Agent Identification Using a Sparse Bayesian Model

Huiping Duan, Hongbin Li, *Senior Member, IEEE*, Jing Xie, Nicolai S. Panikov, and Hong-Liang Cui, *Member, IEEE*

***Abstract*—Identifying agents in a linear mixture is a fundamental problem in spectral sensing applications including chemical and biological agent identification. In general, the size of the spectral signature library is usually much larger than the number of agents really present. Based on this fact, the sparsity of the mixing coefficient vector can be utilized to help improve the identification performance. In this paper, we propose a new agent identification method by using a sparse Bayesian model. The proposed iterative algorithm takes into account the nonnegativity of the abundance fractions and is proved to be convergent. Numerical studies with a set of ultraviolet (UV) to infrared (IR) spectra are carried out for demonstration. The effect of the signature mismatch is also studied using a group of terahertz (THz) spectra.**

***Index Terms*—Agent identification, false alarm, linear mixture, mismatch, signature, sparse Bayesian model, spectral sensing.**

I. Introduction

DETECTION and identification of components from mixtures have been studied in various fields and applications, such as blind source separation for speech recognition [1], spectral unmixing in hyperspectral sensing [2], [3], agent detection with Raman spectroscopy [4], or fluctuation enhanced sensing (FES) [5], [6], and so on. Classical detection algorithms, such as the Matched Subspace Detector (MSD) [7], require the knowledge of the noise and interference characteristics in terms of their probability density functions (pdfs) or statistics estimated from the received data. However, in many practical scenarios, the background information is often unavailable or difficult to estimate, making these detection methods inapplicable. Canonical correlation analysis [8] and the non-negative constrained least squares (NCLS) algorithm [5], [6] have been proposed for detection without using *a priori* knowledge of the background interference.

Due to its simplicity, the linear mixture model is often assumed in identification problems [2], [5], [6], [8]

$$\mathbf{t} = \mathbf{\Phi}\mathbf{w} + \mathbf{v} \tag{1}$$

where $\mathbf{t} \in \mathbb{R}^m$ is a vector containing observations to be analyzed, $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$ is a matrix with each column representing the spectral signature of a possible target, $\mathbf{w} \in \mathbb{R}^n$ is the mixing coefficient vector consisting of the concentration ratio or abundance fraction of each component in the mixture, and $\mathbf{v} \in \mathbb{R}^m$ is the noise standing for the measurement or modeling error. The objective of identification is to determine what components are in the mixture and, furthermore, estimate the components' abundance fractions.

To solve the agent identification problem, a signature library is often employed that includes signatures of all possible targets. Usually the number of signatures in the library is significantly larger than that composing the real mixture. Based on this fact, we propose to exploit the sparsity of the mixing coefficient vector to help improve the identification performance. The iterative algorithm considers the non-negativity of the abundance fractions and is proved to be convergent. Using a sparse Bayesian model, our proposed method obtains an estimate of the abundance fraction vector $\mathbf{w}$ with only a few nonzero entries, and the zero entries of the estimate can be interpreted as the absence of certain agents. Therefore, the mission of agent identification is accomplished by parameter estimation and conducted in a quite simple and direct way. Meanwhile, conventional agent identification methods [5], [6], [8] yield an estimate of $\mathbf{w}$ that is generally not sparse. They have to select and apply a threshold on the estimate to decide if a component is present or not. The choice of the threshold is difficult due to lack of theoretical analysis.

Experiments using a set of ultraviolet (UV) to infrared (IR) spectra are carried out. Results show excellent estimation accuracy of the abundance fractions and good identification performance. In addition, the influence of signature mismatch is also studied. The terahertz (THz) spectra of a group of bacteria are used for demonstration.

This paper is organized as follows. In Section II, the agent identification scheme using a sparse Bayesian model is described, along with a discussion on how to address the issue of signature mismatch. Section III contains the numerical results. Conclusions are drawn in Section IV.

