Signal Processing 96 (2014) 332–345

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

A sparse semi-blind source identification method and its application to Raman spectroscopy for explosives detection

Yuanchang Sun [a,*], Jack Xin [b]

[a] Department of Mathematics and Statistics, Florida International University, Miami, FL 33199, United States
[b] Department of Mathematics, University of California at Irvine, Irvine, CA 92967, United States

ARTICLE INFO

Article history:
Received 12 August 2012
Received in revised form
16 September 2013
Accepted 23 September 2013
Available online 5 October 2013

Keywords:
Semi-blind source separation
Sparsity
Raman spectroscopy
Explosives detection

ABSTRACT

Rapid and reliable detection and identification of unknown chemical substances are critical to homeland security. It is challenging to identify chemical components from a wide range of explosives. There are two key steps involved. One is a non-destructive and informative spectroscopic technique for data acquisition. The other is an associated library of reference features along with a computational method for feature matching and meaningful detection within or beyond the library.

In this paper, we develop a new iterative method to identify unknown substances from mixture samples of Raman spectroscopy. In the first step, a constrained least squares method decomposes the data into a sum of linear combination of the known components and a non-negative residual. In the second step, a sparse and convex blind source separation method extracts components geometrically from the residuals. Verification based on the library templates or expert knowledge helps to confirm these components. If necessary, the confirmed meaningful components are fed back into step one to refine the residual and then step two extracts possibly more hidden components. The two steps may be iterated until no more components can be identified. We illustrate the proposed method in processing a set of the so called swept wavelength optical resonant Raman spectroscopy experimental data by a satisfactory blind extraction of a priori unknown chemical explosives from mixture samples. We also test the method on nuclear magnetic resonance (NMR) spectra for chemical compounds identification.

Published by Elsevier B.V.

1. Introduction

A critical problem in homeland security is reliable and rapid identification of unknown chemical and biological substances in the explosives. Due to the harmful environment caused by the release of the explosive chemicals, non-destructive spectroscopic techniques are typically used to record the optical spectrum without interfering with the samples. Ideally, a standoff detection is performed to acquire the spectral information. Then a search and matching procedure through a prepared spectral database will be carried out for identification. However such an approach would be unsuccessful if the explosives contained chemical compounds outside of the database, which is highly likely as hidden explosives are often unknown a priori. In general, the samples may involve multiple unknown substances besides impurities. Conventional analysis routines are mostly based on least squares fitting whose residuals could remain mysterious. Further analysis calls for the development of blind identification methods to extract major components from the residuals.

Various recent experimental techniques (see [3,7,8] and the references therein) can identify pure chemicals with notable success. These methods are mainly based on Raman spectroscopy, a spectroscopic technique to study the chemical

* Corresponding author. Tel.: +1 3053484959.
E-mail addresses: yuanchang.sun@fiu.edu (Y. Sun), jxin@math.uci.edu (J. Xin).

0165-1684/$ - see front matter Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.sigpro.2013.09.010

composition of the samples [16,28]. Combined with other spectroscopic techniques such as Ultraviolet and Infrared spectroscopy, Raman spectroscopy has been widely used in materials science, biosciences, geosciences (gemology), forensic sciences, nano-technology, and pharmaceutical chemistry [3,5,6,8,19,21,23,30]. For example, the Swept Wavelength Optical Resonant Raman Detector (SWOrRD) at the Naval Research Laboratory developed in 2009 can generate two-dimensional spectral maps of biological agents and chemical substances. The resultant two-dimensional signatures contain much more information than the single illumination wavelengths, which may result in a greater likelihood of successful identification even in complex mixtures [3,8]. Raman spectroscopy is based on Raman scattering, an inelastic scattering process that shifts the frequencies of the incident photons. During the interaction of the incident light with a molecule, a scattered light of lower (Stokes Raman) or higher (anti-Stokes Raman) energy is emitted, allowing the measurements of the molecule's vibrational modes. The appeals of this technique are non-destructiveness and fast sensing capability. It has become a promising tool for standoff distance detection for explosives at airports among other transportation centers.

In general, Raman spectra of a sample are composed of many substances, and they must be identified by an analysis software. If one has the complete knowledge of the kinds of substances in the sample, the least squares fitting can be used to retrieve their concentrations (or volumes) by a linear combination of the known spectra on a template. In most practical situations, one may have to identify the substances and quantify their concentrations at the same time. This becomes a blind source separation (BSS) problem, or recovering pure signal sources from their mixtures without a detailed knowledge of the mixing process. There have been several studies on the BSS of Raman spectra [17,18,22] based on independent component analysis (ICA, [11]) and nonnegative matrix factorization (NMF, [13]). However these methods are non-convex and too general to be robust and reliable in real-world applications. The independence hypothesis of ICA does not hold if chemicals share some common structures and have correlated Raman spectral line shapes. Moreover, these existing approaches do not address how to identify unknown substances from the fitting residuals when partial knowledge of the source signals is available. Such a semi-blind problem is more often encountered in applications and of great importance to practitioners.

In this paper, we shall develop a convex semi-blind source separation method based on partial sparsity of the source spectra. We are concerned with the regime where the sample contains some known and some unknown components. In other words, we have knowledge of some of the components and their concentrations, which is the case of the SWOrRD data. We further assume that the upper bound of the concentrations of the known substances is available, say from experiments or prior knowledge, as is the case of SWOrRD data. Though our method here is designed for Raman spectra, it is applicable to spectra with similar line shapes, such as nuclear magnetic resonance (NMR) data, see [25,26] where more general sparseness source conditions and post-processing methods have been studied. A similar semi-blind source identification approach has been developed for mysterious species of the atmospheric gas mixtures [27]. The method is determined to be an iterative approach. The method for Raman type data contains two major steps. The first step decomposes Raman data X into a linear combination of the reference spectra (known components) plus a remainder, or $X=AS+R$, where the columns of matrix A are known reference spectra, the matrix S contains the non-negative concentrations and has known upper bound. The first step is carried out by solving a constrained least squares problem. The constraints on S help to maintain the nonnegativity of remainder matrix R. The second step performs a nonnegative blind source separation of the remainder matrix R to extract the unknown components. We show that proper sparsity of source signals reduces the general non-convex problem to constrained convex programming permitting solutions with better mathematical properties. Geometrically speaking, this step involves the identification of a convex cone enclosing the data points; Depending on the source sparsity, the cone will be recognized either from its facets or vertexes. Sparse solutions to convex objectives are achieved through ℓ_1 norm minimization and a fast iterative method (the linearized Bregman). The two steps may be iterated. If some of the components from step two are confirmed as chemically meaningful, they are fed back to step one to refine the residual for further extraction of hidden components in step 2.

The paper is organized as follows. In Section 2, we introduce the source sparseness condition and our method. In Section 3, we illustrate the method in processing a set of SWorRD experimental data to identify a prior unknown chemicals. We also test the method and show the numerical results on separation of NMR spectra. Conclusion and future works are discussed in Section 4.

2. The method

In this section, we shall present our method for unknown chemicals detection. Based on the assumption that part of the substances in the chemical mixture and the upper bounds of their concentrations are known, the first step of the method fits these known chemicals to the mixtures. This step solves a constrained least squares optimization.

2.1. Constrained least squares

The following linear model is used for Raman spectra of mixtures

$$X=AS+R, \tag{2.1}$$

where the columns of matrix X are the measured Raman data, the columns of A contain reference spectra of the known chemicals, and those of S matrix contains their concentrations. Matrix R is the fitting residual which might contain hidden spectral structures, the instrument noise, etc. Matrix S is nonnegative on physical ground that its entries represent concentrations. Furthermore, the upper bound of S is known *a priori*. Then for the estimation of S, we minimize the following constrained objective

function:

$$\min_S \|X-AS\|_2^2, \quad \text{s.t. } 0 \le S \le c, \tag{2.2}$$

where the vector c contains the upper bounds of the concentrations of the known substances in the sample. For the Raman data in our numerical experiments, we found that the linear constraint in (2.2) helps to maintain a nonnegative remainder $R = X - AS$. In the residual R, there might be spectral structures (one or many) of chemicals buried in noise, or just random noise. In either case, we factorize the residuals in a blind fashion due to the lack of knowledge of the hidden components. Conventional blind source separation methods such as NMF and ICA are non-convex optimization methods. These methods are for general purpose, yet often unreliable in real-world applications due to non-convexity or sensitivity of their working assumptions. For our problem, we show that proper sparsity of source signals reduces the general non-convex problem to constrained convex programming permitting solutions with better mathematical properties. Sparse solutions to convex objectives are achieved through ℓ_1 minimization.

2.2. Blind source separation with convex cones

The remainder matrix R is factorized as

$$R = WM, \tag{2.3}$$

where the columns of the matrix W are the identified substances, M is their concentrations in the sample. All the matrices are nonnegative. For the purpose of illustration, we shall call $W \in \mathbb{R}^{p\times n}$ the source matrix, $M \in \mathbb{R}^{n\times m}$ the mixing matrix. The dimensions of the matrices are expressed in terms of three numbers: (1) p the number of available samples, (2) m the number of mixture signals, and (3) n the number of source signals. For the Raman data we considered in the paper, there are more mixtures than sources, i.e., $m \ge n$. The goal is to recover W and M for a given R. This is also known as an nonnegative matrix factorization, or NMF problem.

