0812	0.05954	50	0	0
9 × 10,000	$5.62 \times 10^{-18}$	$1.30 \times 10^{-36}$	$8.13 \times 10^{-22}$	$4.63 \times 10^{-44}$	0.0835	0.08635	50	0	0
9 × 50,000	$5.52 \times 10^{-18}$	$1.38 \times 10^{-36}$	$2.62 \times 10^{-22}$	$5.69 \times 10^{-45}$	0.1821	0.12421	50	0	0
9 × 100,000	$5.57 \times 10^{-18}$	$1.31 \times 10^{-36}$	$3.79 \times 10^{-22}$	$1.68 \times 10^{-44}$	0.3066	0.1855	50	0	0
9 × 500,000	$5.40 \times 10^{-18}$	$1.36 \times 10^{-36}$	$9.78 \times 10^{-23}$	$1.84 \times 10^{-45}$	1.5029	1	50	0	0
9 × 1,000,000	$5.32 \times 10^{-18}$	$1.21 \times 10^{-36}$	$1.49 \times 10^{-22}$	$2.10 \times 10^{-45}$	3.0258	3.1	50	0	0
11 × 1000	$4.05 \times 10^{-18}$	$4.77 \times 10^{-37}$	$1.75 \times 10^{-20}$	$3.13 \times 10^{-42}$	0.0672	0.02826	50	0	0
11 × 5000	$4.16 \times 10^{-18}$	$5.04 \times 10^{-37}$	$2.03 \times 10^{-21}$	$9.47 \times 10^{-44}$	0.096	0.10778	50	0	0
11 × 10,000	$4.21 \times 10^{-18}$	$4.87 \times 10^{-37}$	$1.31 \times 10^{-21}$	$7.33 \times 10^{-44}$	0.0916	0.00928	50	0	0
11 × 50,000	$4.09 \times 10^{-18}$	$5.02 \times 10^{-37}$	$6.61 \times 10^{-22}$	$2.97 \times 10^{-44}$	0.2137	0.05729	50	0	0
11 × 100,000	$4.12 \times 10^{-18}$	$3.79 \times 10^{-37}$	$1.90 \times 10^{-22}$	$6.92 \times 10^{-45}$	0.355	0.25758	50	0	0
11 × 500,000	$4.14 \times 10^{-18}$	$3.96 \times 10^{-37}$	$1.18 \times 10^{-22}$	$1.66 \times 10^{-45}$	1.8358	0.69888	50	0	0
11 × 1,000,000	$4.21 \times 10^{-18}$	$4.37 \times 10^{-37}$	$8.91 \times 10^{-23}$	$8.11 \times 10^{-46}$	3.667	5.4	50	0	0
10 × 1,000,000	$4.71 \times 10^{-18}$	$7.54 \times 10^{-37}$	$8.46 \times 10^{-23}$	$8.89 \times 10^{-46}$	3.4245	2.5	50	0	0
20 × 1,000,000	$1.83 \times 10^{-18}$	$5.10 \times 10^{-38}$	$1.12 \times 10^{-22}$	$6.32 \times 10^{-46}$	6.6053	9.3	50	0	0
40 × 1,000,000	$7.84 \times 10^{-19}$	$5.33 \times 10^{-39}$	$4.22 \times 10^{-23}$	$9.61 \times 10^{-47}$	14.8988	64.8	50	0	0
60 × 1,000,000	$7.32 \times 10^{-19}$	$4.62 \times 10^{-38}$	$1.05 \times 10^{-22}$	$2.18 \times 10^{-46}$	20.2509	628.9	50	0	0
80 × 1,000,000	$6.51 \times 10^{-19}$	$6.39 \times 10^{-39}$	$7.51 \times 10^{-23}$	$1.40 \times 10^{-46}$	27.0654	101.7	50	0	0
100 × 1,000,000	$5.13 \times 10^{-19}$	$2.83 \times 10^{-39}$	$4.65 \times 10^{-23}$	$5.57 \times 10^{-47}$	33.7978	136.6	50	0	0