II. Proposed Agent Identification Scheme

Since the size of the signature library is much larger than the number of agents present in the real sample, the true abundance fraction vector $\mathbf{w}$ is a sparse vector with many zero entries. By assuming $\mathbf{w}$ as a random variable with some prior pdf and incorporating the sparsity constraint into the estimator, a Bayesian approach can be developed with improved estimation accuracy over conventional methods. In the following, we solve the sparse abundance fraction vector in the Bayesian framework, which has been investigated extensively in the machine

Manuscript received January 25, 2011; accepted March 03, 2011. Date of publication March 22, 2011; date of current version August 24, 2011. The associate editor coordinating the review of this paper and approving it for publication was Prof. Kiseon Kim.

H. Duan and J. Xie are with L. C. Pegasus Corporation, Hillside, NJ 07205 USA (e-mail: duan0002@ntu.edu.sg; jingxie@lcpegasus.com).

H. Li is with Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: hli@stevens.edu).

N. S. Panikov is with the Department of Biology, Northeastern University, Boston, MA 02115 USA (e-mail: n.panikov@neu.edu).

H.-L. Cui is with the Department of Physics, Polytechnic Institute of New York University, Brooklyn, NY 11201 USA (e-mail: hcui@poly.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSEN.2011.2130521

learning community [9]–[11]. Henceforth, the Gaussian distribution $p(\mathbf{x}) = |2\pi\boldsymbol{\Sigma}|^{-(1/2)} \exp[-(1/2)(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]$ is written in an abbreviated form as $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

A. Weight Prior

To promote sparsity, a super-Gaussian prior with the following form is adopted in [12]:

$$p(\mathbf{w}) \propto \exp\left(-\sum_{i=1}^{n} |w_i|^q\right) \tag{2}$$

where $q \in [0, 1]$. Given such a prior, the maximum *a posteriori* (MAP) estimator for the abundance fraction vector can be shown to be

$$\hat{\mathbf{w}}_{\mathrm{MAP}} = \arg\min_{\mathbf{w}} \|\mathbf{t} - \boldsymbol{\Phi}\mathbf{w}\|_2^2 + \lambda \sum_{i=1}^{n} |w_i|^q \tag{3}$$

where $\|\cdot\|_2$ represents the l_2 norm and λ is a tradeoff parameter to balance between the degree of sparsity and the data fitting error. As the value of q approaches 0, (3) reduces to the following optimization for the MAP estimator:

$$\min_{\mathbf{w}} \|\mathbf{t} - \boldsymbol{\Phi}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_0 \tag{4}$$

where the l_0 quasi-norm $\|\cdot\|_0$ counts the number of nonzero entries in $\mathbf{w}$. Formulation (4) is a non-convex optimization problem. It can be solved by an exhaustive search which is highly impractical. Instead, when $p = 1$, an alternative optimization is

$$\min_{\mathbf{w}} \|\mathbf{t} - \boldsymbol{\Phi}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1 \tag{5}$$

where $\|\cdot\|_1$ represents the l_1 norm defined as $\|\mathbf{w}\|_1 = \sum_{i=1}^{n} |w_i|$. Note that (5) is convex and can be solved efficiently [13]–[15]. Using (5), the true sparse weight vector can be identified with a high probability under mild conditions, for example, when the amplitudes of the nonzero weights are sufficiently large [16], or the degree of coherence between columns of the matrix $\boldsymbol{\Phi}$ is small enough [17]. Analysis of such conditions and properties of (5) is a major subject in the community of compressive sensing [18], where a theoretic framework exists.

The sparse Bayesian learning of [9]–[11] makes use of a hierarchical weight prior

$$p(w_i) = \int p(w_i|\alpha_i)p(\alpha_i)\mathrm{d}\alpha_i \tag{6}$$

where w_i denotes the ith element of $\mathbf{w}$ and the conditional density $p(w_i|\alpha_i)$ is a zero-mean Gaussian distribution with variance α_i^{-1}. The hyperparameter α_i is the inverse variance (precision) of the Gaussian density function. With the Gamma-distributed hyperprior [9]

$$p(\alpha_i) = \frac{1}{\Gamma(a)} b^a \alpha_i^{a-1} \mathrm{e}^{-b\alpha_i}, \quad \alpha_i > 0 \tag{7}$$