2.2.1. Vertex representation

Various methods have been developed to solve BSS problems by exploiting the natures of source signals. For example, independent component analysis (ICA) recovers statistically independent signals. However, the independence should not be assumed on the Raman spectra of the chemical substances when they share common structural features (the line shapes of their Raman spectra are similar). A better working assumption for the data is a so-called partial sparseness condition. Namely, the source signals are only required to be non-overlapping at some locations of acquisition variable. This sparseness condition was first known in the 1990s [1,29] in the study of blind hyper-spectral unmixing of remote sensing, where the source condition is called pixel purity assumption (PPA) [2]. In 2005, Naanaa and Nuzillard [20] used this assumption to separate the signals in NMR spectroscopy. In fact, this condition is well suited to many chemical substances including the Raman spectra studied in this paper. Such a sparseness condition leads to a dramatic mathematical simplification of a general non-negative matrix factorization problem (2.3) which is non-convex. Geometrically speaking, the problem of finding the mixing matrix M reduces to the identification of a minimal cone containing the rows of matrix R. The latter can be achieved by linear programming. In the context of hyper-spectral unmixing, the resulting geometric (cone) method is the so called N-findr [29], and is now a benchmark in hyper-spectral unmixing. Next we shall review the essentials of the partial spareness condition and the geometric cone method.

Simply speaking, the key sparseness assumption on the source signals is that each source has a stand alone peak at some location of acquisition variable where the other sources are zero. More precisely, the source matrix $W \ge 0$ is assumed to satisfy the following condition:

Assumption. For each $j \in \{1,2,\ldots,n\}$ there exists an $i_j \in \{1,2,\ldots,p\}$ such that $w_{i_j,j} > 0$ and $w_{i_j,k} = 0$, $(k = 1,\ldots,i-1, i+1,\ldots,n)$.

Eq. (2.3) can be rewritten in terms of rows as

$$R^i = \sum_{k=1}^{n} w_{i,k} M^k \quad i = 1,\ldots,p, \tag{2.4}$$

where R^i denote the ith row of R, and M^k the kth row of M. The source assumption implies that $R^{i_j} = w_{i_j,j} M^j$ $j = 1,\ldots,n$ or $M^j = (1/w_{i_j,j}) R^{i_j}$. Hence Eq. (2.4) is rewritten as

$$R^i = \sum_{k=1}^{n} \frac{w_{i,k}}{w_{i_k,k}} M^{i_k}, \tag{2.5}$$

which says that every row of R is a nonnegative linear combination of the rows of M. The identification of M's rows is equivalent to identifying a convex cone of a finite collection of vectors. The cone encloses the data rows in matrix R, and is the smallest of such cones. Such a minimal enclosing convex cone can be found by linear programming methods. Mathematically, the following optimization problems are suggested to estimate the mixing matrix:

$$\min \text{ score} = \left\| \sum_{i=1, i\ne l}^{p} R^i \lambda_i - R^l \right\|_2, \quad \text{s.t. } \lambda_i \ge 0, \quad l = 1,\ldots,p. \tag{2.6}$$

A score is associated with each row of R. A row with a low score is unlikely to be a row of M because this row is roughly a nonnegative linear combination of the other rows of R. On the other hand, a high score means that the corresponding row is far from being a nonnegative linear combination of other rows. The n rows from R with highest scores are selected to form M, the mixing matrix. Fig. 1 is used to demonstrate this vertex-represented cone containing all data points. As depicted in the figure, the cone can be identified by its vertices. Some source signals however do not satisfy the stand-alone peak condition, and this makes the estimation of M's rows intractable to the convex cone method proposed above. Among the cases which stand-alone peak condition are violated, we shall consider a regime where data points lie inside a cone or on its facets, see Fig. 2 for an example. Note that there is no data

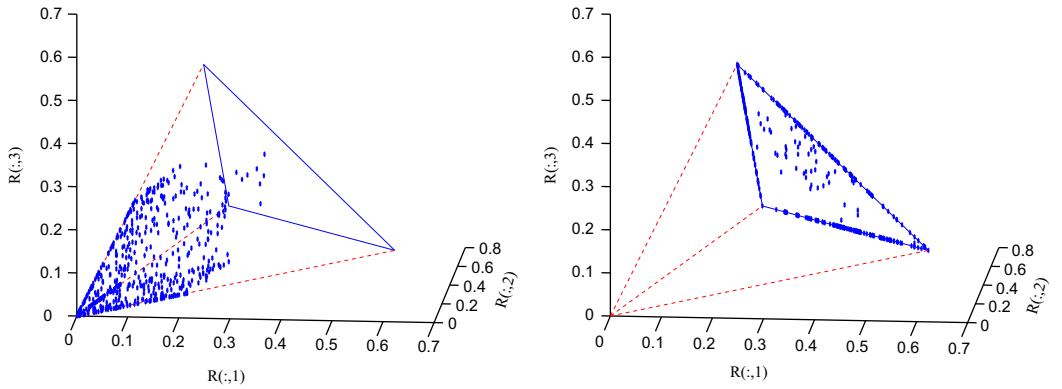

Fig. 1. A cloud of data points (rows of R) rescaled to lie on a plane determined by three vertices of the cone (right).

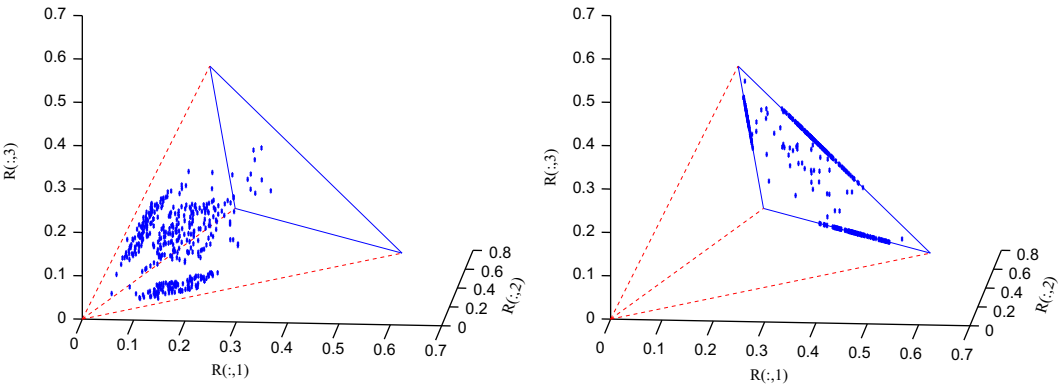

Fig. 2. Scatter plots of rows of X. The cloud of data points (left) is also rescaled to lie on the plane determined by the vertices of the cone (right).

points on the vertices of the cone, we need to develop an approach to identify such a cone by its facets in this case.

2.2.2. Facet representation of a convex cone

Given the point cloud as shown in Fig. 2, the facets of the cone are lying on flat submanifolds (planes in this example of three-dimensional data, and hyperplanes in higher dimension). The identification of these flat geometric structures will help to determine the vertices of the cone, i.e., the rows of the mixing matrix M. The problem is boiled down to the identification of flat structures from a point cloud, which has been an important topic in computer vision. Various approaches have been proposed to recognize (extract) shapes from point clouds in 2D and 3D cases. Classical methods include Hough transform and moving least squares. The Hough transform is a feature extraction technique used in image analysis, computer vision, and digital image processing [24]. The simplest case of Hough transform is a linear transform for detecting straight lines, and it is essentially a voting process where each point belonging to the line (or other pattern) votes for all the possible lines passing through that point. The votes are accumulated in an accumulator array, and the line(s) receiving the maximum vote is taken to be the desired line(s). The main advantages of the transform are its robustness to noise and outliers in the data, and discontinuities in the line (pattern) – both of which are frequently encountered in real-world image and data. The disadvantages of the Hough transform are its demand for a tremendous amount of computing power and large storage. The high cost prohibits them from being used in processing of high dimensional data such as the data we have tested in the paper. For the application under consideration, the dimension of the data can be up to tens or hundreds, hence a more practical approach should be employed. Here we opt for the moving least squares (MLS). The MLS method was proposed by Lancaster and Salkauskas [12] for smoothing and interpolating data. The idea is to start with a weighted least squares formulation for an arbitrary fixed point in $\mathbb{R}^d$, and then move this point over the entire point cloud, where a weighted least squares fit is computed and evaluated for each point individually. We shall briefly review some essentials of the MLS next.