**Figure 4.** Time vs.  $n$  comparison on the simulated data.

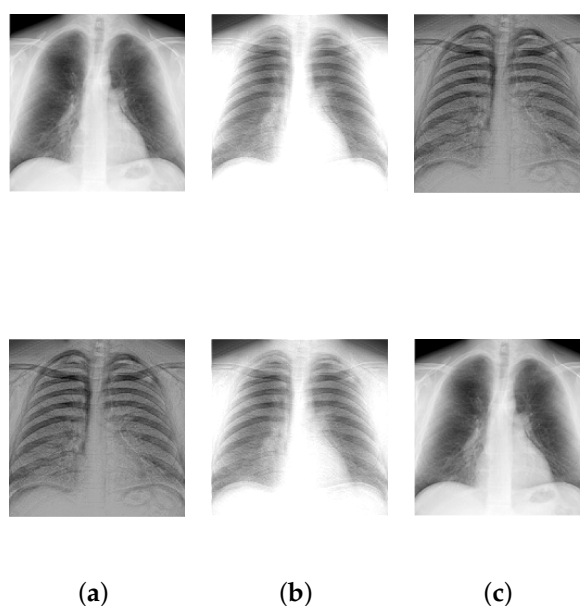
## 5.2. Image Mixture Separation

**Setup:** In the following experiments, image data available from the literature and online repositories are considered (see Table 2). Each source image of an image set is reshaped into one row vector. Then, the reshaped images are row-wise stacked together to generate the  $\mathbf{S}$  matrix. The source matrices are pre-processed in order to satisfy the locally dominant assumption. Next, for each source matrix  $\mathbf{S}$ , 100 random  $\mathbf{A}$  matrices are generated, and correspondingly 100 random  $\mathbf{X}$  matrices are analyzed using the proposed approach. Table 2 summarizes the details of the image sets that are considered in this subsection. The first column conveys the name of the image set that is being considered. The column corresponding to  $\mathbf{n}$  indicates the total number of sources, and the column corresponding to  $\mathbf{N}$  indicates the total data points (or column vectors) in  $\mathbf{X}$ . The value of  $\psi$  is set to  $\nu$  in this experiment.

**Table 2.** Image Mixture Separation Data.

Image Set	$\mathbf{n}$	$\mathbf{N}$
Chest X-rays	2	26,896
Scenery	3	65,536
CT Scans	5	16,384
Zip Codes	7	12,672
Finger Print	9	90,000

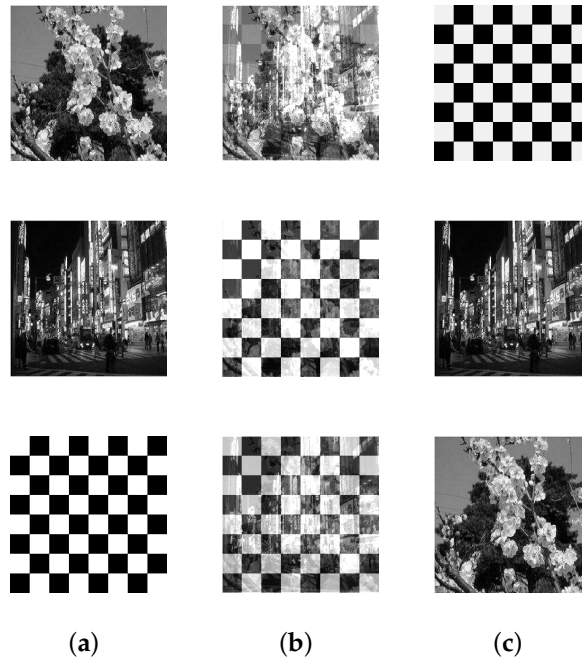
**Results:** Table 3 displays the results after executing the proposed approach on the 100 mixture instances of each image set. The columns have the notation similar to the earlier experiment, except that the  $\nu$ Time column units are in seconds. Moreover, Figures 5–9 depict the results. Low mean errors ( $e_A$  and  $e_S$ ) over the 100 runs are obtained for all the image sets. This shows that the method precisely recovers  $\mathbf{A}$  and  $\mathbf{S}$  matrices. A low value in the corresponding variance column indicates the high level of consistency of the proposed approach. The solution time, specifically for the finger print data set, indicates the applicability of the proposed approach for big data with complex image mixing scenarios. Based on the results, it can be seen that the 50 percentile criterion eliminates a good amount (more than 50%) of the redundant columns. However, in some instances (at most 7 percentage in one instance), the criterion eliminated some of the extreme vectors.



