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

Computational Modeling of Spectral Data Fitting with Nonlinear Distortions

Yuanchang Sun · Wensong Wu · Jack Xin

Received: date / Accepted: date

Abstract Substances such as chemical compounds and biological agents are invisible to human eyes, they are usually captured by sensing equipments with their spectral fingerprints. Although spectra of pure substances can be identified by visual inspection, the spectra of their mixtures take a variety of complicated forms. Given the knowledge of spectral references of the constituent substances, the task of data fitting is to solve their weights, which usually can be solved by a least squares. Complications occur if the basis functions (reference spectra) may not be used directly to best fit the data. In fact, random spectral distortions such as shifting, compression, and expansion have been observed in some spectra. In this paper, we formulate mathematical model for such distortions and build them into data fitting algorithms. If minimal knowledge of the distortions is available, a deterministic approach termed *augmented least squares* is developed by fitting the spectral references together with their derivatives to the mixtures. If the distribution of the distortions is known *a prior*, we propose to solve the problem with maximum likelihood estimators which incorporate the distortions into the variance matrix. The proposed methods are substantiated with numerical examples including data from Raman spectroscopy (RS), nuclear magnetic resonance (NMR), and differential optical absorption spectroscopy (DOAS) and show satisfactory results.

Yuanchang Sun · Wensong Wu
Department of Mathematics and Statistics, Florida International University, Miami, FL 33199, USA.

E-mail: yuasun@uci.edu; wensong.wu@fiu.edu

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

1 Introduction

Given a (or multiple) noisy spectral observation(s) of a sample containing several sources (e.g., chemicals), one can retrieve the weights of source signals in the mixtures by the least squares type of methods, the most widely used data analysis tools with numerous applications. Suppose the observations are linear mixtures of the source signals

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

where $\boldsymbol{X} \in \mathbb{R}^{m\times p}$ is the $m \times p$ observation matrix (mixture matrix), each row $\boldsymbol{x}_i$ stands for a single observation; $\boldsymbol{S}$ is the basis (or source) matrix with each $\boldsymbol{s}_i$ being a pure source signal; $\boldsymbol{A}$ is the mixing matrix of size $n \times m$, and $\boldsymbol{N}$ is the noise matrix. Given $\boldsymbol{S}, \boldsymbol{X}$, the least squares solution of the mixing matrix is given by $\hat{\boldsymbol{A}} = \boldsymbol{X}\boldsymbol{S}^{\mathrm{T}}(\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}})^{-1}$ which is the minimizer of $\min_{\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{A}\,\boldsymbol{S}\|_2^2$. Least squares has been very successful with data fitting problems. However, complications occur if the basis functions (rows of $\boldsymbol{S}$) may not be used directly to best fit the data. For example, the reference spectra may have either shifting or compression (or expansion) when the source objects are mixed, these effects are sometimes also called spectral variability (distortions). Then the reference spectra need to be properly registered before fitting.

In this paper, we consider a regime where the spectral signals possess such spectral variability in addition to the noise and measurement errors. This has been a challenging problem in signal and image processing, for example, in hyperspectral (HSI) demixing [3], the endmembers (source signals) exhibit either mismatchs to their lab measured spectral signatures (for example, peak shifts) or random distortion (some peaks are being compressed or expanded), these spectral variability

maybe caused by a number of factors such as lighting conditions, geographic locations, seasons, etc. Similar phenomena has been observed in Raman spectroscopy, where spectral lines of source species can have random shifts along the wavenumber axis. In differential optical absorption spectroscopy (DOAS), spectra are collected by moving a grating motor. Changes in grating position and temperature lead to changes in dispersion of the light beam on the detector. In nuclear magnetic resonance spectroscopy (NMR), spectral variations may exist in the data as the instruments age or the changing environment. To handle the spectral variabilities, experimentalists often adjust the reference spectra on a lookup table by knowledge of the sensing process. In hyperspectral processing, researchers have proposed various methods for the spectral variability for unmixing the observed pixels, the methods include deterministic approaches and statistical approaches (readers are referred to [3] and references therein). For example, authors in [8] design a selection criterion to include as many as possible variety of reflectance spectra of similar endmembers, the hyperspectral data is then unmixed with all possible different combinations of endmember signatures. In hyperspectral target detection, a robust matched filter was designed in [6] by allowing the mismatch between the target source spectrum and its reference within a predefined ϵ sphere. In [9], an ℓ_1 minimization based approach for robust template matching (data matching) in hyperspectral classification and target detection. Other statistical approach includes Bayesian spectral mixture analysis method; In [2] the method uses the endmember signature probability distribution in the analysis for maximally capturing the spectral variability of an image with the least number of endmembers. In the context of image fusion, spectral distortion is introduced and usually difficult to control when combining relevant information from two or more images into a single image from multiple sensors. In [10], a total variation based evaluation method is developed to reduce the distortion effects and hence improve the fusion quality. In dealing with the problem of separating overlapped texts in degraded documents where distortions can be due to back-to-front interferences (or see-through,show-through/bleed-through), the authors in [11] proposed a linear convolutional density model in the blind fashion. The proposed method outperforms standard blind source separation techniques such as independent component analysis and produces results comparable to those of the best performing among recent state-of-the-art methods. In [12] a perturbed linear model was proposed where the spectral variability in the endmembers was represented by a spatially varying perturbation.

Although the spectral distortions in hyperspectral image processing and other contexts have been extensively studied, the assumptions and methods are all seeming to be limited to specific applications and could be *ad hoc* in many cases. Less has been done for other spectral mixtures including Raman, NMR, and DOAS spectroscopy. In this work, we shall assume a database containing only one reference spectrum for each source, and the objective is to improve the estimates of the mixing matrix. This really sets our work apart from the methods aforementioned. To our best knowledge, the perturbation modeling (in hyperspectral imaging) introduced in [12] is probably the closest to our work in the paper, it is however in the class of blind signal separation where both the endmembers (source signals) and their abundances are to be estimated. Given the fact that the random shifts is rather small comparing to the signal length, a natural idea is to include the derivatives of the templates into the fitting basis (matrix S), we then have an augmented source matrix whose additional columns are the derivatives of the spectral references. When more information is known about the shifts such as their statistical distributions, then we should include this information into the approach. Under this assumption, we consider two scenarios both of which has random shifts in source spectral lines. In the first case, the random shifts from source spectra are assumed to be independent and identically distributed. In the second scenario, the random shifts are assume to have serial correlations (the shifts in mixtures from one source follow an autoregressive model), which cause correlations between mixtures. In the first case, we solve the problem in a sequential manner, i.e., treat one observation at a time to achieve the mixing coefficients of the source for that particular observation. While in the second case the observations can not be treated separately because the shifts are serially correlated (the mixtures are coupled all together).