Consider the rows of $R \in \mathbb{R}^{p\times m}$. Note that each row can be viewed as a point in $\mathbb{R}^m$, and there are p points. We

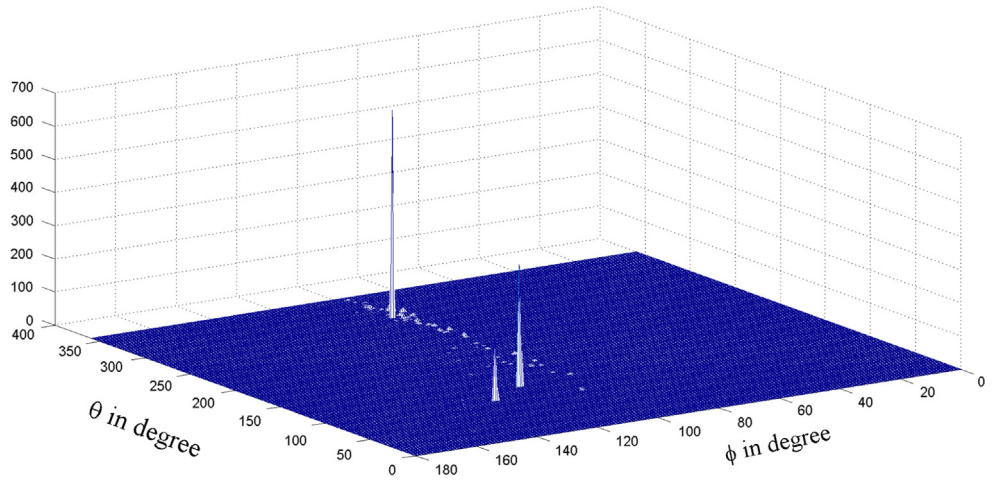

Fig. 3. The voting results on the coefficients by linear MLS, the three peaks imply there are three planes in the data which is plotted in Fig. 2.

shall denote $\mathbf{x}_i = [R(i,1), R(i,2), \ldots, R(i,m-1)] \in \mathbb{R}^{m-1}$, the i-th row of R without the last component. Let $y_i = R(i,m)$. Then the problem formulation is: given p points $\mathbf{x}_i$ where $i \in [1 \ldots p]$. In the weighted least squares, for a fixed point $\overline{\mathbf{x}} \in \mathbb{R}^{m-1}$ we will minimize

$$J = \sum_i \theta(\|\overline{\mathbf{x}} - \mathbf{x}_i\|)\|f(\mathbf{x}_i) - y_i\|^2 \tag{2.7}$$

where f is taken from the space of polynomials of total degree d in q dimensions, denoted as $\prod_d^q$, and can be written as

$$f(\mathbf{x}) = \mathbf{b}(\mathbf{x})^{\mathrm{T}}\mathbf{c} = \mathbf{b}(\mathbf{x}) \cdot \mathbf{c}, \tag{2.8}$$

where $\mathbf{b}(\mathbf{x}) = [b_1(\mathbf{x}), \ldots, b_k(\mathbf{x})]^{\mathrm{T}}$ is the polynomial basis vector and $\mathbf{c} = [c_1, \ldots, c_k]$ is the vector of unknown coefficients, which we wish to minimize in (2.7). Here are some examples for polynomial bases: (a) for $d = 2, q = 2$, $\mathbf{b}(\mathbf{x}) = [1, x, y, x^2, xy, y^2]^{\mathrm{T}}$, (b) for a linear fit $(d = 1, q = 3)$, $\mathbf{b}(\mathbf{x}) = [1, x, y, z]^{\mathrm{T}}$, and (c) for fitting a constant in arbitrary dimensions, $\mathbf{b}(\mathbf{x}) = [1]$. In general, the number k of elements in $\mathbf{b}(\mathbf{x})$ (and therefore in $\mathbf{c}$) is given by $k = (d+q)!/d!q!$, see [14]. (2.7) is similar to conventional least squares only that now the cost function is weighted by $\theta(e)$ where e_i are the Euclidean distances between $\overline{\mathbf{x}}$ and the positions of data points $\mathbf{x}_i$.

The unknown coefficients we wish to obtain from the solution to (2.7) are weighted by distance to $\overline{\mathbf{x}}$ and therefore a function of $\overline{\mathbf{x}}$. Thus, the weighted least square approximation is written as

$$f_{\overline{\mathbf{x}}}(\mathbf{x}) = \mathbf{b}(\mathbf{x})^{\mathrm{T}}\mathbf{c}(\mathbf{x}) = \mathbf{b}(\mathbf{x}) \cdot \mathbf{c}(\mathbf{x}) \tag{2.9}$$

Many choices for the weighting function θ have been proposed in the literature, such as Gaussian $\theta(e) = \exp(-e^2/h^2)$ where h is a spacing parameter which can be used to smooth out small features in the data, see [15]. Another popular weighting function is $\theta(e) = 1/(e^2 + \varepsilon^2)$.

Remark. The choice of weight θ is crucial to make the MLS accurate and robust with respect to noise and non-uniformity which are typical to the data under consideration. After extensive testing, we find that some commonly used weighting functions, such as $\theta(e) = \exp(-e^2/h^2), \theta(e) = (1 - e/h)^4 (4e/h + 1)$ or $\theta(e) = 1/(e^2 + \varepsilon^2)$, work well for our problem, h is a spacing parameter which can be used to smooth out small features in the data. They all give accurate local approximation for the hyperplanes. We opt for the choice of $\theta(e) = \exp(-e^2/h^2)$.

Analogous to conventional least squares, we take partial derivatives of the cost function J with respect to $\mathbf{c}(\overline{x})$

$$2\sum_i \theta(e_i)\mathbf{b}(\mathbf{x_i})[\mathbf{b}(\mathbf{x_i})^{\mathrm{T}} - y_i] = 2\sum_i [\theta(e_i)\mathbf{b}(\mathbf{x_i})\mathbf{b}(\mathbf{x_i})^{\mathrm{T}}c(\overline{x}) - \theta(e_i)b(x_i)] = 0,$$

where $e_i = \|\overline{x} - x_i\|$. We divide by the constant and rearrange to obtain

$$\sum_i \theta(e_i)\mathbf{b}(x_i)b(x_i)^{\mathrm{T}}c(\overline{\mathbf{x}}) = \sum_i \theta(e_i)\mathbf{b}(\mathbf{x_i})y_i,$$

and solve for the coefficients

$$c(\overline{\mathbf{x}}) = \left[\sum_i \theta(e_i)\mathbf{b}(\mathbf{x_i})\mathbf{b}(\mathbf{x_i})^{\mathrm{T}}\right]^{-1} \sum_i \theta(e_i)\mathbf{b}(\mathbf{x_i})y_i.$$

Note that the coefficients $c(\overline{\mathbf{x}})$ are local and need to be recomputed for every $\overline{\mathbf{x}}$.

The idea of MLS is to start with weighted least squares formulation for an arbitrary fixed point, and then move this point over the entire domain, where a weighted least squares fit is computed and evaluated for each point individually. For the data structure considered here, linear polynomials will be used for the fitting. MLS on each data point generates a coefficient vector $\mathbf{c}$. It is then followed by a voting process on these coefficients, the idea of voting is similar to that in Hough transform. Fig. 3 shows the result of the voting on the coefficients computed by MLS, the presence of three large peaks means that there are three planes. Note that these planes pass through the origin. Hence the equation of such a hyperplane is given by the following:

$$0 = \mathbf{P} \cdot \mathbf{n} = xn_x + yn_y + zn_z, \tag{2.10}$$

or

$$0 = x \cdot \cos\theta \cdot \sin\phi + y \cdot \sin\theta \cdot \sin\phi + z \cdot \cos\phi, \tag{2.11}$$

The parameters (normal direction described by angles $\theta$and ϕ) can be easily read off from the plot.

Now we have a convex cone of facet description, and please be reminded that we still need to solve the vertexes of the cone to estimate M. Note that the vertexes are on the intersections of the facets (hyperplanes), then we shall propose to first find the intersections then project them to a cut-plane in the section enclosing positive numbers (e.g., $x+y+z=1$ in 3D). This will deliver one solution of M's rows. In the example above, the equations of three planes read follows:

$$\begin{cases} 5.65x-9.90y-z=0 \\ 2.68x-5.01y+z=0 \\ 1.01x--3.34y-z=0 \end{cases}$$

To solve for the vertexes, any two equations from the above intersect with the projection plane $x+y+z=1$ to generate the following three solutions written as columns of a matrix:

$$\hat{M}=\begin{bmatrix} 0.8686 & 0.7670 & 0.3669 \\ 0.4850 & 0.3855 & 0.3669 \\ 0.1017 & 0.5130 & 0.8546 \end{bmatrix}$$

$$M=\begin{bmatrix} 0.7670 & 0.8686 & 0.3669 \\ 0.3835 & 0.4826 & 0.3669 \\ 0.5113 & 0.0967 & 0.8561 \end{bmatrix},$$

where M is the ground truth, and the first row of $\hat{M}$ is scaled to be same as that of M for easy comparison.

Remark 1. For a point cloud of dimension higher than three, the peaks cannot be visualized, and their locations cannot be read off directly. However, we can detect peaks by looking for local maxima of the surface. Another alternative is to do a low dimensional projection, and try to read the peak locations off.