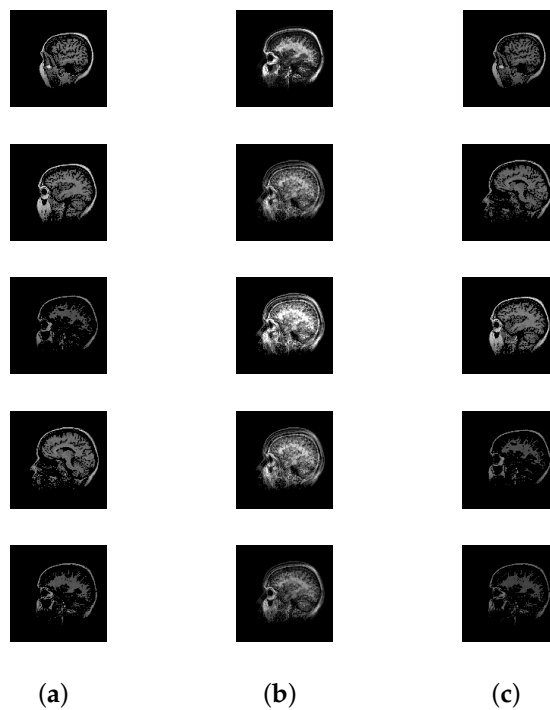
**Figure 5.** Mixing and unmixing of chest X-rays. (a) Original; (b) Mixture; (c) Recovered.

**Table 3.** Image mixture separation results.

Image Set	mErrA	vErrA	mErrS	vErrS	mTime	vTime	mRed	vRed	nMiss
Chest X-rays	$2.45 \times 10^{-17}$	$2.81 \times 10^{-34}$	$1.33 \times 10^{-22}$	$2.02 \times 10^{-44}$	0.0755	$4.89 \times 10^{-5}$	81.632	0.0101	0
Scenery	$2.06 \times 10^{-17}$	$5.30 \times 10^{-35}$	$4.10 \times 10^{-22}$	$4.12 \times 10^{-44}$	0.109	$9.81 \times 10^{-5}$	73.2417	0.0033	7
CT Scan	$1.19 \times 10^{-17}$	$9.09 \times 10^{-36}$	$4.80 \times 10^{-22}$	$1.20 \times 10^{-44}$	0.0679	$1.34 \times 10^{-4}$	89.7026	0.0012	4
Zip Codes	$7.36 \times 10^{-18}$	$2.61 \times 10^{-36}$	$4.96 \times 10^{-22}$	$8.17 \times 10^{-45}$	0.0787	$3.74 \times 10^{-5}$	74.0513	0.0047	6
Finger Print	$5.72 \times 10^{-18}$	$1.19 \times 10^{-36}$	$1.67 \times 10^{-22}$	$2.23 \times 10^{-45}$	0.2716	$1.57 \times 10^{-4}$	55.952	$6.49 \times 10^{-6}$	0



**Figure 6.** Mixing and unmixing of scenery. (a) Original; (b) Mixture; (c) Recovered.



**Figure 7.** Mixing and unmixing of a CT scan. (a) Original; (b) Mixture; (c) Recovered.



Figure 8. Mixing and unmixing of zip codes. (a) Original; (b) Mixture; (c) Recovered.

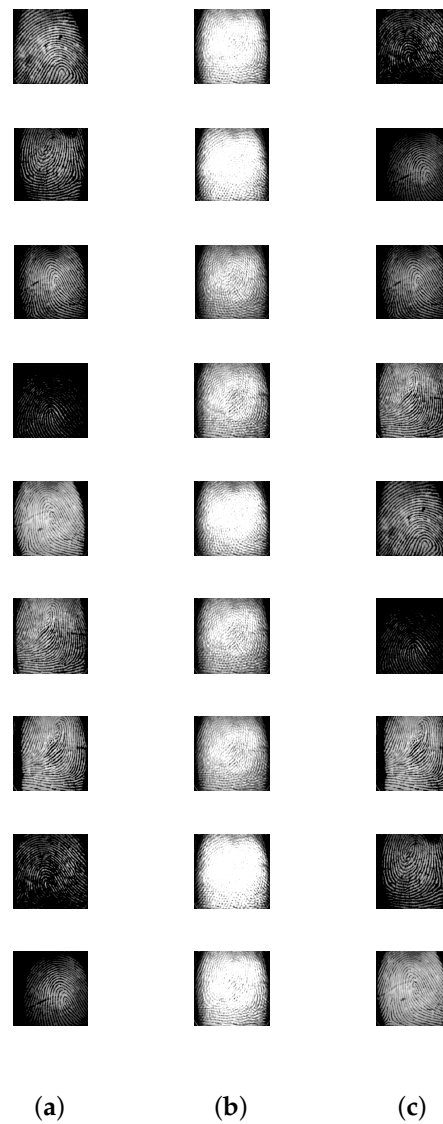


Figure 9. Mixing and unmixing of a finger print. (a) Original; (b) Mixture; (c) Recovered.