The paper is structured as follows; In section 2, we present the methods and their mathematical details. In section 3, numerical investigations are performed to validate the methods proposed, in addition comparisons between the methods are provided; Concluding remarks and future works are given in section 4.

2 The Methods

We are interested in modeling the shift and compression (or expansion) effects (as shown in Fig. 2.5) occurring in sensing, and build them into data matching algorithms. Let us consider modeling shift effect by writing each row of $\boldsymbol{X}$ as $\boldsymbol{x}_i(\nu), i = 1, \cdots, m$, ν can be wavelength, frequency, or time depending on physical origins of the

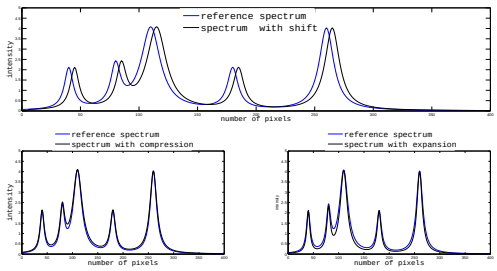


Fig. 1 Shift, compression, and expansion in signals.

problems. It gives the row entries when discretized. In the model, each source signal may have different shifts in different mixtures. Let the shift on row $\boldsymbol{s_j} = \boldsymbol{s_j}(\nu)$ of $\boldsymbol{S}$ be denoted by ξ_{ij} which means the shift of source $\boldsymbol{s_j}$ in mixture $\boldsymbol{x_i}$. The nonlinear mixing model with shift adjustment is $\boldsymbol{x}_i = \sum_{j=1}^{n} \boldsymbol{a}_{ij}\boldsymbol{s}_j(\nu + \xi_{ij}) + \boldsymbol{N}_i$, and the related minimization problem is

$$(\boldsymbol{a}_{ij}, \xi_{ij}) = \arg\min \|\boldsymbol{x}_i - \sum_{j=1}^{n} \boldsymbol{a}_{ij}\boldsymbol{s}_j(\nu + \xi_{ij})\|_2 \ , \qquad (2.1)$$

and it is non-convex. A similar problem was studied in [7] based on alternating least squares estimation with Levenberg-Marquart iterative method. In the context of image registration, the author in [5] propose a convex approximation of (2.1). Both two methods assume that each source signal has the same shift in all mixtures. In practice, shifts of the same source may be different in different mixtures, and this is a more difficult and complicated problem.

2.1 Augmented Least Squares

The idea of augmented least squares is to fit the reference spectra with their derivatives to the observations. Given that the shifts are small comparing with the signal length, we shall use a truncated Taylor expansion of $s_j\nu + \xi_{ij}$ as an approximation,

$$\boldsymbol{s}_j(\nu + \xi_{ij}) = \boldsymbol{s}_j(\nu) + \xi_{ij}\boldsymbol{s}'_j(\nu) + \frac{1}{2}\xi_{ij}^2\boldsymbol{s}''_j(\nu) + \cdots \ , \quad (2.2)$$

given the existence of the derivatives of $\boldsymbol{s}_i$. In practice, it usually uses the first several terms as an approximation, in many tested cases, the first and second derivatives work well, more terms may be included as needed. Knowing that the each mixture actually is linear combination of all the sources and their derivatives, we include both the reference spectra and their derivatives to a form an augmented fitting basis as

$$\tilde{\boldsymbol{S}} = [\boldsymbol{s_1}; \boldsymbol{s'_1}; \boldsymbol{s_2}; \boldsymbol{s'_2}; \cdots; \boldsymbol{s_n}; \boldsymbol{s'_n}]$$

where only the first derivatives are included. We will then solve the following augmented least squares problem (AgLS)

$$\min_{\tilde{\boldsymbol{A}}} \|\boldsymbol{X} - \tilde{\boldsymbol{A}}\tilde{\boldsymbol{S}}\|_2^2 \ , \qquad (2.3)$$

for the linear mixture model

$$\boldsymbol{X} = \tilde{\boldsymbol{A}}\,\tilde{\boldsymbol{S}} + \boldsymbol{N} \qquad (2.4)$$

The solution is $\tilde{\boldsymbol{A}} = \boldsymbol{X}\tilde{\boldsymbol{S}}^{\mathrm{T}}(\tilde{\boldsymbol{S}}\tilde{\boldsymbol{S}}^{\mathrm{T}})^{-1}$, here $\tilde{\boldsymbol{A}}$ contain weights of all the source signals and their derivatives.

Besides modeling the shift effect, augmented least squares is able to handle compression/expansion effects occurring in source signals for data matching. In fact, let the compression/expansion scale on each source signal $\boldsymbol{s}_j$ be denoted by v_j whose value is close to 1. The nonlinear mixing model with distortion adjustment is:

$$\boldsymbol{x}_i = \sum_{j=1}^{n} a_{ij}\boldsymbol{s}_j(v_j\nu) + \boldsymbol{N}_i \ , \text{where } i = 1, \cdots, m \ , \qquad (2.5)$$

and the related minimization problem is

$$(\boldsymbol{A}, \boldsymbol{v}) = \arg\min \|\boldsymbol{x}_i - \sum_{j=1}^{n} a_{ij}\boldsymbol{s}_j(v_j\nu)\|_2 \ . \qquad (2.6)$$

This minimization is non-convex. Using the same idea as the augmented least squares, let $v_j = 1 + \delta_j$, then

$$\boldsymbol{s}_j(v_j\nu) = \boldsymbol{s}_j(\nu + \delta_j\nu) = \boldsymbol{s}_j(\nu) + \boldsymbol{s}'_j(\nu)\delta_j\nu + \cdots \ .$$

Then the rest follows the equations (2.3), (2.4).