where a is the shape parameter, b is the inverse scale parameter and $\Gamma(a) = \int_0^\infty t^{a-1}\mathrm{e}^{-t}\,\mathrm{d}t$ is the gamma function, the hierarchical prior after carrying out the integration in (6) is a student-t distribution [10]. Selecting various values for a and b will result in special hyperpriors and hence different prior distributions. For example, with $a = 0$ and $b = 0$, the hyperprior turns into a log-uniform distribution: $p(\ln\alpha_i) = \alpha_i p(\alpha_i) = 1/\Gamma(0)$. This is a noninformative hyperprior and employed in the relevance vector machine (RVM) [10] to promote sparsity. With $a = 1$, the hyperprior turns into an exponential distribution, which leads to a Laplacian prior [11]. Compared with the Bayesian model using the Laplacian prior and the l_1-norm constrained minimization (5), which have to tune some parameters during estimation, the Bayesian model using the noninformative hyperprior can estimate all the parameters automatically. Therefore, in the following sections, we choose $a = 0$ and $b = 0$ in the sparse Bayesian model to estimate the abundance fraction vector.

B. Sparse Bayesian Model

In the Bayesian framework, based on a prior distribution which is formulated over the unknown parameters to describe the problem of interest, the posterior distribution can be derived by applying Bayes' rule [19].

First, the noise vector $\mathbf{v}$ in (1) is assumed to be Gaussian distributed with zero mean and variance $\sigma^2\mathbf{I}_m$, where $\mathbf{I}_m$ denotes the m-dimensional identity matrix. A prior over the abundance fraction vector is assumed to be

$$\mathbf{w}|\alpha \sim \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1}) \tag{8}$$

where $\mathbf{w}|\boldsymbol{\alpha}$ means the random vector $\mathbf{w}$ conditioned on the parameter $\boldsymbol{\alpha}$, $\boldsymbol{\alpha} = [\alpha_1\alpha_2\cdots\alpha_n]^T$ contains all hyperparameters and $\mathbf{A} = \mathrm{diag}(\boldsymbol{\alpha})$ is a diagonal matrix indicating the inverse variance (precision) of each mixing coefficient. The likelihood of the observed data can be written as

$$\mathbf{t}|\mathbf{w}, \sigma^2 \sim \mathcal{N}(\boldsymbol{\Phi}\mathbf{w}, \sigma^2\mathbf{I}_m). \tag{9}$$

By Bayes' rule, the posterior over the weight vector can be shown to be (see Appendix A for the derivation)

$$\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{10}$$

where

$$\begin{aligned} \boldsymbol{\mu} &= \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}^T\mathbf{t} \\ \boldsymbol{\Sigma} &= (\sigma^{-2}\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \mathbf{A})^{-1}. \end{aligned} \tag{11}$$

C. Abundance Fraction Vector Estimation

Given $\boldsymbol{\alpha}$ and σ^2, the abundance fraction vector can be estimated by minimizing the Bayesian mean square error (MSE) [20]

$$\begin{aligned} \mathrm{Bmse}(\hat{\mathbf{w}}|\boldsymbol{\alpha}, \sigma^2) &= \mathrm{E}\left(\|\mathbf{w}-\hat{\mathbf{w}}\|^2|\boldsymbol{\alpha}, \sigma^2\right) \\ &= \iint \|\mathbf{w}-\hat{\mathbf{w}}\|^2 p(\mathbf{w}, \mathbf{t}|\boldsymbol{\alpha}, \sigma^2)\,\mathrm{d}\mathbf{w}\mathrm{d}\mathbf{t}. \end{aligned} \tag{12}$$

The solution is well known to be the posterior mean [20]

$$\hat{\mathbf{w}}_{\mathrm{Bmse}} = \mathrm{E}(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2). \tag{13}$$

That is, the optimal Bayesian MSE estimator is the mean of the posterior pdf $p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2)$. With the Gaussian posterior (10), it can be shown easily that the Bayesian MSE estimator is equivalent to the MAP estimator [20]. Therefore, we have

$$\hat{\mathbf{w}}_{\mathrm{Bmse}} = \boldsymbol{\mu} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \sigma^2\mathbf{A})^{-1}\boldsymbol{\Phi}^T\mathbf{t}. \tag{14}$$

Compared with the maximum-likelihood (ML) estimator $\hat{\mathbf{w}}_{\mathrm{ML}} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{t}$, the Bayesian MSE estimator (14) has an extra term related with the hyperparameters so it depends on both the data and the prior knowledge. This is expected to be able to improve the estimation accuracy.