### 5.3. Comparative Experiment-I

**Setup:** Given  $n$  and  $N$ , a random source matrix  $\mathbf{S}$  that does not satisfy the locally dominant assumption is generated in this experiment. For the generated  $\mathbf{S}$  matrix, 100 random mixing matrices  $\mathbf{A}$  are generated. Then, using the  $\mathbf{X} = \mathbf{AS}$  equation, 100 random mixture matrices are generated. This study is executed for  $n = 5, 7, 9, 11, 13$  and  $15$ , and for  $N = 10,000$ . The value of  $\psi$  is set to zero in this experiment. Well known methods from the SCA literature are compared with the proposed approach. The methods that are used for the comparison are N-FINDR ([44]), VCA ([45]), MVSA ([56]). The objective of this experiment is to highlight that the above three (like the other typical algorithms in the literature) do not have the capability to test the locally dominant assumption from the knowledge of  $\mathbf{X}$ . To the best of my knowledge, only an exhaustive search similar to the one presented in [24] can do such a test. However, the proposed approach does not require such an exhaustive search.

**Results:** Table 4 displays the results after executing the proposed and selected approaches on the 100 randomly generated mixture instances. The column corresponding to **mErrA** (**vErrA**) indicates the mean (variance) of error  $e_A$  over the 100 instances for the three methods used from the literature. The lines in the column corresponding to **ErrA** indicate that the proposed approach was unable to identify any mixing matrix. The reason for the lines is the non-existence of the locally dominant assumption in the source data. This information is captured in the column corresponding to **TnMiss**. The numbers in the **TnMiss** column indicate the total number of times the proposed approach exited with the “no locally dominant sources” token. Based on the results, it can be seen that the other algorithms try to find the best match for the columns of  $\mathbf{A}$ . However, they are unable to validate the locally dominant assumption. This is due to the fact that no such test is available in the literature. However, in all the scenarios, the proposed approach was able to conclude that the input data is not a mixture of sources that contain the locally dominant signals.

**Table 4.** Comparative experiment-I separation results.

n	N	VCA		MVSA		N-FINDR		Proposed	
		mErrA	vErrA	mErrA	vErrA	mErrA	vErrA	ErrA	TnMiss
5	10,000	0.0755	$7.08 \times 10^{-5}$	0.0813	$6.65 \times 10^{-5}$	0.0905	$4.14 \times 10^{-5}$	—	100
7	10,000	0.056	$9.63 \times 10^{-6}$	0.0567	$2.6 \times 10^{-5}$	0.0604	$7.23 \times 10^{-6}$	—	100
9	10,000	0.0422	$2.96 \times 10^{-6}$	0.0402	$1.13 \times 10^{-5}$	0.0441	$1.34 \times 10^{-6}$	—	100
11	10,000	0.0333	$8.56 \times 10^{-7}$	0.0314	$4.16 \times 10^{-6}$	0.0342	$7.16 \times 10^{-7}$	—	100
13	10,000	0.0269	$3.28 \times 10^{-7}$	0.0252	$1.56 \times 10^{-6}$	0.0274	$3.6 \times 10^{-7}$	—	100
15	10,000	0.0223	$1.52 \times 10^{-7}$	0.0212	$7.53 \times 10^{-7}$	0.0226	$1.8 \times 10^{-7}$	—	100