2.2 Augmented Maximum Likelihood Estimators

In this augmented least squares method, no prior information such as the statistical distributions, means, etc., of the random shifts are assumed. Hence it is a rather general approach and can be applied in many situations such as when the knowledge of the random shifts are limited. If certain statistics about the shifts are known a priori, we should take into account the information.

Let $\xi_{1i}, \xi_{2i}, \cdots, \xi_{mi}$ be the random shifts from the *ith* source signal in all the mixtures. We rewrite model (2.4) by including the source spectral and their first derivatives along with random shifts variables.

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} + \big(\boldsymbol{A} \odot \boldsymbol{\Xi}\big)\boldsymbol{S'} + \boldsymbol{N} \ , \qquad (2.7)$$

where matrix $\boldsymbol{S'}$ contains the first derivatives of entries of $\boldsymbol{S}$, and columns of $\boldsymbol{\Xi}$ are vectors $[\xi_{1i}, \xi_{2i}, \cdots, \xi_{mi}]^{\mathrm{T}}, i =$

$1, \cdots, n$. Symbol $\odot$ means the element-wise multiplication. Below we have the model in form of matrix elements,

$$x_{ij} = \sum_{k=1}^{n} a_{ik}s_{kj} + \sum_{k=1}^{n} a_{ik}\xi_{ik}s'_{kj} + n_{ij} \ . \tag{2.8}$$

Clearly, the randomness and uncertainty of the observations include two parts: noises $\boldsymbol{N}$ and the random shift $\big(\boldsymbol{A} \odot \boldsymbol{\Xi}\big)\boldsymbol{S'}$. Then the mixture matrix $\boldsymbol{X}$ has mean $\boldsymbol{AS}$ and follow the same distribution as $\big(\boldsymbol{A} \odot \boldsymbol{\Xi}\big)\boldsymbol{S'} + \boldsymbol{N}$.

Let $\Gamma_k \in \mathbb{R}^{m\times p}$ with $\Gamma_{k,ij} = a_{ik}s'_{kj}, k = 1, \cdots, n$. Notice that $\Gamma_k = \Gamma_k(A)$ depends on the unknown mixing matrix A, then we have

$$x_{ij} = \sum_{k=1}^{n} a_{ik}s_{kj} + \sum_{k=1}^{n} \Gamma_{k,ij}\xi_{ik} + N_{ij} \tag{2.9}$$

Below we discuss two cases according to the correlations among the mixtures.

2.2.1 Statistical Model Case I: Heteroscedasticity

Assume that ξ_{ik} are independent and identical Gaussian $\mathcal{N}(0, \sigma_k^2), k = 1, \cdots, n$, while $N_{ij} \sim \mathcal{N}(0, \tau^2)$, and ξ_{ik} and N_{ij} are independent. Clearly, we have the expectation of x_{ij}, $\mathbf{E}(x_{ij}) = \sum_{k=1}^{n} a_{ik}s_{kj}$, and the covariance between entries of $\boldsymbol{X}$ are

$$\begin{aligned}\mathbf{Cov}(x_{ij}, x_{i'j'}) &= \mathbf{Cov}(\sum_{k=1}^{n} \Gamma_{k,ij}\xi_{ik}, \sum_{k=1}^{n} \Gamma_{k,i'j'}\xi_{ik}) \\ &= \sum_{k=1}^{n} \Gamma_{k,ij}\Gamma_{k,i'j'}\sigma_k^2 \ , j \neq j' \\ &= \sum_{k=1}^{n} a_{ik}^2 s'_{kj}s'_{kj'}\sigma_k^2.\end{aligned}$$

And we know $\mathbf{Cov}(x_{ij}, x_{i'j'}) = 0$, if $i \neq i'$ since different rows of matrix $\boldsymbol{X}$ are uncorrelated. However, elements within each rows are correlated, for if we rewrite the linear model (2.4) in terms of their rows

$$\boldsymbol{x}_i = \sum_{j=1}^{n} a_{ij}\boldsymbol{s}_j + \sum_{j=1}^{n} a_{ij}\xi_{ij}\boldsymbol{s'}_j + \boldsymbol{N}_i \ , \tag{2.10}$$

apparently the ith row (observation) $\boldsymbol{x}_i$ is independent to the k-th row $\boldsymbol{x}_k$ for $i \neq k$.

Let $\mathrm{vec}(\boldsymbol{X}) \in \mathcal{R}^{mp\times 1}$ be the vectorization of the mixture matrix $\boldsymbol{X}$. Note that $\boldsymbol{X} = [\boldsymbol{x}_1^{\mathrm{T}}, \boldsymbol{x}_2^{\mathrm{T}}, \ldots, \boldsymbol{x}_m^{\mathrm{T}}]^{\mathrm{T}}$ with $\boldsymbol{x}_i$ being a row vector of length p, we shall stack rows of $\boldsymbol{X}$ to form the $\mathrm{vec}(\boldsymbol{X})$, that is

$$\mathrm{vec}(\boldsymbol{X}) = [x_{11}, x_{12}, \cdots, x_{1p}, \cdots, x_{m1}, \cdots, x_{mp}]^{\mathrm{T}}. \tag{2.11}$$

Then $\mathrm{vec}(\boldsymbol{X}) \sim \mathcal{N}_{mp}(\mathrm{vec}(AS), \boldsymbol{V})$, where $\mathcal{N}_{mp}$ denotes $mp-$variate normal distribution, and

$$\boldsymbol{V} = \tau^2\boldsymbol{I}_{mp\times mp} + \sum_{k=1}^{n} \Omega_k\sigma_k^2 \ = \begin{pmatrix} \boldsymbol{V_1} & & \mathrm{O} \\ & \ddots & \\ \mathrm{O} & & \boldsymbol{V_m} \end{pmatrix}_{mp\times mp}$$

$\boldsymbol{\Omega_k} \in \mathcal{R}^{mp\times mp} = \mathrm{diag}\Big((\boldsymbol{\Gamma}_k)_{1,:}^{\mathrm{T}}(\boldsymbol{\Gamma}_k)_{1,:}, \cdots, (\boldsymbol{\Gamma}_k)_{m,:}^{\mathrm{T}}(\boldsymbol{\Gamma}_k)_{m,:}\Big)$, here $(\boldsymbol{\Gamma}_k)_{i,:} \in \mathcal{R}^{1\times p} = a_{ik}\boldsymbol{s'}_{k,:}$. Hence $\boldsymbol{\Omega_k}$ is a block diagonal matrix, and $\boldsymbol{V_i}$ is $p \times p$ matrix.