2.3. Sparse source recovery

With the recovery of M, we solve for W next. An existing method [20] is to directly compute $W=RM^{+}$, where M^{+} is the pseudo-inverse of M. It is however sensitive to noise, and tends to introduce errors and artifacts of negative values. Although a nonnegative least squares can produce a nonnegative W, spurious peaks might be introduced in the results. To benefit from the sparseness source condition in the first step, a more reliable method is to solve a nonnegative ℓ_1 optimization. Although the source signals (columns of W) are not sparse, the rows of W possess sparsity. Hence, we seek the sparsest solution for each row W^i of W:

$$\min \|W^i\|_0 \quad \text{s.t. } W^iM=R^i, \quad W^i\geq 0. \tag{2.12}$$

Here $\|\cdot\|_0$ (0-norm) represents the number of nonzeros. Because of the non-convexity of the 0-norm, we minimize the ℓ_1-norm as a convex relaxation:

$$\min \|W^i\|_1 \quad \text{s.t. } W^iM=R^i, \quad W^i\geq 0, \tag{2.13}$$

which is in the form of linear programming. The fact that the data may in general contain noise suggests us to solve the following unconstrained optimization problem:

$$\min_{W^i\geq 0} \mu\|W^i\|_1+\frac{1}{2}\|R^i-W^iM\|_2^2, \tag{2.14}$$

for which Bregman iterative method [9,31] with a proper projection onto non-negative convex subset is used to obtain a solution. In this paper, we shall use the linearized Bregman iteration to solve (2.18) due to its efficiency. For each row W^i of W, we introduce $u=(W^i)^{\mathrm{T}}, f=(R^i)^{\mathrm{T}}, B=M^{\mathrm{T}}$, then (2.18) is equivalent to

$$\min_{u\geq 0} \mu\|u\|_1+\tfrac{1}{2}\|f-Bu\|_2^2. \tag{2.15}$$

The l_2 norm in (2.18) and (2.15) is to model the unknown measurement error or noise as Gaussian. When there is minimal measurement error, one must assign a tiny value to μ to heavily weigh the fidelity term $\|f-Bu\|_2^2$ in order for $Bu=f$ to be nearly satisfied. The linearized Bregman method can be written iteratively by introducing an auxiliary variable v^j:

$$\begin{cases} v^{j+1}=v^j-B^{\mathrm{T}}(Bu^j-f), \\ u^{j+1}=\delta\cdot shrink_+(v^{j+1},\mu), \end{cases} \tag{2.16}$$

where $u^0=v^0=O$, $\delta>0$ is the step size, and shrink$_+$ is for computing nonnegative solutions,

$$shrink_+(v,\mu)=\begin{cases} v-\mu, & \text{if } v>\mu, \\ 0, & \text{if } v<\mu. \end{cases} \tag{2.17}$$

To summarize, we use the following flow chart to show the pipeline of the method proposed in this work.

Algorithm 2 (*Semi-BSS method*). Input: the measured data matrix X and a list of known chemical spectra in matrix A.

1. (Decomposition) Decompose the data into a sum of linear combination of the known components and a non-negative residual, $X=AS+R$. The method is a constrained least squares. The next steps are to extract unknown chemical spectra from the residual.
2. (BSS with convex cones) Suppose that the residual matrix $R=WM$ where the columns of the matrix W are the identified substances, M is their concentrations in the sample. To recover M, we use convex cone method based on vertex component analysis or facet component analysis. Linear programming may be used to identify the vertexes of the cone (which are rows of M) if they are among the data. For the scenario where data points (rows of R) lie either inside or on the facets of a convex cone, yet none of them are located on edges (vertices). The cone structure will be reconstructed from its facets instead of the vertices. We use moving least squares to identify the facets, other methods include Hough transform.
4. (Plane fitting and intersecting) For the convex cone of facet representation, we first obtain the normal directions of the planes where the facets lie on, then the equations of the planes. Obtain the m intersections of any $m-1$ planes out of m planes with the plane $\mathbf{x}^T\cdot\mathbf{1}=1$, and form the m rows of mixing matrix M.

5. (Source recovering) We seek the sparsest solution for each row W^i of W. We shall solve following unconstrained optimization problem,

$$\min_{W^i \geq 0} \mu \| W^i \|_1 + \tfrac{1}{2} \| R^i - W^i M \|_2^2, \tag{2.18}$$

for which Bregman iterative method with a proper projection onto non-negative convex subset is used to obtain a solution.

6. (Verification) For the recovered sources W, verification based on the library templates or expert knowledge helps to confirm these components. If necessary, the confirmed meaningful components are fed back into step 1 to refine the residual and then steps 2–7 extract possibly more hidden components. The steps may be iterated until no more components can be identified.

3. Numerical experiments

In this section, we shall apply our method to the so-called SWOrRD Raman data and NMR data, and present the computational results.

3.1. Experimental data

The Swept Wavelength Optical Resonant Raman Detector (SWOrRD) at the Naval Research Laboratory is a spectroscopy system which is able to produce two-dimensional spectral data of biological agents and chemical substances. A target substance is illuminated with a specific laser wavelength and this generates a resonance Raman spectrum. When the laser wavelength is varied and the process is repeated, the resonance Raman spectra (one at each laser wavelength), forms a two-dimensional plot (signature) where one axis is the input laser wavelength and the other axis is the wavenumber of the Raman spectrum. Fig. 4 is an example of 2D SWOrRD Raman spectra of a mixture of ethanol, ethylene glycol, acetonitrile, and water. The data matrix X corresponds to this 2D spectra. Each column X is a Raman spectrum of the mixture at a specific laser wavelength. For SWOrRD data, there are much more mixtures than the number of sources. The spectral features may be slightly altered from one wavelength to another. To reduce this effect, we do not necessarily use all the data, and so we sample the spectra at a subset of input laser wavelengths.

The other dataset is NMR spectra. NMR technique is widely used by chemists and biochemists to investigate the properties of organic molecules, though it is applicable to any kind of sample that contains nuclei possessing spin. 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. The real part of the spectrum can be presented as the sum of symmetrical, positive valued, Lorentzian-shaped peaks. Laboratory NMR spectrometer usually produce spectra of high resolution. However, as the instruments age, some pixels may become less responsive due to damage or dust particles being present. The resulting spectra will contain limited information due to the resolution loss. In this case, the stand-alone peak condition can be easily violated, i.e., no data points are on the vertexes of the cone. They lie on the facets and inside of the cone. And the method proposed can be used to retrieve the mixing matrix from the facets representation.

Next, we report the computational results from our proposed method.

3.2. Results of detection of a prior unknown chemicals

We tested our method on two sets of SWORrD data. The samples consist of several liquid substances, some of them are commonly used for synthesizing explosives. The first data set includes 21 mixed Raman spectra at different

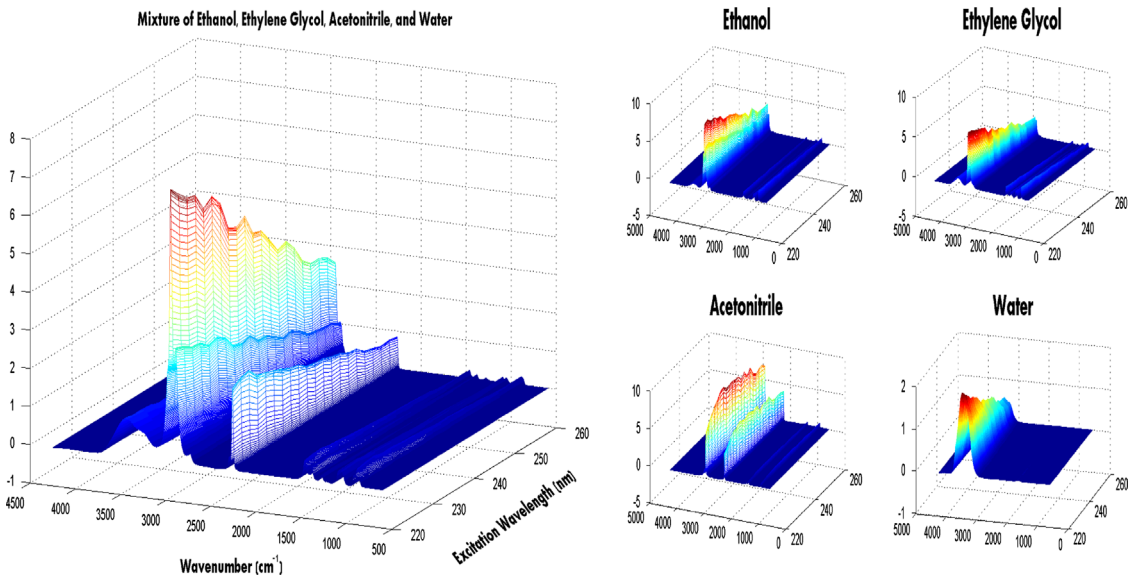

Fig. 4. SWOrRD 2D signatures of four pure substances are shown in boxes on the right. A signature of a mixture of these pure substances is shown on the left.

excitation wavelengths. This sample is known to contain methanol and its concentration is below 1/3. There are also two *a priori* unknown liquid chemicals. The method first fits the known reference spectrum of methanol to the Raman data. Then it identifies two unknown chemicals from the fitting residuals. The results are shown in a series of plots, Figs. 5–8. From all the data available, we found that the 5 spectra at consecutive laser wavelengths of (248, 250, 252, 254, 256) nm produce the best results. The theoretical underpinning of optimal selection of the data is under further study. Fig. 5 shows the line shape of the spectrum of the mixture, and the spectral reference of methanol. The residual after fitting methanol to the data is plotted in Fig. 6, where some structure can be seen. Then further identification of the two hidden chemicals was made by the convex BSS method, and the results are presented in Fig. 7. Compared to the ground truth, the results are satisfactory in that the recovered spectral structures are recognizable as ethanol and acetonitrile.