### 5.4. Comparative Experiment-II

**Setup:** In this experiment, a random source matrix  $\mathbf{S}$  that satisfies the locally dominant assumption, is generated. For the generated  $\mathbf{S}$  matrix, 100 random mixing matrices  $\mathbf{A}$  are generated. Then, using the  $\mathbf{X} = \mathbf{AS}$  equation, 100 random mixture matrices are generated. In each mixture matrix, 5% of the columns are randomly selected, and a uniform noise between 0 and 0.01 is added to all the elements of the selected columns. This study is executed for  $n = 5, 7, 9, 11, 13$  and  $15$ , and for  $N = 10,000$ . Well known methods from the SCA literature are compared with the proposed approach. The methods that are used for the comparison are N-FINDR ([44]), VCA ([45]), MVSA ([56]). The objective of this experiment is to comparatively assess the performance of the proposed and selected methods in noisy data. In this experiment, the value of  $\psi$  is defined as follows:  $\psi = \rho |\Theta^{-1} \mathbf{e}|$ , where  $\mathbf{e} \in \mathbb{R}^n$  is a vector of all ones, and  $\rho$  takes the following values:  $0, 0.2, 0.4, \dots, 1$ .

**Results:** Table 5 displays the results after executing the proposed and selected approaches on the 100 randomly generated noisy mixture instances. The columns corresponding to VCA, MVSA and N-FINDR present the average error  $e_A$  over the 100 instances. The proposed approach is executed 100 times for each value of  $\rho$ , and the average error  $e_A$  for each value of  $\rho$  is archived. The column



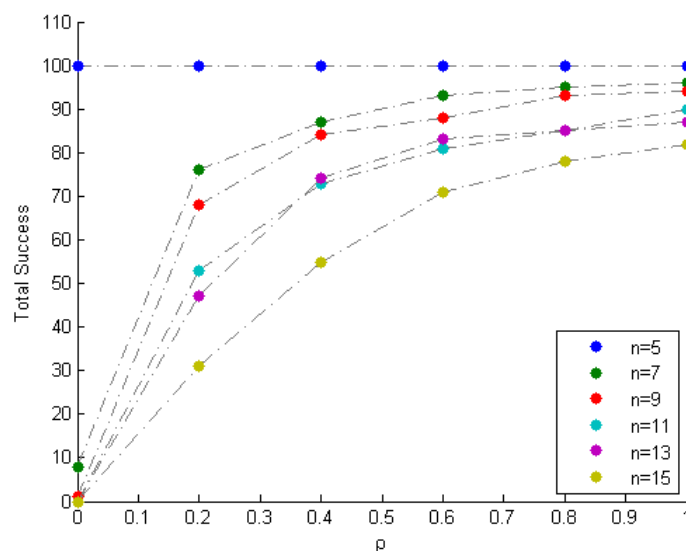
corresponding to the proposed approach presents the best of the average errors  $e_A$  over the values of  $\rho$ . Table 6 (Figure 10) indicates the total number of times the method fails (succeeds) to identify the mixing matrix, for various values of  $\rho$ . Based on the low value of error in the proposed column, and the trends depicted in Figure 10, it can be seen that the proposed approach recovers **A** and **S** in the majority of the noisy instances for higher values of  $\rho$ . Moreover, as  $n$  increases, the complexity of mixing increases, and thus the proposed approach requires a higher value of  $\rho$  for the recovery of **A** and **S**.

**Table 5.** Comparative experiment-II separation results.

n	N	VCA	MVSA	N-FINDR	Proposed
5	10,000	0.0798	0.0872	0.0954	$6.43 \times 10^{-10}$
7	10,000	0.0573	0.045	0.0658	$4.73 \times 10^{-10}$
9	10,000	0.0407	0.0309	0.0467	$3.77 \times 10^{-10}$
11	10,000	0.0321	0.0265	0.0357	0.0012
13	10,000	0.0255	0.0231	0.0284	0.0007
15	10,000	0.0212	0.0197	0.0235	0.0021

**Table 6.** Effect of parameter  $\rho$  on the total number of failures.

n	N	$\rho = 0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 1$
5	10,000	0	0	0	0	0	0
7	10,000	92	24	13	7	5	4
9	10,000	99	32	16	12	7	6
11	10,000	100	47	27	19	15	10
13	10,000	100	53	26	17	15	13
15	10,000	100	69	45	29	22	18