The log-likelihood function is

$$\begin{aligned}\mathcal{L}(A, \sigma^2, \tau^2|\boldsymbol{X}, \boldsymbol{S}) = &-\frac{1}{2}\ln\det(\boldsymbol{V}) - \frac{1}{2}\ln(2\pi) \\ &- \frac{1}{2}\Big(\mathrm{vec}(\boldsymbol{X} - \boldsymbol{AS})^{\mathrm{T}}\boldsymbol{V}^{-1}\mathrm{vec}(\boldsymbol{X} - \boldsymbol{AS})\Big)\end{aligned}$$

Notice that $\boldsymbol{V} = \boldsymbol{V}(\sigma^2, A)$ is determined by σ^2 and A. If $\boldsymbol{V}, \sigma^2$, and τ^2 are known, the maximized likelihood estimator (AgMLE) of $\boldsymbol{A}$ is the solution of m generalized least squares, where $\hat{\boldsymbol{A}_{i,:}}$, the AgMLE of $\boldsymbol{A_{i,:}}$ (a row vector containing the weights of source signals in the ith mixture) for the ith mixture is

$$\hat{\boldsymbol{A}_{i,:}} = \arg\min \|\boldsymbol{V_i}^{-1/2}(\boldsymbol{X_{i,:}} - \boldsymbol{A_i}\boldsymbol{S})^{\mathrm{T}}\|_2^2 \ , \tag{2.12}$$

and it has a closed form $\hat{\boldsymbol{A}_{i,:}} = \boldsymbol{X_{i,:}}\boldsymbol{V_i^{-1}}\boldsymbol{S^{\mathrm{T}}}(\boldsymbol{S}\boldsymbol{V_i^{-1}}\boldsymbol{S^{\mathrm{T}}})^{\boldsymbol{-1}}$, and $\mathbf{Var}(\hat{\boldsymbol{A}}_{i,:}) = (\boldsymbol{S}\boldsymbol{V_i^{-1}}\boldsymbol{S^{\mathrm{T}}})^{\boldsymbol{-1}}$. We propose the following iterative approach for AgMLE of $(\boldsymbol{A}, \sigma^2, \tau^2)$,

1, Start with an initial $\boldsymbol{A^{(0)}}$ obtained by the ordinary least squares, $\min \|\boldsymbol{X} - \boldsymbol{AS}\|_2^2$ whose solution is $\boldsymbol{A^{(0)}} = \boldsymbol{X}\boldsymbol{S^{\mathrm{T}}}(\boldsymbol{S}\boldsymbol{S^{\mathrm{T}}})^{\boldsymbol{-1}}$. Then $\Gamma_k^{(0)}$ whose entry is $\Gamma_{k,ij}^{(0)} = A_{ik}^{(0)}S_{kj}$.

2, Obtain $\boldsymbol{\xi} = (\xi_{ik})$ by solving

$$(\xi_{ik}) = \arg\min \|(\boldsymbol{X} - \boldsymbol{A^{(0)}}\boldsymbol{S})_{i,:} - \sum_{k}(\Gamma_k)_{i,:}\xi_{i,k}\|_2^2 \ .$$

Then $\sigma_k^{(0)} = \sqrt{\sum_{i=1}^{m}\xi_{ik}^2}/m$, $\tau^0 = \frac{1}{\sqrt{mp}}\|(\boldsymbol{X} - \boldsymbol{A^{(0)}}\boldsymbol{S})_{i,:} - \sum_{k}(\Gamma_k)_{i,:}\xi_{i,k}\|_2$; $\boldsymbol{V}_i^{(0)} = \sum_{k=1}^{n}\boldsymbol{\Gamma}_k)_{i,:}^{\mathrm{T}}(\boldsymbol{\Gamma}_k)_{i,:}(\sigma_k^{(0)})^2$.

3, Update $\boldsymbol{A}^{(0)}$ by

$$\boldsymbol{A}_{i,:}^{(1)} = \arg\min \|\boldsymbol{V_i}^{-1/2}(\boldsymbol{X_{i,:}} - \boldsymbol{A_i^{(0)}}\boldsymbol{S})^{\mathrm{T}}\|_2^2 \ .$$

4, Set $\boldsymbol{A}^{(0)} \leftarrow \boldsymbol{A}^{(1)}$ and iterate step 2-3 until it converges.

2.2.2 Statistical Model Case II: Autoregressive Model

Here we consider a scenario where the mixtures are correlated. The correlation might due to many factors such as the dispersion of the chemical compounds examined which is generally a function of time, hence shift from one source existing in the observation at current time will have influence on the shift in the mixture of future experiments. Suppose the observations are associated with time, meaning that we shall have a time series of mixtures. Then we assume that the correlation between adjacent time the random shifts are stronger than that between shifts over longer time. Under such condition, we propose to use autoregressive model AR(1) to characterize the random shifts exhibited in the mixtures. We still assume an iid noise $N_{ij} \sim \mathcal{N}(0,\tau^2), i = 1,\cdots,m, j = 1,\cdots,p$. For any source $k = 1,\cdots,n$, we assume $\xi_{1k},\xi_{2k},\cdots,\xi_{mk}$, the shifts of source k in all mixtures, are correlated according to

$$\xi_{ik} = \rho_k \xi_{i-1,k} + u_{ik}\ , |\rho_k| < 1 \tag{2.13}$$

with $u_{ik} \sim \mathcal{N}(0,\sigma_k^2), \xi_{0,k} = 0$. This implies that $\mathbf{Var}(\xi_{ik}) = \frac{\sigma_k^2}{1-\rho_k^2}, \mathbf{Corr}(\xi_{ik},\xi_{i-t,k}) = \rho_k^t$. We also assume that ξ_{ik} and N_{ij} are independent, furthermore, $(\xi_{1k},\cdots,\xi_{mk}) \perp (\xi_{1k'},\cdots,\xi_{mk'})$ for $k \neq k'$. Following the model (2.9), we have

$$\begin{aligned}
\mathbf{Var}(x_{ij}) &= \mathbf{Var}(\sum_{k=1}^{n}(\Gamma_{k,ij}\xi_{ik}) + \tau^2\ , \\
&= \sum_{k=1}^{n}\Gamma_{k,ij}^2\frac{\sigma_k^2}{1-\rho_k^2} + \tau^2\ , \\
\mathbf{Cov}(x_{ij},x_{i'j'}) &= \mathbf{Cov}(\sum_{k=1}^{n}(\Gamma_{k,ij}\,\xi_{ik}, \sum_{k=1}^{n}\Gamma_{k,i'j'}\,\xi_{i'k}) \\
&= \sum_{k=1}^{n}\mathbf{Cov}(\Gamma_{k,ij}\,\xi_{ik}, \Gamma_{k,i'j'}\,\xi_{i'k}) \\
&= \sum_{k=1}^{n}\Gamma_{k,ij}\Gamma_{k,i'j'}\frac{\sigma_k^2}{1-\rho_k^2}\rho_k^{|i-i'|}\ , j \neq j'\ .
\end{aligned}$$