As a comparison, we present here the results by non-negative matrix factorization (NMF). The NMF was first introduced by Lee and Seung [13], sometimes called also PMF (positive matrix factorization), decomposes the data matrix R as a product of two matrices W and M having only nonnegative elements. The NMF was first introduced in its modern formulation as a method to decompose images. NMF will be applied to extract signals from the fitting residual after fitting the known spectral references to the data, Consider the following cost function:

$$J(W,M) = \| R - WM \|^2 = \sum \| [R]_{i,k} - [WM]_{ik} \|^2$$

$$\text{s.t. } W_{ij} \geq 0,\ M_{j,k} \geq 0. \tag{3.1}$$

This is a non-convex problem, so the NMF algorithms may or may not converge to the same meaningful solutions on each run, depending on the random initial conditions and the kind of algorithms we use. Using the gradient descent approach for this cost function in respect of elements $W_{i,j}$ of W, we obtain

$$\Delta W_{i,j}(l+1) = W_{i,j}(l+1) - W_{i,j}(l) \tag{3.2}$$

$$\Delta W_{i,j}(l+1) = -\eta_{ij}\frac{\partial J}{\partial W_{i,j}} = -\eta_{ij}\left([RM^{\mathrm{T}}]_{i,j} - [WMM^{\mathrm{T}}]_{i,j}\right). \tag{3.3}$$

Assuming that elements $W_{i,j}$ are fixed, we obtain additive update rule for elements M_{jk} of M

$$\Delta M_{j,k}(l+1) = M_{j,k}(l+1) - M_{j,k}(l) \tag{3.4}$$

$$\Delta M_{j,k}(l+1) = -\tilde{\eta}_{jk}\frac{\partial J}{\partial M_{j,k}} = -\tilde{\eta}_{jk}\left([WA^{\mathrm{T}}]_{j,k} - [W^{\mathrm{T}}WM]_{j,k}\right). \tag{3.5}$$

To ensure automatically non-negativity constraints on W and M, Lee and Seung proposed to choose a specific learning rates in [13]

$$\eta_{ij} = \frac{W_{i,j}}{[WMM^{\mathrm{T}}]_{ij}}, \quad \tilde{\eta}_{jk} = \frac{M_{jk}}{[W^{\mathrm{T}}WM]_{jk}} \tag{3.6}$$

which leads to a simple multiplicative update rules:

$$W_{i,j} \leftarrow W_{i,j}\frac{[RM^{\mathrm{T}}]_{i,j}}{[WMM^{\mathrm{T}}]_{i,j}}, \quad M_{j,k} \leftarrow W_{j,k}\frac{[W^{\mathrm{T}}R]_{j,k}}{[W^{\mathrm{T}}WM]_{j,k}} \tag{3.7}$$

Study shows that this standard NMF (without any auxiliary constraints) provides sparseness of its component, one can achieve some control this sparsity by imposing additional constraints. Among many other works, the spareness constrained NMF was studied by Hoyer [10] where an ℓ_1 norm was imposed either on source or the

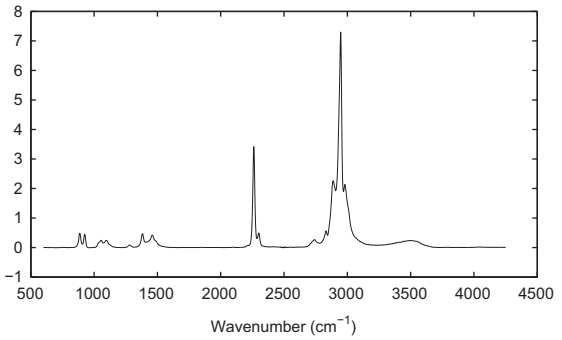

Fig. 6. One fitting residual from the constrained least squares data fitting.

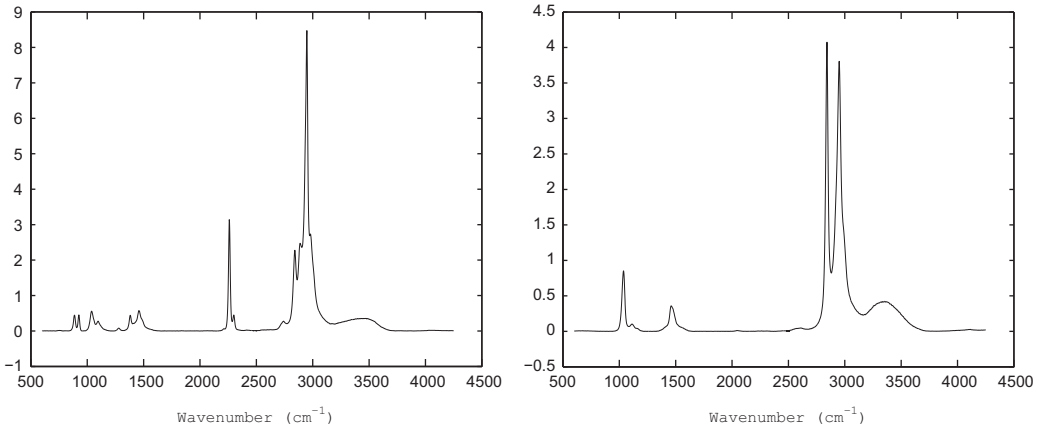

Fig. 5. A Raman spectrum of the mixture of methanol and two other liquid chemicals is shown on the left, while the Raman spectral reference of methanol is plotted on the right.

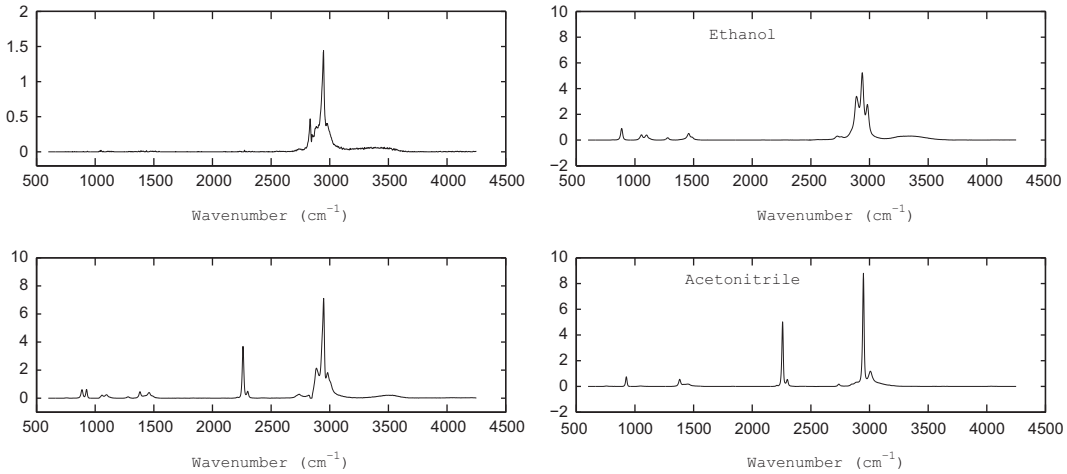

Fig. 7. Left column is the two identified spectral structures, right column shows the Raman spectral references for ethanol and acetonitrile. The value of μ used in step 2 is 0.09.

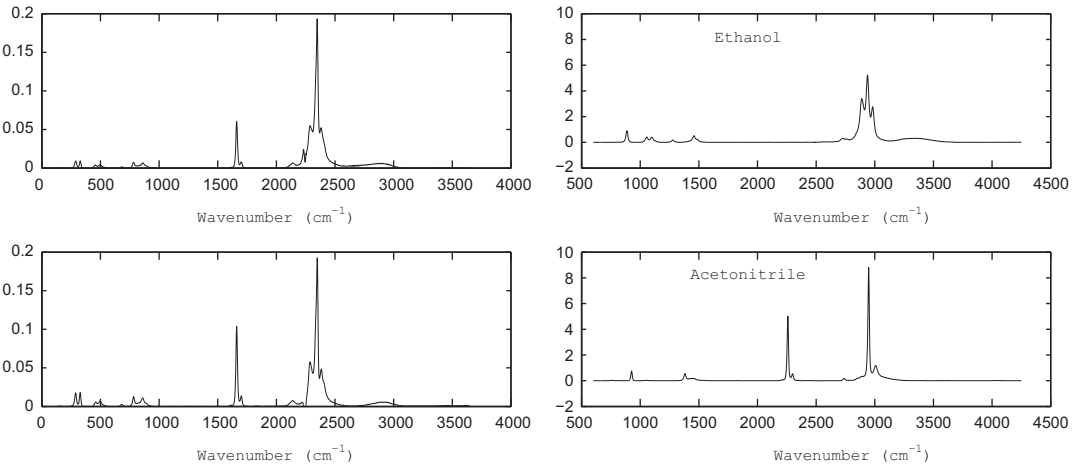

Fig. 8. Left column is the results computed by NMF, right column shows the Raman spectral references for ethanol and acetonitrile.