**Figure 10.** Total success vs.  $\rho$  (the tolerance parameter).

### 6. Discussion and Conclusions

The SCA approaches are relatively new to the BSS problem when compared to the ICA approaches. The main critique that often appears with respect to locally dominant SCA approaches is the validity of the locally dominant criterion. In this paper, a mathematical modeling-based approach is proposed that can validate the existence of the locally dominant criterion from the given mixture matrix. That is,

the formulation can be used not only to identify the mixing matrix, but also to validate the assumption presented in Equation (2). Although the approach is proposed for the determined case, it can also be applied to the overdetermined cases. The columns of the matrix  $X$  are proportional to the number of constraints in the proposed LP. Thus, big data often leads to LP with many redundant constraints. We propose the usage of interior point methods, when the total number of constraints is very high [57]. Moreover, LP decomposition-based approaches for SCA can also be developed to improve the solution time [58]. In addition to that, the LP presolve theory [59] can be designed to eliminate the redundant constraints in the proposed SCA approach. Roughly speaking, the point correntropic ranking method may be seen as a novel probabilistic approach for removing LP redundant constraints. The proposed method of estimating the point correntropy is computationally cheap, and can be applied to the big data scenarios. From the simulated data study, it can be seen that if the input data is uniformly distributed, then the 50 percentile criterion can be raised to a higher value. However, from the image data set, it is clear that the real world data is rarely uniformly distributed. Thus, the 50 percentile criterion is a good estimate to avoid the LP resolving. From the comparative experiments, it can be concluded that the proposed approach validates the locally dominant assumption in non-noisy and noisy mixing scenarios. To summarize, the proposed approach provides new insights into the BSS problem.

**Acknowledgments:** The author would like to acknowledge the support provided by Deanship of Research (DSR) at King Fahd University of Petroleum & Minerals (KFUPM) for funding this work through Junior Faculty Grant No. JF141010.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Joho, M.; Mathis, H.; Lambert, R.H. Overdetermined blind source separation: Using more sensors than source signals in a noisy mixture. In Proceedings of the Independent Component Analysis and Blind Signal Separation, Helsinki, Finland, 19–22 June 2000; pp. 81–86.
2. Winter, S.; Sawada, H.; Makino, S. Geometrical interpretation of the PCA subspace approach for overdetermined blind source separation. *EURASIP J. Adv. Signal Process.* **2006**, *2006*, 071632.
3. Bofill, P.; Zibulevsky, M. Underdetermined blind source separation using sparse representations. *Signal Process.* **2001**, *81*, 2353–2362.
4. Zhen, L.; Peng, D.; Yi, Z.; Xiang, Y.; Chen, P. Underdetermined blind source separation using sparse coding. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 3102–3108.
5. Herault, J.; Jutten, C.; Ans, B. Detection de Grandeurs Primitives dans un Message Composite par une Architecture de Calcul Neuromimetique en Apprentissage Non Supervise. In 1985—GRETSI—Actes de Colloques; Groupe d'Etudes du Traitement du Signal et des Images: Juan-les-Pins, France, 1985; pp. 1017–1022.
6. Syed, M.; Georgiev, P.; Pardalos, P. A hierarchical approach for sparse source blind signal separation problem. *Comput. Oper. Res.* **2012**, *41*, 386–398.
7. Hyvärinen, A.; Karhunen, J.; Oja, E. *Independent Component Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2004; Volume 46.
8. Amari, S.I.; Cichocki, A.; Yang, H.H. A new learning algorithm for blind signal separation. In Proceedings of the 8th International Conference on Neural Information Processing Systems, Denver, CO, USA, 27 November–2 December 1996; pp. 757–763.
9. Comon, P. Independent component analysis, a new concept? *Signal Process.* **1994**, *36*, 287–314.
10. Hyvärinen, A. New approximations of differential entropy for independent component analysis and projection pursuit. In Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems, Denver, CO, USA, 1–6 December 1998; pp. 273–279.
11. Bell, A.J.; Sejnowski, T.J. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* **1995**, *7*, 1129–1159.
12. Chai, R.; Naik, G.R.; Nguyen, T.N.; Ling, S.H.; Tran, Y.; Craig, A.; Nguyen, H.T. Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 715–724.