We take the vectorization of the mixture matrix $\boldsymbol{X}$ as defined in (2.11), then $\mathrm{vec}(\boldsymbol{X}) \sim \mathcal{N}_{mp}(\mathrm{vec}(AS),\boldsymbol{V})$, and

$$\boldsymbol{V} = \tau^2\boldsymbol{I}_{mp\times mp} + \sum_{k=1}^{n}\Omega_k\frac{\sigma_k^2}{1-\rho_k}\ ,$$

where

$$\boldsymbol{\Omega_k} = \begin{pmatrix} \boldsymbol{\Omega_{k,11}} & \boldsymbol{\Omega_{k,12}} & \cdots & \boldsymbol{\Omega_{k,1m}} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Omega_{k,m1}} & \boldsymbol{\Omega_{k,m2}} & \cdots & \boldsymbol{\Omega_{k,mm}} \end{pmatrix}_{mp\times mp}\ , \tag{2.14}$$

here $\boldsymbol{\Omega_{k,ij}} = (\boldsymbol{\Gamma}_k)_{i,:}^{\mathrm{T}}(\boldsymbol{\Gamma}_k)_{j,:}\rho_k^{|i-j|}$ which is no longer a diagonal matrix. We write

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V_{11}} & \boldsymbol{V_{12}} & \cdots & \boldsymbol{V_{1m}} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{V_{m1}} & \boldsymbol{V_{m2}} & \cdots & \boldsymbol{V_{mm}} \end{pmatrix}_{mp\times mp} \tag{2.15}$$

where $\boldsymbol{V_{ij}} \in \mathbb{R}^{p\times p} = \sum_{k}^{n}(\boldsymbol{\Gamma}_k)_{i,:}^{\mathrm{T}}(\boldsymbol{\Gamma}_k)_{j,:}\rho_k^{|i-j|}\frac{\sigma_k^2}{1-\rho_k} + \tau^2\boldsymbol{I}_{p\times p}\mathbb{I}(i=j)$, where $\mathbb{I}$ is the indicator function.

The log-likelihood function takes the same form as (2.12). We propose the following iterative algorithm for estimating the parameters,

1, Start with an initial $\boldsymbol{A^{(0)}}$ obtained by the ordinary least squares.

2, Obtain $\boldsymbol{\xi} = (\xi_{ik})$ by solving

$$\hat{\xi}_{ik} = \arg\min \|(\boldsymbol{X} - \boldsymbol{A^{(0)}}\boldsymbol{S})_{i,:} - \sum_{k}(\Gamma_k^{(0)})_{i,:}\xi_{i,k}\|_2^2\ .$$

Then we regress $\hat{\xi}_{ik}$ on $\hat{\xi}_{i-1,k}$ for each k and we obtain $\rho_k^{(0)} = \frac{\sum_{i=2}^{m}\hat{\xi}_{ik}\hat{\xi}_{i-1,k}}{\sum_{i=2}^{m}\hat{\xi}_{i-1,k}^2}$, and $\sigma_k^{(0)} = \frac{\sqrt{\sum_{i=1}^{m}\xi_{ik}^2}}{m}$, $\tau^0 = \frac{1}{\sqrt{mp}}\|(\boldsymbol{X} - \boldsymbol{A^{(0)}}\boldsymbol{S})_{i,:} - \sum_{k}(\Gamma_k)_{i,:}\xi_{i,k}\|_2$.

3, $\boldsymbol{V}_{ij}^{(0)} = \sum_{k=1}^{n}\boldsymbol{\Gamma}_k)_{i,:}^{\mathrm{T}}(\boldsymbol{\Gamma}_k)_{j,:}[\rho_k^{(0)}]^{|i-j|}\frac{(\sigma_k^{(0)})^2}{1-\rho_k^{(0)}} + (\tau^{(0)})^2\boldsymbol{I}_{p\times p}\mathbb{I}(i=j)$. Then update $\boldsymbol{A}^{(0)}$ by

$$\boldsymbol{A}^{(1)} = \arg\min \|\boldsymbol{V^{(0)}}^{-1/2}\mathbf{Vec}(\boldsymbol{X} - \boldsymbol{A^{(0)}}\boldsymbol{S})^{\mathrm{T}}\|_2^2\ .$$

4, Set $\boldsymbol{A}^{(0)} \leftarrow \boldsymbol{A}^{(1)}$ and iterate step 2-3 until it converges.

3 Experimental Investigation

We report here the numerical examples and results of synthetic data as well as realistic data. Note that the methods require certain smoothness of the data to guarantee the continuity of the derivatives, and this condition is satisfied approximately by the test data. We also show here the comparisons with the perturbed linear mixture modeling (PLMM) proposed in [12]. Here are some essentials of the their method. Let $\boldsymbol{\delta}_{ij}$ be the perturbation of the jth source in the ith mixture, then for each mixed signal the PLMM model takes the following form

$$\boldsymbol{x}_i = \sum_{j=1}^{n}a_{ij}\cdot(\boldsymbol{s}_j + \boldsymbol{\delta}_{ij}) + \boldsymbol{N}_i\ , \text{for}\quad i = 1,\ldots,m\ , \tag{3.1}$$