D. Hyperparameters Estimation

The hyperparameters need to be determined before calculating the optimal Bayesian MSE estimator (14). One way to find the hyperparameters is to use the type-II maximum likelihood [9].

Specifically, the marginal likelihood is (see Appendix A for the derivation)

$$\mathbf{t}|\boldsymbol{\alpha}, \sigma^2 \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{t}}) \tag{15}$$

where $\mathbf{\Sigma}_{\mathbf{t}} = \sigma^2\mathbf{I}_m + \mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T$. By differentiating the log marginal likelihood with respect to $\boldsymbol{\alpha}$ and σ^2 and setting them to zero, the hyperparameters are given by (see Appendix B for the derivation)

$$\begin{aligned} \alpha_i &= \frac{1}{\mu_i^2 + \Sigma_{ii}}, \quad i = 1, 2, \ldots, n \\ \sigma^2 &= \frac{\|\mathbf{t} - \mathbf{\Phi}\mathbf{w}\|^2}{m - \sum_{i=1}^{n} \gamma_i} \end{aligned} \tag{16}$$

where $\gamma_i = 1 - \alpha_i\Sigma_{ii}$, μ_i is the ith entry of $\boldsymbol{\mu}$ and Σ_{ii} is the ith diagonal element of $\mathbf{\Sigma}$. Since the calculation of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ requires the values of α_i and σ^2, the hyperparameters and the posterior statistics are solved in an iterative way as follows.

Step 1) Initialize $\boldsymbol{\alpha}$ and σ^2.
Step 2) Calculate $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ using (11).
Step 3) Update $\boldsymbol{\alpha}$ and σ^2 using (16).
Step 4) Removing signatures corresponding to $\alpha_i = \infty$ or $\mu_i < 0$ from $\mathbf{\Phi}$.
Step 5) Repeat Steps 2–4 until convergence.

To help understand the behavior of this iterative procedure, we consider the case where columns of $\mathbf{\Phi}$ are orthonormal so that $\mathbf{\Phi}^T\mathbf{\Phi} = \mathbf{I}_n$. In this case, $\mu_i = (\mathbf{\Phi}^T\mathbf{t})_i/(1 + \sigma^2\alpha_i)$. It is apparent that as α_i approaches infinity, μ_i tends to be zero. In addition, (16) infers that larger entries in $\boldsymbol{\mu}$ are always updated by multiplying with larger coefficients during iteration. Therefore, the above procedure can reinforce some of the entries in $\boldsymbol{\mu}$ while suppressing others to generate sparsity. Note that in Step 4, besides removing signatures with $\alpha_i = \infty$, those agents with negative abundance fraction estimations are also determined to be absent. This is based on the physical meaning that the abundance fraction must be nonnegative. Later experiments will show that this is an effective way to reduce the error of the abundance fraction estimation and the false-alarm rate.

As shown in Appendix C, by expectation maximization (EM) [21], the same update formulations can be derived for the hyperparameters. If we do not force the negative abundance estimations to zero, the convergence of the above iteration can be guaranteed by the inherent property of the EM algorithm. In Appendix C, we prove that in the updating procedure of the sparse Bayesian method, forcing the negative abundance estimations to zero will not change the convergence of the iteration.

E. Agent Identification

In the sparse estimate of $\mathbf{w}$, irrelevant components are exactly zero. Therefore, the criterion for identifying agents is quite straightforward. Zero coefficients indicate the absence of the corresponding components. Nonzero coefficients are the estimated abundance fractions of agents present in the mixture. In summary, the agent identification scheme is as follows.