13. Naik, G.R.; Baker, K.G.; Nguyen, H.T. Dependence independence measure for posterior and anterior EMG sensors used in simple and complex finger flexion movements: Evaluation using SDICA. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 1689–1696.
14. Naik, G.R.; Al-Timemy, A.H.; Nguyen, H.T. Transradial amputee gesture classification using an optimal number of sEMG sensors: An approach using ICA clustering. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *24*, 837–846.
15. Deslauriers, J.; Ansado, J.; Marrelec, G.; Provost, J.S.; Joannette, Y. Increase of posterior connectivity in aging within the Ventral Attention Network: A functional connectivity analysis using independent component analysis. *Brain Res.* **2017**, *1657*, 288–296.
16. O’Muircheartaigh, J.; Jbabdi, S. Concurrent white matter bundles and grey matter networks using independent component analysis. *NeuroImage* **2017**, doi:10.1016/j.neuroimage.2017.05.012.
17. Hand, B.N.; Dennis, S.; Lane, A.E. Latent constructs underlying sensory subtypes in children with autism: A preliminary study. *Autism Res.* **2017**, *10*, 1364–1371.
18. Arya, Y. AGC performance enrichment of multi-source hydrothermal gas power systems using new optimized FOPPID controller and redox flow batteries. *Energy* **2017**, *127*, 704–715.
19. Al-Ali, A.K.H.; Senadji, B.; Naik, G.R. Enhanced forensic speaker verification using multi-run ICA in the presence of environmental noise and reverberation conditions. In Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuching, Malaysia, 12–14 September 2017; pp. 174–179.
20. Comon, P.; Jutten, C. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*; Academic Press: Cambridge, MA, USA, 2010.
21. Nascimento, J.M.; Dias, J.M. Does independent component analysis play a role in unmixing hyperspectral data? *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 175–187.
22. Syed, M.; Georgiev, P.; Pardalos, P. Robust Physiological Mappings: From Non-Invasive to Invasive. *Cybern. Syst. Anal.* **2015**, *1*, 96–104.
23. Georgiev, P.; Theis, F.; Cichocki, A.; Bakardjian, H. Sparse component analysis: A new tool for data mining. In *Data Mining Biomedicine*; Springer: Boston, MA, USA, 2007; pp. 91–116.
24. Naanaa, W.; Nuzillard, J. Blind source separation of positive and partially correlated data. *Signal Process.* **2005**, *85*, 1711–1722.
25. Chan, T.H.; Ma, W.K.; Chi, C.Y.; Wang, Y. A convex analysis framework for blind separation of non-negative sources. *IEEE Trans. Signal Process.* **2008**, *56*, 5120–5134.
26. Chan, T.H.; Chi, C.Y.; Huang, Y.M.; Ma, W.K. A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing. *IEEE Trans. Signal Process.* **2009**, *57*, 4418–4432.
27. Syed, M.N.; Georgiev, P.G.; Pardalos, P.M. Blind Signal Separation Methods in Computational Neuroscience. In *Modern Electroencephalographic Assessment Techniques: Theory and Applications*; Humana Press: New York, NY, USA, 2015; pp. 291–322.
28. Ma, W.K.; Bioucas-Dias, J.M.; Chan, T.H.; Gillis, N.; Gader, P.; Plaza, A.J.; Ambikapathi, A.; Chi, C.Y. A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Process. Mag.* **2014**, *31*, 67–81.
29. Bioucas-Dias, J.M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 354–379.
30. Yin, P.; Sun, Y.; Xin, J. A geometric blind source separation method based on facet component analysis. *Signal Image Video Process.* **2016**, *10*, 19–28.
31. Lin, C.H.; Chi, C.Y.; Wang, Y.H.; Chan, T.H. A Fast Hyperplane-Based Minimum-Volume Enclosing Simplex Algorithm for Blind Hyperspectral Unmixing. *IEEE Trans. Signal Process.* **2016**, *64*, 1946–1961.
32. Zhang, S.; Agathos, A.; Li, J. Robust Minimum Volume Simplex Analysis for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6431–6439.
33. Naanaa, W.; Nuzillard, J.M. Extreme direction analysis for blind separation of nonnegative signals. *Signal Process.* **2017**, *130*, 254–267.
34. Sun, Y.; Ridge, C.; Del Rio, F.; Shaka, A.; Xin, J. Postprocessing and sparse blind source separation of positive and partially overlapped data. *Signal Process.* **2011**, *91*, 1838–1851.
35. Aharon, M.; Elad, M.; Bruckstein, A. On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them. *Linear Algebra Appl.* **2006**, *416*, 48–67.