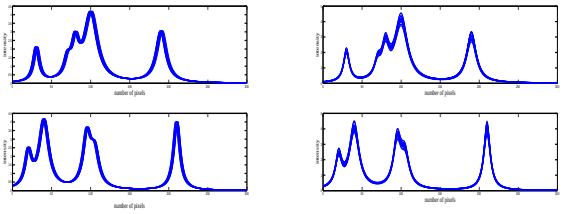


Fig. 2 Reference spectra (red lines) and 15 corresponding instances (blue) under spectral distortions. Left: random shifts; right: unknown perturbations.

and the cost function is

$$\mathcal{J}(a_{ij}, \boldsymbol{\delta}_{ij}) = \|\boldsymbol{x}_i - \sum_{j=1}^{n} a_{ij} \cdot (\boldsymbol{s}_j + \boldsymbol{\delta}_{ij})\|_2^2. \tag{3.2}$$

Note that $\boldsymbol{\delta}_{ij}$ is a vector of the same size of $\boldsymbol{s}_j$, the symbol $\cdot$ means multiplication. In [12], some other penalty terms are added to the cost function, and the alternating direction method of multipliers is used to solve this non-convex optimization. Here we shall use alternating least squares to obtain a solution.

3.1 Synthesized examples

In this part, we show the results of simulated examples for all the three scenarios: the augmented least squares (AgLS), heteroscedasticity, and autoregressive model. There are 2 sources, and 100 observations. The source spectra are shown in the left plot of Fig. 2. The estimation of the parameters are presented below and in the Figs. 3-5. The comparisons with ordinary least squares (OLS) are also shown in the figures. It can be seen that the result of the AgLS are better than OLS, but are less accurate than the two other methods when the statistics of the shifts are known. However, the AgLS is a rather general method and computationally less complicated and expensive. We also test the effectiveness of AgLS on linear compression and expansion for signals in Fig. 2.5. The results are shown in Fig. 6, where the augmented least squares is able to deal with mild linear compression or expansion.

We present here the comparison between our method AgLS and PLMM proposed in [12]. First we test the two methods on the signals with random shifts (the left plot of Fig. 2), and the result is depicted in Fig. 7. The AgLS method outperforms PLMM and this is expected since the data are known *a prior* to have random shifts. In the case of spectral distortion being additive perturbation (the right plot from Fig. 2), PLMM method delivers a slightly better result than AgLS as shown in Fig. 8.

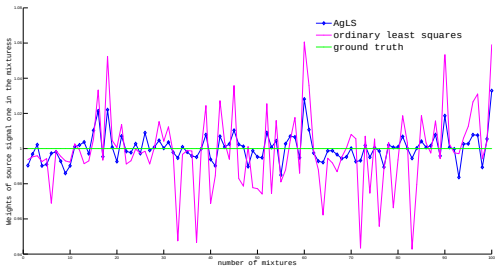


Fig. 3 The result of the augmented least squares (AgLS), and its comparison to the ordinary least squares. The stand deviations of the random shifts of source signals are $\sigma_1 = \sigma_2 = 1$. The standard deviation of the noise is 0.05, the ground truth of the mixing coefficients (weights) of source 1 is one.

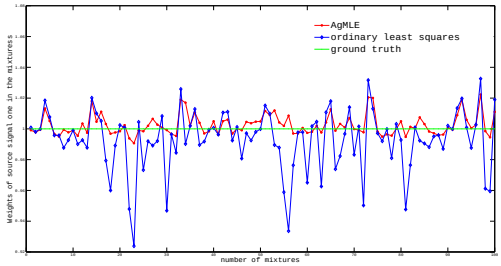


Fig. 4 Heteroscedasticity: the result of the augmented maximum likelihood estimator (AgMLE), and its comparison to the ordinary least squares. The stand deviations of the random shifts of source signals are $\sigma_1 = 1$ and $\sigma_2 = 1$, whose estimations are $\hat{\sigma}_1 = 1.0722, \hat{\sigma}_2 = 0.7617$. The standard deviation of the noise is $\tau = 0.05$ and its estimation is $\hat{\tau} = 0.1806$, the ground truth of the mixing coefficients of source 1 is one.

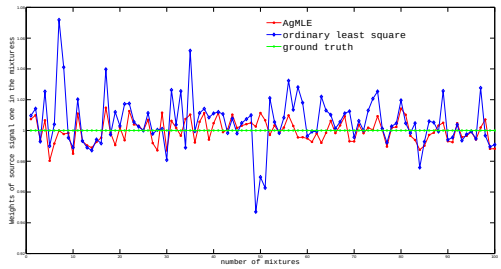


Fig. 5 Autoregressive model: the result of AgMLE for serial correction between mixtures, and its comparison to the ordinary least squares. The parameters used in AR(1) model are $\sigma_1 = 1, \sigma_2 = 1, \tau = 0.05, \rho_1 = 0.5, \rho_2 = 0.4$. Their estimations are $\hat{\sigma}_1 = 0.9714, \hat{\sigma}_2 = 1.0616, \hat{\rho}_1 = 0.6168, \hat{\rho}_2 = 0.3887, \hat{\tau} = 0.2367$.

3.2 Data from NMR, Raman, and DOAS spectroscopy

1, Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful and popular tool for chemists and biochemists to investigate and determine the structures and properties of molecules. In the example of

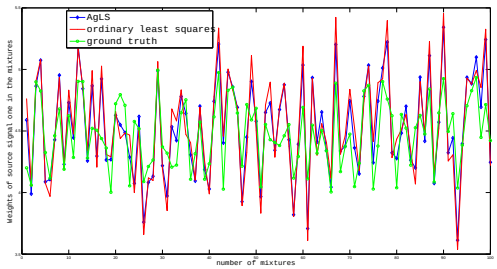


Fig. 6 The results of AgLS on linear compression effect and comparisons to the ordinary least squares. Here the compression/expansion parameters v_j are in the range (0.80, 1.20), the ground truth of mixing matrix $\boldsymbol{A}$ are from uniform distribution of [4,5]. The absolute fitting error for the augmented least squares is 3.147, for the ordinary least squares is 3.632.

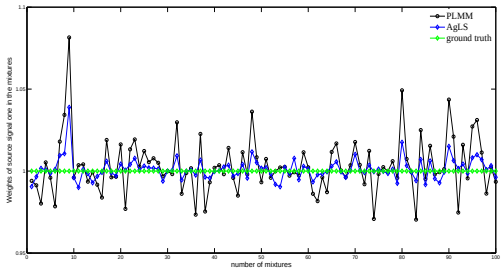


Fig. 7 Comparisons of AgLS with PLMM for signals with random shifts.