Step 1) Construct the regression matrix $\mathbf{\Phi}$ by signatures of all possible targets stored in the spectral signature library and the observation vector $\mathbf{t}$ by the spectrum of the sample to be identified.
Step 2) Solve the weight vector $\mathbf{w}$ by $\hat{\mathbf{w}} = \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is evaluated by the iterative procedure described in Section II-D.
Step 3) Identify the presence and absence of agents by the value of $\hat{\mathbf{w}}$. If $\hat{w}_i$ is equal to zero, the agent associated with the spectrum in the ith column of $\mathbf{\Phi}$ is absent. Otherwise, the agent is determined to be present with the estimated abundance fraction $\hat{w}_i$.

F. Signature Mismatch

Stochastic variations of signatures usually occur when the target spectra are measured in real systems and environments. We show an example here. The THz spectra of nine bacteria are measured three times. The principal component analysis (PCA) plot for these spectra is shown in Fig. 1. It is observed that the principal components (extracted features) of different bacteria are far away from each other but those for the three measurements of the same bacteria gather together closely. We compute the spectral correlation coefficients of the same bacteria and find that all of them are approaching one (larger than 0.999) while those values between different bacteria are much smaller. The large correlation coefficients imply that we can model the nominal signature (the one stored in the signature library) as follows:

$$\tilde{\boldsymbol{\phi}}_i = \kappa_i\boldsymbol{\phi}_i + \mathbf{\Delta}_i, \quad i = 1, 2, \ldots, n \tag{17}$$

where $\boldsymbol{\phi}_i$ represents the true signature (the one generating the mixture), κ_i denotes a multiplicative factor whose value is around one, and $\mathbf{\Delta}_i$ denotes the additive fluctuation at different frequency bins, which is modeled as a zero-mean Gaussian distributed random vector with the covariance matrix equal to $\delta_i\mathbf{I}_m$. Hence, the data model using the nominal signatures is reformulated as

$$\mathbf{t} = \tilde{\mathbf{\Phi}}\mathbf{w} + \mathbf{v} = \mathbf{\Phi}\tilde{\mathbf{w}} + \tilde{\mathbf{v}} \tag{18}$$

where

$$\begin{aligned} \tilde{\mathbf{w}} &= \mathbf{K}\mathbf{w} \\ \mathbf{K} &= \mathrm{diag}\left([\kappa_1 \quad \kappa_2 \quad \cdots \quad \kappa_n]\right) \\ \tilde{\mathbf{v}} &= \sum_{i=1}^{n} \mathbf{\Delta}_i w_i + \mathbf{v}. \end{aligned} \tag{19}$$

We can see that the zero elements in $\tilde{\mathbf{w}}$ and $\mathbf{w}$ are at the same positions, so that the unknown multiplicative factor has no influence on the sparsity of the abundance fraction vector. Moreover, the new noise term $\tilde{\mathbf{v}}$ follows the Gaussian distribution with zero mean and variance $\sum_{i=1}^{n} w_i^2\delta_i + \sigma^2$. Therefore, the fluctuation of signatures can be attributed to the measurement noise and

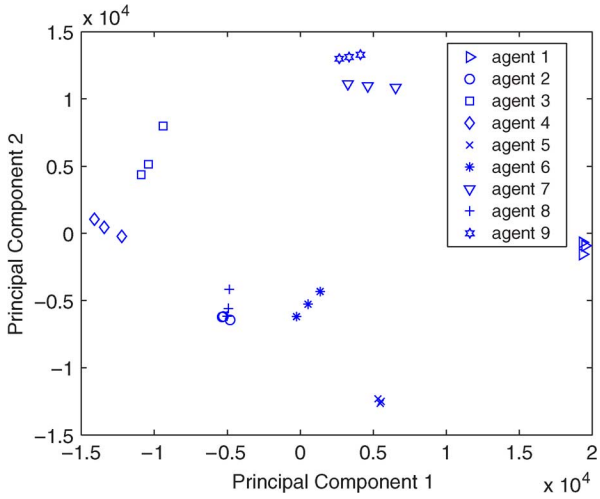


Fig. 1. PCA plot for the THz spectra of nine bacteria.