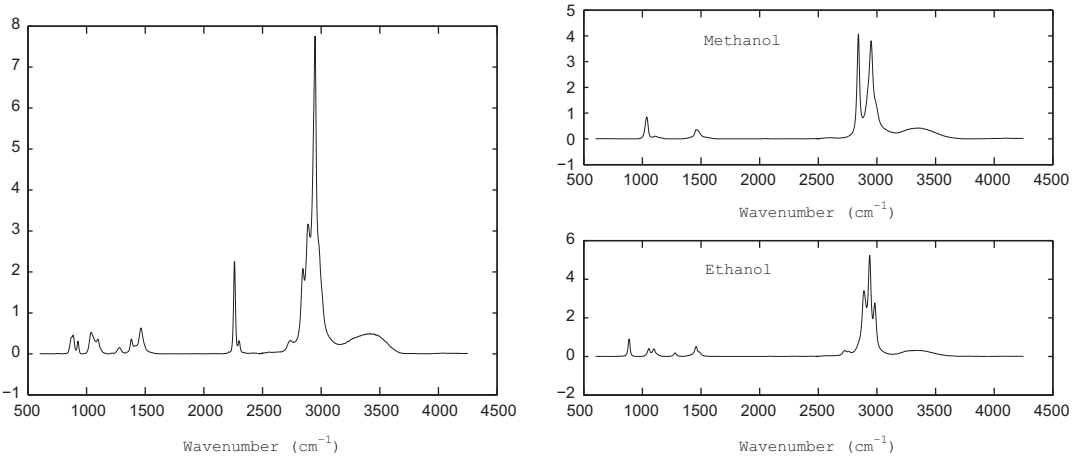

Fig. 9. A Raman spectrum of the mixture of methanol, ethanol and two other chemicals is shown on the left, while the Raman spectral references of methanol and ethanol are plotted on the right. $\mu = 0.09$ is used for the source recovery in step 2.

mixing matrix depending on the problem. Despite their successes in some BSS problems, the NMF and its extensions are non-convex optimization methods, hence often are unreliable in real-world applications due to non-convexity. To show the improvements of our proposed method, we present some NMF results of two extracted spectral structures in Fig. 8. It is clear that our method delivers better results.

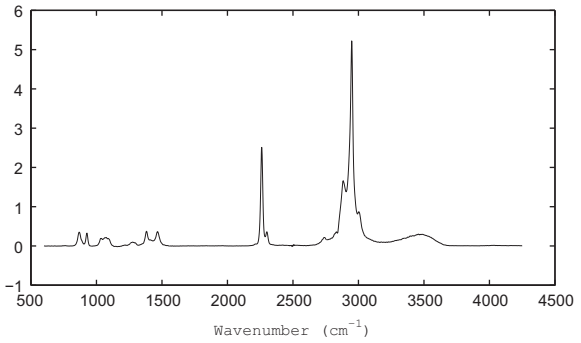

Fig. 10. One fitting residual from the constrained least squares data fitting.

In the second example, we try to identify two hidden chemicals from a mixture of four chemicals. The known chemicals are methanol and ethanol. The total concentration of the two is below 1/2, which is known from the sensing hardware. The computational results are presented in Figs. 9–13. Although no apparent Raman spectral structure can be identified from the first plot in Fig. 11, the second structure is easily recognizable as acetonitrile upon comparison with reference spectrum. Next we subtract the confirmed acetonitrile from the residual by solving the constrained least squares in step 1, then feed the new residual back to step 2 for extracting more hidden chemicals. The extracted structure in Fig. 13 can be easily identified as ethylene glycol comparing to the ground truth. For a comparison, we also provide NMF results of the recovered spectral structures in Fig. 12.

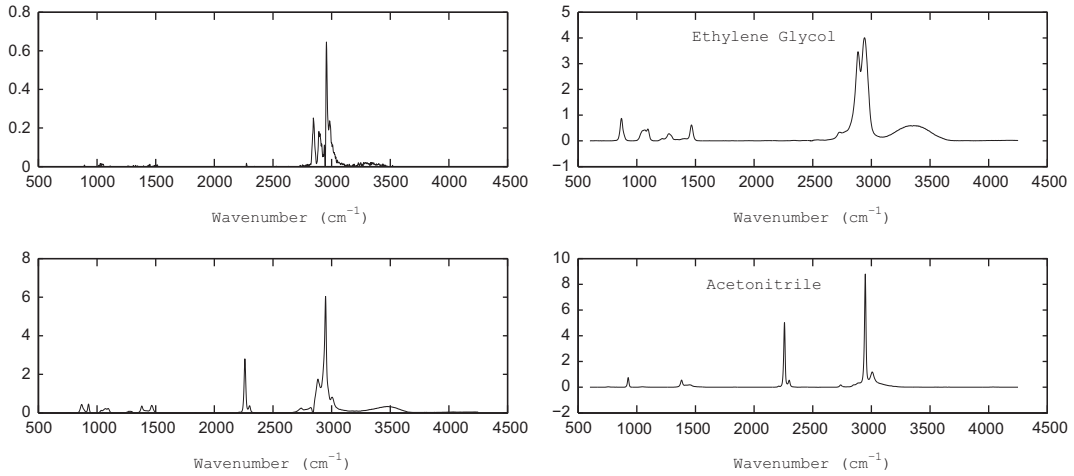

Fig. 11. Left column is the two identified spectral structures, right column shows the Raman spectral references for acetonitrile and ethylene glycol.

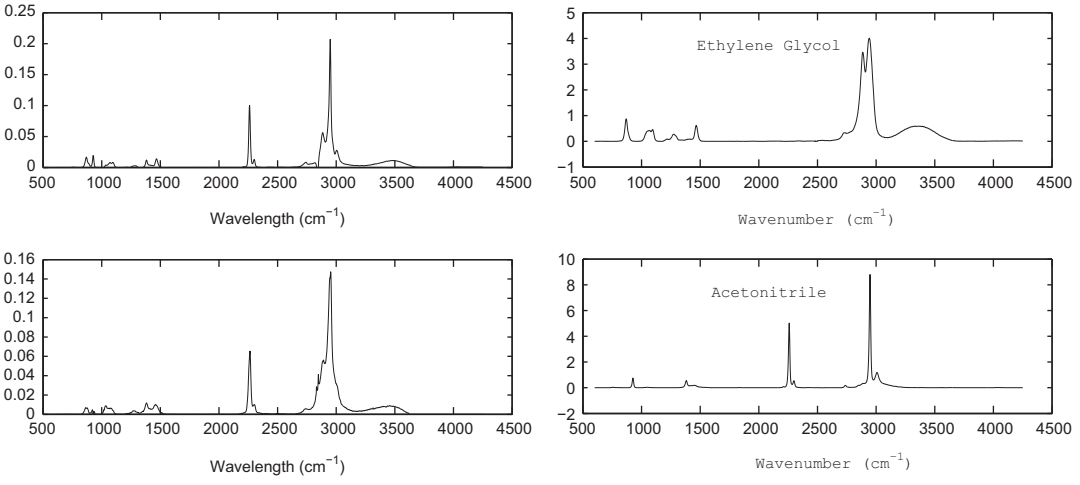

Fig. 12. Left column is the two identified spectral structures by NMF, right column shows the Raman spectral references for acetonitrile and ethylene glycol.

Computations are performed on a Dell laptop with 6 G RAM and 1.6 GHz i7 CPU. The cpu time for Example 1 is 7.613 s, and it is 10.327 s for Example 2. The method proves to be efficient and can be of use in rapid detection of chemical substances.

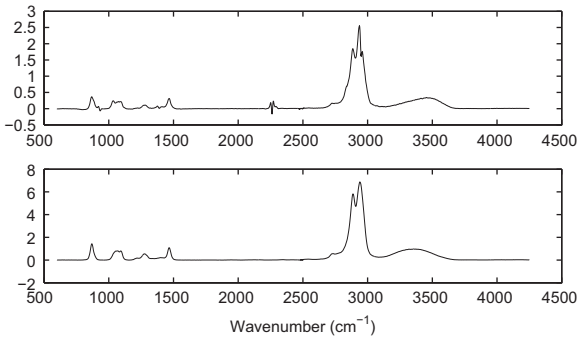

Fig. 13. Top panel is the recovered structure after subtracting acetonitrile. The spectral line shape is easily recognizable as ethylene glycol, whose spectral reference is shown in the bottom plot.