36. Georgiev, P.; Theis, F.; Ralescu, A. Identifiability conditions and subspace clustering in sparse BSS. In *Independent Component Analysis and Signal Separation*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 357–364.
37. Drumetz, L.; Veganzones, M.A.; Henrot, S.; Phlypo, R.; Chanussot, J.; Jutten, C. Blind hyperspectral unmixing using an Extended Linear Mixing Model to address spectral variability. *IEEE Trans. Image Process.* **2016**, *25*, 3890–3905.
38. Amini, F.; Hedayati, Y. Underdetermined blind modal identification of structures by earthquake and ambient vibration measurements via sparse component analysis. *J. Sound Vib.* **2016**, *366*, 117–132.
39. Gribonval, R.; Schnass, K. Dictionary Identification—Sparse Matrix-Factorization via  $l_1$ -Minimization. *IEEE Trans. Inf. Theory* **2010**, *56*, 3523–3539.
40. Kreutz-Delgado, K.; Murray, J.; Rao, B.; Engan, K.; Lee, T.; Sejnowski, T. Dictionary learning algorithms for sparse representation. *Neural Comput.* **2003**, *15*, 349–396.
41. Nascimento, J.M.; Bioucas-Dias, J.M. Blind hyperspectral unmixing. In Proceedings of the SPIE Conference on Image and Signal Processing for Remote Sensing XIII, Florence, Italy, 18–20 September 2007.
42. Duarte, L.T.; Moussaoui, S.; Jutten, C. Source separation in chemical analysis: Recent achievements and perspectives. *IEEE Signal Process. Mag.* **2014**, *31*, 135–146.
43. Sun, Y.; Xin, J. Nonnegative Sparse Blind Source Separation for NMR Spectroscopy by Data Clustering, Model Reduction, and  $l_1$  Minimization. *SIAM J. Imag. Sci.* **2012**, *5*, 886–911.
44. Winter, M.E. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In Proceedings of the Imaging Spectrometry V, Denver, CO, USA, 19–21 July 1999; pp. 266–275.
45. Nascimento, J.M.; Dias, J.M. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 898–910.
46. Santamaria, I.; Pokharel, P.P.; Principe, J.C. Generalized correlation function: Definition, properties, and application to blind equalization. *IEEE Trans. Signal Process.* **2006**, *54*, 2187–2197.
47. Liu, W.; Pokharel, P.P.; Principe, J.C. Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Trans. Signal Process.* **2007**, *55*, 5286–5298.
48. Singh, A.; Principe, J.C. Using correntropy as a cost function in linear adaptive filters. In Proceedings of the 2009 International Joint Conference on Neural Networks, Atlanta, GA, USA, 14–19 June 2009; pp. 2950–2955.
49. Zhao, S.; Chen, B.; Principe, J.C. Kernel adaptive filtering with maximum correntropy criterion. In Proceedings of the 2011 International Joint Conference on Neural Networks, San Jose, CA, USA, 31 July–5 August 2011; pp. 2012–2017.
50. Chen, B.; Xing, L.; Liang, J.; Zheng, N.; Principe, J.C. Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion. *IEEE Signal Process. Lett.* **2014**, *21*, 880–884.
51. Chen, B.; Xing, L.; Zhao, H.; Zheng, N.; Principe, J.C. Generalized correntropy for robust adaptive filtering. *IEEE Trans. Signal Process.* **2016**, *64*, 3376–3387.
52. Chen, B.; Liu, X.; Zhao, H.; Principe, J.C. Maximum correntropy Kalman filter. *Automatica* **2017**, *76*, 70–77.
53. Syed, M.N.; Pardalos, P.M.; Principe, J.C. On the optimization properties of the correntropic loss function in data analysis. *Optim. Lett.* **2014**, *8*, 823–839.
54. Kuhn, H.W. The hungarian metRhod for the assignment problem. In *50 Years of Integer Programming 1958–2008*; Springer: Berlin, Germany, 2010; pp. 29–47.
55. Cplex, I. *User-Manual CPLEX*; IBM Software Group: New York, NY, USA, 2011; Volume 12.
56. Li, J.; Bioucas-Dias, J.M. Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data. In Proceedings of the 2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, 7–11 July 2008; Volume 3, doi:10.1109/IGARSS.2008.4779330.
57. Terlaky, T. *Interior Point Methods of Mathematical Programming*; Springer Science & Business Media: Berlin, Germany, 2013; Volume 5.
58. Dantzig, G.B.; Thapa, M.N. *Linear Programming 2: Theory and Extensions*; Springer Science & Business Media: Berlin, Germany, 2006.
59. Andersen, E.D.; Andersen, K.D. Presolving in linear programming. *Math. Program.* **1995**, *71*, 221–245.