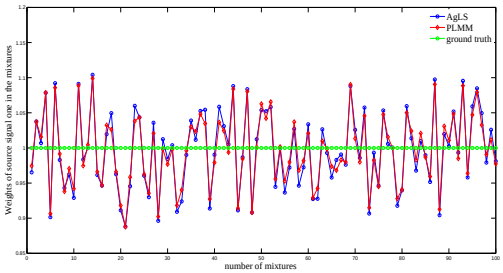


Fig. 8 Comparisons of AgLS with PLMM for signals with spectral distortions from the right plot in Fig. 2.

NMR data, we use the true spectra of four chemical compounds, mannitol, $\beta-$cyclodextrine, $\beta-$sitosterol, and menthol (see Fig. 9 for their spectra). The coefficients of the mixing matrix A are generated from a uniform distribution in the range of [5,10] for each mixture. Then we obtain the mixtures by simulating the mixing process in (1.1) with randomly shifted source spectra and by adding Gaussian noise. The results of the estimations are listed in Figs. 10,11. It once again shows that the methods proposed are better than the ordinary least squares.

2, Raman spectroscopy. In a second example, we apply the methods to the Raman spectra. As shown in Fig. 12, a Raman spectrum gives a collection of peaks that correspond to the characteristic vibrational frequencies of the materials being examine, thus providing unique information on the molecular structure and chemical composition of the matters. We test our methods on Raman data provided by NSF/ATD program [1]. The sample consists of several liquid substances, some of them are commonly used for making explosives. The dataset include 21 mixed Raman spectra at different incident laser wavelengths. Note that we do not have the ground truth of the mixing coefficients, moreover we do not have knowledge of the statistics of the random shifts. In this situation, the augmented least squares clearly is the most suitable method. As a comparison we show the comparisons to the case 2 where IID Gaussian distributed shifts are assumed. We also compare our methods to the nonlinear least solver from [7] The results of augmented least squares are close to those of nonlinear least squares (see Fig. 13), although the augmented least squares is less complicated and easier to implement. On the other hand, the augmented maximum likelihood results are much more deviated from other methods for the mixtures number 1-15 (Fig. 14). This observation shows that shifts are not Gaussian or the existence of other nonlinear distortion such as compression or expansion in the signals.

3, Differential optical absorption spectroscopy (DOAS) is based on the light absorption property of matter to identify broadband and narrow band spectral structures, and analyze atmospheric trace gases concentrations. Fig. 15 shows spectra reference of trace gases HONO and NO_2 and a DOAS spectrum of mixture sample containing these two gases. It is found that the augmented least squares is able to achieve better results comparing to the ordinary least squares, on the other hand, the augmented maximum likelihood estimator (heteroscedasticity) produces same results as the ordinary least squares, in fact, the matrix $\boldsymbol{V}$ is actually diagonally dominant, and the shift effect is insignificant. The better results of augmented least squares may imply that there are other spectral distortions such as linear compression or expansion. This once again shows that the augmented least squares are able to handle not only the shift but also other distortions in the spectra.

4 Conclusion and Future Works

In this paper, we deal with data matching where the source signals have random shifts and/or other nonlinear distortions. The modeling of these distortions is based on the truncated Taylor expansion of the signals. The computational approach is an augmented least squares which fits the reference spectra of the signals and their

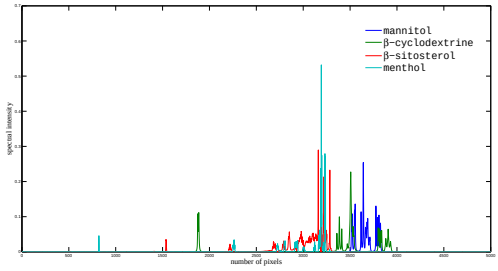


Fig. 9 The reference spectra of mannitol, $\beta-$cyclodextrine, $\beta-$sitosterol, and menthol.

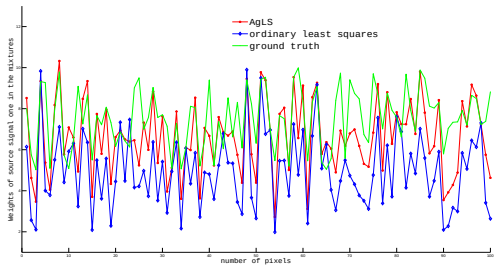


Fig. 10 Augmented least squares: the estimation of the weights of mannitol in the 100 mixtures.

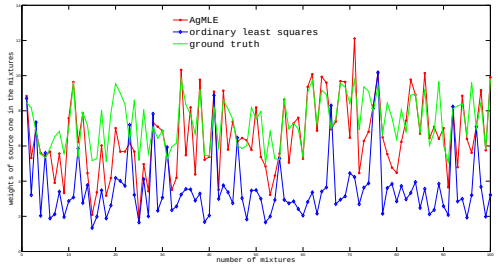


Fig. 11 AgMLE: the estimation of the weights of mannitol in the 100 mixtures. The parameters used in the example: $\sigma_i = 2, i = 1, \cdots, 4, \tau = 0.05$. The computed results of the parameters are $\hat{\sigma}_1 = 0.9353, \hat{\sigma}_2 = 1.0264, \hat{\sigma}_3 = 0.9229, \hat{\sigma}_4 = 1.0766, \hat{\tau} = 0.1940$.

derivatives to the mixtures. If the random distortions are Gaussian, an augmented maximum likelihood estimator (AgMLE) is developed. The methods prove to work well with real-world data from NMR and Raman as well DOAS spectroscopy. In future work, we shall study non-Gaussian priors (such as Laplacian or hyper Laplacian) in maximum likelihood estimator to model the shifts and other distortions. For signals with jumps and/or high oscillations, the methods based on derivatives would fail to work. These signals can be found in intensity image of a building with trees and vegetation for image registrations (see Fig. 18). One idea is to model shifting by shift operator that is characterized by sparse matrix of ones and zeros. The unknowns to be

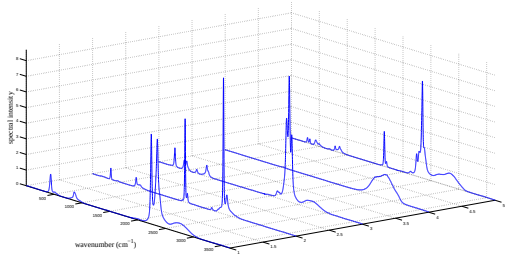


Fig. 12 Raman spectra: from left to right, Methanol, Acetonitrile, Ethanol, Water, and their mixture.