3.3. Detection of chemical compounds from their NMR spectra

In a third example we apply the method to NMR data. We used true NMR spectra of four compounds mannitol, β-cyclodextrine, β-sitosterol, and menthol as source signals (data are from [20]). Suppose that mannitol is known to be contained in the mixture sample, and the upper bounds of its mixing coefficients are also assumed to be given. To mimic the information loss in the data, we remove a few peaks from the mixture data, and replace the removed peaks with zeros. The left plot in Fig. 14 shows the original mixture, and the right plot is a mixture spectrum with damaged pixels. We add Gaussian noise (SNR = 60 dB) to the data to test the robustness of the method. The three peaks in Fig. 15 imply that three planes are detected in the scattered plot of Fig. 14, and the coordinates of the peaks correspond to the normal directions of these planes, and they can be easily read off in this example. The recovered source spectra and their ground truth are shown in Fig. 16. the results are reasonably good in the presence of damaged pixels in the data. For comparison, NMF results are provided in Fig. 17.

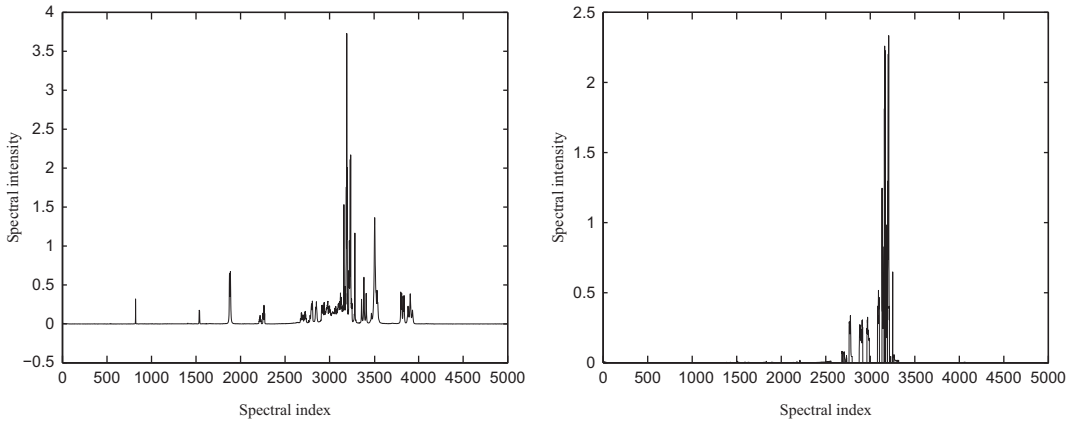

Fig. 14. Left panel is one of the original mixture spectrum; in the right plot, some pixels are removed and replaced by zeros to mimic the information loss by an aged instrument.

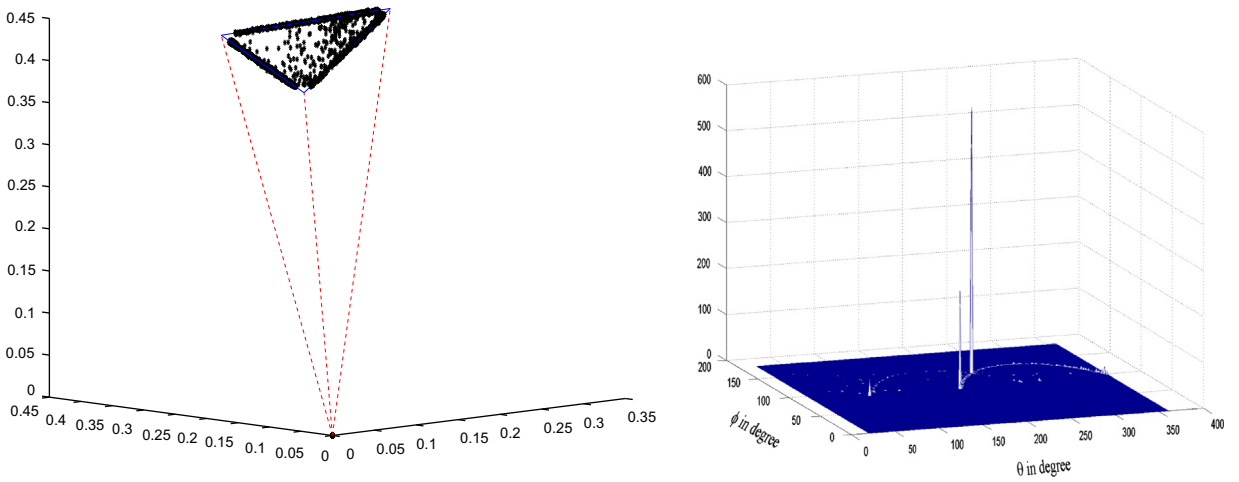

Fig. 15. The three peaks (right) correspond to the normal directions of the three planes in the scattered plot on the left.

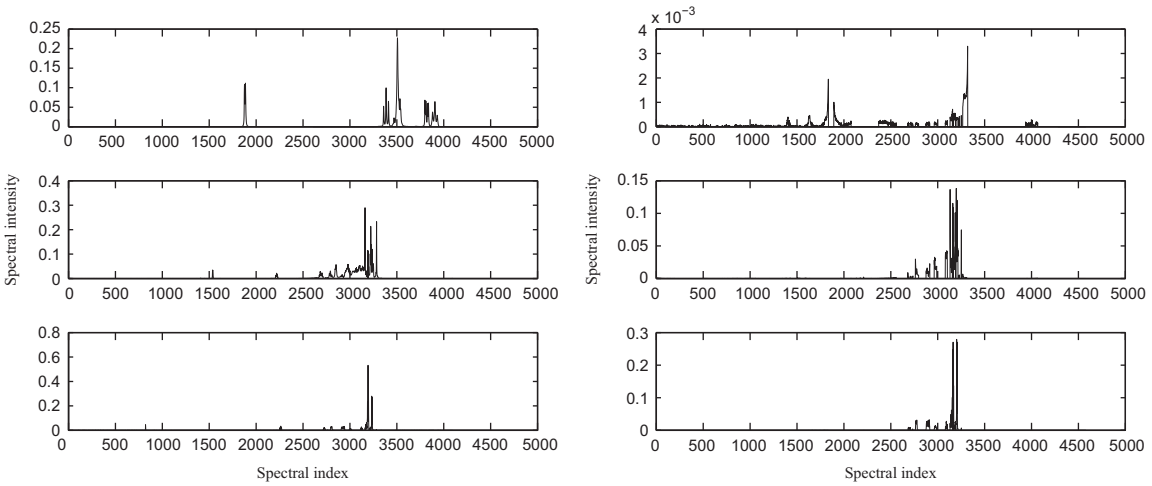

Fig. 16. Left: the true source signals. Right: source signals recovered by our method.

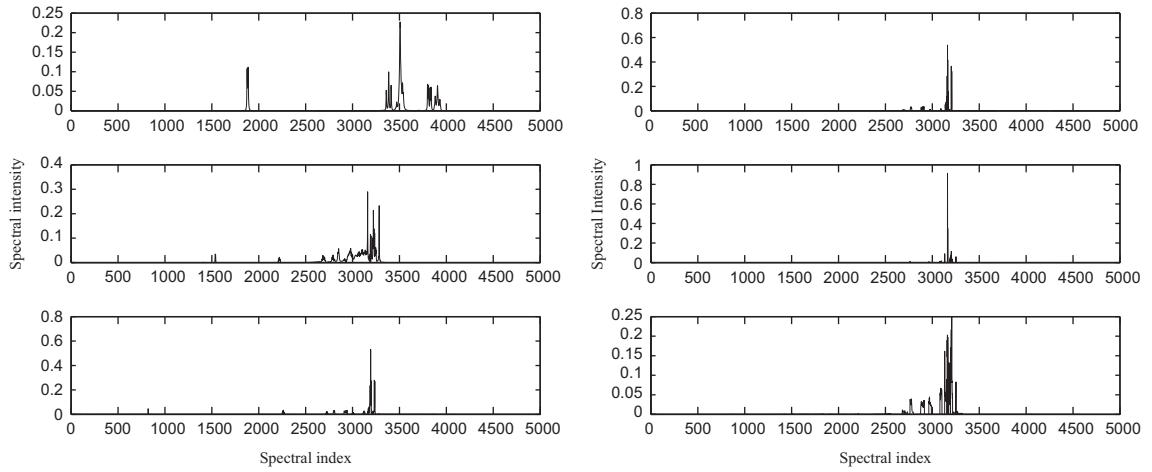

Fig. 17. Left: the true source signals. Right: source signals recovered by NMF.

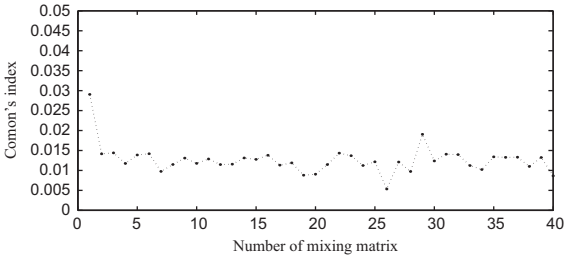

Fig. 18. Performance of our method on 40 random 3×3 mixing matrices. The three sources in Example 3 are used.