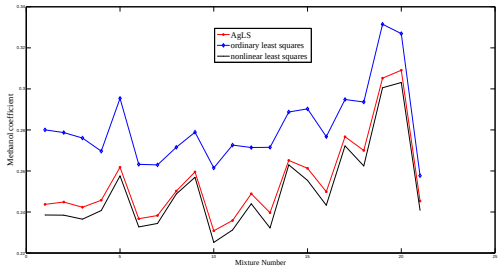


Fig. 13 The estimation of mixing coefficients of Methanols computed by AgLS. As a comparison, we show the results by nonlinear least squares.

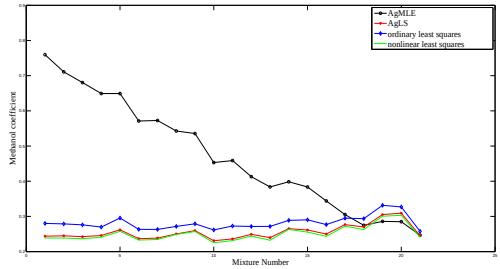


Fig. 14 The estimation of mixing coefficients of Methanols computed by AgMLE and comparison to other methods.

estimated are the mixing matrix and the shift matrix, and they are multiplied together, hence the problem is non-convex. A future work is to study convex approximation of the problems.

Acknowledgments

The authors thank Professor Barbara Finlayson-Pitts and Dr. Lisa Wingen for the experimental DOAS data, and Naval Research Lab for the Raman Data. J. Xin acknowledges support of NSF grant DMS-1222507. Y. Sun thanks support of grant 800006981 from Simons Foundation.

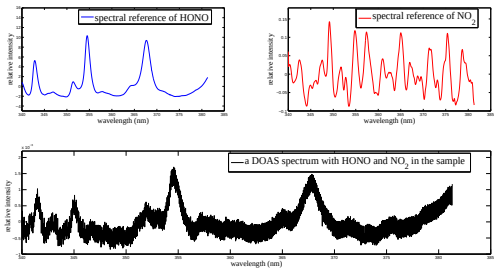


Fig. 15 The spectral references of trace gases HONO and NO_2, and their noisy mixture.

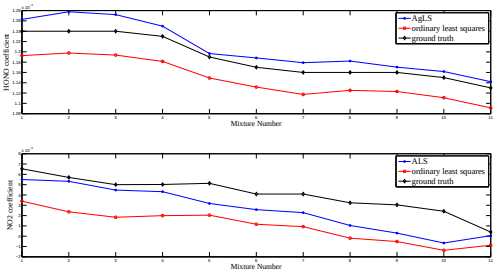


Fig. 16 Coefficients of HONO and NO_2 by AgLS and comparisons to other methods.

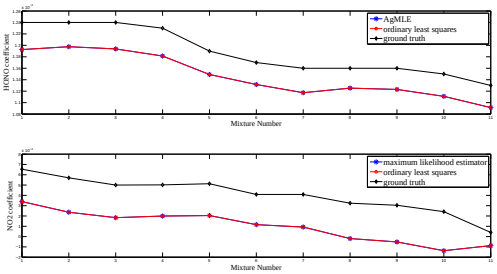


Fig. 17 The coefficients of HONO and NO_2 by AgMLE and comparisons to other methods.

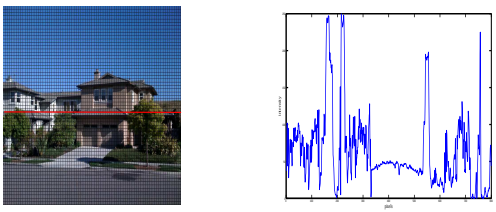


Fig. 18 Signal with jumps and oscillations (right) is a slice (red line) of an image containing buildings and plants.

References

1. http://grassmann.math.colostate.edu/ATD/home.html.
2. Song, C., *Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate endmember variability?*, Remote Sensing Environ., vol. 95, no. 2, pp. 248–263, 2005.
3. A. Zare and K. C. Ho, *Endmember variability in hyperspectral analysis: Addressing Spectral Variability During Spectral Unmixing*, IEEE Signal Processing Mag., vol. 31, no. 1, pp. 95–104, 2014.
4. C-I Chang, ed., *Hyperspectral Data Exploitation: Theory and Applications*, Wiley-InterScience, 2007.
5. E. Esser, *A Convex Model of Image Registration*, UCLA CAM Report, 10-04, 2010.
6. D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, *Robust matched filters for target detection in hyperspectral imaging data*, ICASSP 2007. IEEE International Conference, Vol. 1, pp. I-529-I-532.
7. J. Stutz and U. Platt, *Numerical analysis and estimation of the statistical error of differential optical absorption spectroscopy measurements with least squares methods*, Appl. Optics., **35**, pp. 6041–6053, 1996.
8. D.A. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, R.O. Green, *Mapping Chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models*, Remote Sensing of Environment, 65, pp. 267–279, 1998.
9. Guo, Z., and Osher, S., *Template Matching via ℓ_1 Minimization and Its Application to Hyperspectral Data,* Inverse Problem and Imaging, Vol 5, No. 1, 2011, pp. 19–35.
10. Zhang, Y.J., *Adaptive region-based image fusion using energy evaluation model for fusion decision*, Signal, Image and Video Processing, Vol 1, Issue 3, 2007, pp. 215–223.
11. Tonazzini, A., Savino, P., and Salerno, E., *A non-stationary density model to separate overlapped texts in degraded documents*, Signal, Image and Video Processing, Vol. 9, Suppl. 1, 2015, pp. 155–164.
12. Thouvenin, P-A., Dobigeon, N., Tourneret, J-Y., *Hyperspectral unmixing with spectral variability using a perturbed linear mixing model*, 23^{rd} European Signal Processing Conference (EUSIPCO), Nice, France, Aug.31-Sep.4, 2015, pp. 819–823.