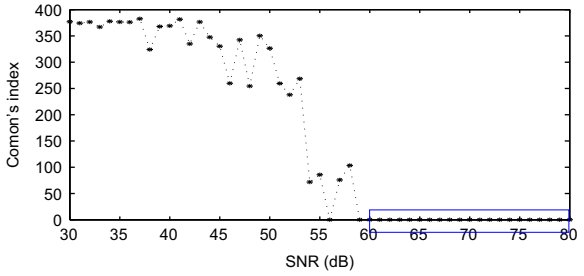

Fig. 19. Test the robustness of our method in the presence of noise. The portion in the blue box is zoomed in and depicted in Fig. 20. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

To test the performance of our method, we compute the Comon's index [4]. The index is defined as follows: let A and $\hat{A}$ be two nonsingular matrices with L_2-normalized columns. Then the distance between A and $\hat{A}$ denoted by $\varepsilon(A,\hat{A})$ which reads

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

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

where $D = A^{-1}\hat{A}$, and d_{ij} is the entry of D. In [4] Comon proved that A and $\hat{A}$ are considered nearly equivalent in the

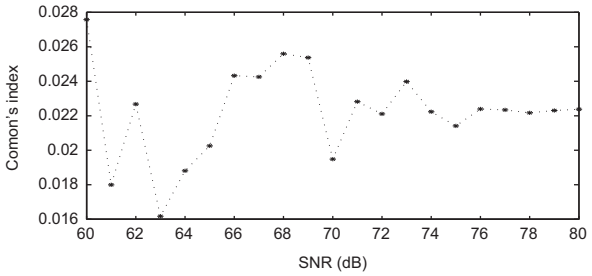

Fig. 20. The Comon's index V.S. noise of SNR=60–80 dB.

sense of BSS (i.e., $\hat{A} = AP\Lambda$) if $\varepsilon(A,\hat{A}) \approx 0$. Figs. 18 and 19 show Comon's indices between the true mixing matrices and the computed matrices by our method. For the result in Fig. 18, we compute the Comon's indices using the three sources in Example 3 and 40 4×4 random mixing matrices. Clearly the Comon's indices are very small suggesting the equivalence in the sense of BSS of the true mixing matrices and the computed ones. Figs. 19 and 20 show the performance of our method in the presence of noise. The three sources in Example 3 are combined to generate three noisy mixtures by adding white Gaussian noises with SNR varying from 30 dB to 80 dB. The reliability of our method is manifested from the plots.

4. Conclusions and future work

A semi-blind sparse and convex source identification method is developed to extract meaningful spectral structures from mixture data of Raman spectroscopy and NMR spectroscopy. The method is designed to identify potentially hidden chemicals after fitting the known spectral references to the data. Our method can be useful for extracting unknown source signals from the residuals after known reference spectra have been first deployed to fit the data. The major strength of the technique is its ability to be used either with known reference spectra for quantification or without reference spectra for identification of unknown/hidden chemical substances. Numerical results on SWOrRD data and NMR data showed the promising potential of our method on explosives/chemicals detection.

The model considered in the paper is a linear and stationary model which assumes no shift and(or) squeeze in the spectral lines. A future line of work will study how to build this nonlinear effect into the identification model. Given that the shift and(or) squeeze amount is small, one may explore the idea of image registration. We also plan to study more reliable and efficient methods for the residuals decomposition, as their success highly depends on a viable working assumption on the pure signals.

The semi-blind source identification problem we addressed here also has analogues in detecting atmospheric trace gases with the so called differential optical absorption spectroscopy. Analysis of the fitting residuals for mysterious species arises there as well, and is non-convex in general. Recently the authors achieved some success in this direction based on a similar semi-blind framework, although the details of fitting and blind source identification process are quite different [27].

Acknowledgments

This work was partially supported by NSF-ADT Grant DMS-0911277. The authors thank Naval Research Laboratory for the experimental Raman data, and Dr. J. Grun for helpful discussions.

References

[1] J. Boardman, Automated spectral unmixing of AVRIS data using convex geometry concepts, in: Summaries of the IV Annual JPL Airborne Geoscience Workshop, JPL Pub. 93–26, vol. 1, 1993, pp. 11–14.
[2] C.-I. Chang (Ed.), Hyperspectral Data Exploitation: Theory and Applications, Wiley-Interscience, 2007.
[3] G. Comanescu, C.K. Manka, J. Grun, S. Nikitin, D. Zabetakis, Identification of explosives with two-dimensional ultraviolet resonance Raman spectroscopy, Applied Spectroscopy 62 (8) (2008) 833–839.
[4] P. Comon, Independent component analysis—a new concept? Signal Processing 36 (1994) 287–314.
[5] K. Faulds, W.E. Smith, D. Graham, Evaluation of surface-enhanced resonance Raman scattering for quantitative DNA analysis, Analytical Chemistry 76 (2) (2004) 412–417.
[6] C.A. Froud, I.P. Hayward, J. Laven, Advances in the Raman depth profiling of polymer laminates, Applied Spectroscopy 57 (12) (2003) 1468–1474.
[7] M. Gaft, L. Nagli, Standoff laser based spectroscopy for explosives detection, Proceedings of SPIE 6737 (2007) 673903.
[8] J. Grun, C.K. Manka, S. Nikitin, D. Zabetakis, G. Comanescu, D. Gillis, J. Bowles, Identification of bacteria from two-dimensional resonant-Raman spectra, Analytical Chemistry 79 (2007) 5489–5493.
[9] Z. Guo, S. Osher, Template matching via ℓ_1 minimization and its application to hyperspectral target detection (Technical Report 09-103), UCLA, ⟨http://www.math.ucla.edu/applied/cam/⟩, 2009.
[10] P. Hoyer, Non-negative matrix factorization with sparseness constraints, Journal of Machine learning Research 5 (2004) 1457–1469.
[11] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley and Sons, New York, 2001.
[12] P. Lancaster, K. Salkauskas, Surfaces generated by moving least squares methods, Mathematics of Computation 87 (1981) 141–158.
[13] D.D. Lee, H.S. Seung, Learning of the parts of objects by non-negative matrix factorization, Nature 401 (1999) 788–791.
[14] D. Levin, The approximation power of moving least-squares, Mathematics of Computation 67 (1998) 1517–1531.
[15] D. Levin, Mesh-independent surface, Geometric Modeling for Scientific Visualization 3 (2003) 37–49.
[16] D.A. Long, The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules, John Wiley and Sons, Ltd., 2002.
[17] H. Li, T. Adali, W. Wang, Non-negative matrix factorization with orthogonality constraints and its application to Raman spectroscopy, Journal of VLSI Signal Processing Systems for Signal Image and Video Technology 48 (2007) 83–97.
[18] S. Miron, M. Dossot, C. Carteret, S. Margueron, D. Brie, Joint processing of the parallel and crossed polarized Raman spectra and uniqueness in blind nonnegative source separation, Chemometrics and Intelligent Laboratory Systems 105 (1) (2011) 7–18.
[19] J.V. Miller, E.G. Bartick, Forensic analysis of single fibers by Raman spectroscopy, Applied Spectroscopy 55 (12) (2001) 1729–1732.
[20] W. Naanaa, J.-M. Nuzillard, Bind source separation of positive and partially correlated data, Signal Processing 85 (9) (2005) 1711–1722.
[21] I. Notingher, S. Verrier, S. Haque, J.M. Polak, L.L. Hench, Spectroscopic study of human lung epithelial cells (A549) in culture: living cells versus dead cells, Biopolymers 72 (4) (2003) 230–240.
[22] C.J. Rowlands, S.R. Elliott, Improved blind-source separation for spectra, Journal of Raman Spectroscopy 42 (2011) 1761–1768.
[23] S. Šašić, D.A. Clark, Defining a strategy for chemical imaging of industrial pharmaceutical samples on Raman Line-Mapping and global illumination instruments, Applied Spectroscopy 60 (5) (2006) 494–502.
[24] L.G. Shapiro, G.C. Stockman, Computer Vision, Prentice-Hall Inc., 2001.
[25] Y. Sun, C. Ridge, F. del Rio, A.J. Shaka, J. Xin, Postprocessing and Sparse Blind Source Separation of Positive and Partially Overlapped Data, Signal Processing 91 (8) (2011) 1838–1851.
[26] Y. Sun, J. Xin, A Recursive Sparse Blind Source Separation Method and its Application to Correlated Data in NMR Spectroscopy of Biofluids, Journal of Scientific Computing 51 (3) (2012) 733–753.

[27] Y. Sun, L.M. Wingen, B.J. Finlayson-Pitts, J. Xin, A Semi-blind Source Separation Method for Differential Optical Absorption Spectroscopy of Atmospheric Gas Mixtures, arXiv:1109.6370[math.NA], 2012.
[28] G. Turrell, Infrared and Raman Spectra of Crystals, Academic Press, 1972.
[29] M.E. Winter, N-findr: an algorithm for fast autonomous spectral endmember determination in hyperspectral data, Proceedings of SPIE 3753 (1999) 266–275.
[30] Y. Yao, Q. Li, J. Zhang, R. Liu, L. Jiao, Y.T. Zhu, Z. Liu, Temperature-mediated growth of single-walled carbon-nanotube intramolecular junctions, Nature Materials 6 (2007) 283–286.
[31] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithm for ℓ_1-minimization with applications to compressive sensing, SIAM Journal on Imaging Sciences 1 (143) (2008) 143–168.