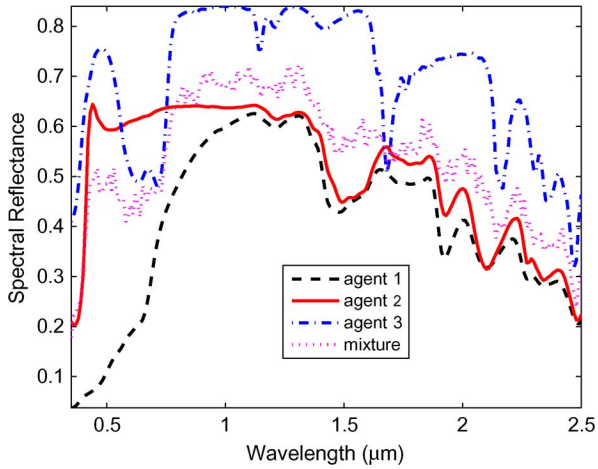


Fig. 2. Spectral reflectance of the three substances and the mixture with the SNR of 30 dB.

implies the decrease of the signal-to-noise ratio (SNR). Meanwhile, it can be observed that if the multiplicative perturbation exists, the influence of perturbations on the SNR is related with the magnitudes of the nonzero entries in $\mathbf{w}$. If the sum of the abundance fractions of all components is equal to one, the noise power is

$$\begin{aligned} \mathrm{P}_{\tilde{\mathbf{v}}} &= \sum_{i=1}^{n} w_i^2 \delta_i + \sigma^2 \\ &\le \|\mathbf{w}\|_2^2 \cdot \max_i \delta_i + \sigma^2 \\ &\le \|\mathbf{w}\|_1^2 \cdot \max_i \delta_i + \sigma^2 \\ &= \max_i \delta_i + \sigma^2. \end{aligned} \tag{20}$$

III. Numerical Studies

In our numerical studies, the abundance fractions of agents are estimated by the following methods: the sparse Bayesian model (SBM) using the noninformative hyperprior and taking into account the nonnegativity of the abundance fractions, as described in Section II-D, the NCLS algorithm and the orthogonal matching pursuit (OMP) method. Among these approaches, the NCLS algorithm has been used in [5] and [6] for agent identification based on the least squares optimization and the nonnegative constraint. The OMP method selects the relevant atoms by correlating them with the residual in a constructive way [13].

A. Experiments With Accurate Spectral Signatures

To illustrate the identification performance of the proposed method, in the following experiments, the UV to IR spectral reflectance of 30 materials are downloaded from the USGS spectral library (http://speclab.cr.usgs.gov/spectral.lib06/) and used to form the regression matrix $\mathbf{\Phi}$. A spectral mixture is manually formed by superimposing the reflectance of three materials with the equivalent abundance fraction 1/3. As typical SNRs of the actual spectrometers are not below 30 dB when the water absorption bands have been removed [3], our simulations assume that the observations are corrupted by the white Gaussian noise (WGN) with the SNR of 30 dB, where [3], [4]

$$\mathrm{SNR} = 10 \log_{10} \frac{\|\mathbf{\Phi w}\|_2^2}{m\sigma^2}. \tag{21}$$

The spectral reflectance of the three substances and the mixture are plotted in Fig. 2.

Based on the definitions given in [6], the probability of detection is computed by

$$\mathrm{P_D} = \frac{\sum_{i=1}^{30} \mathrm{N}\{|\hat{w}_i - w_i| < \gamma w_i | w_i \neq 0\}}{\sum_{i=1}^{30} \mathrm{N}\{w_i \neq 0\}} \tag{22}$$

while the probability of false alarm is computed by

$$\mathrm{P_F} = \frac{\sum_{i=1}^{30} \mathrm{N}\{\hat{w}_i > 0.5\gamma | w_i = 0\}}{\sum_{i=1}^{30} \mathrm{N}\{w_i = 0\}} \tag{23}$$

where $\mathrm{N}\{\cdot\}$ represents the number of trials in which the condition in the subsequent bracket is satisfied and γ is the threshold ranging within [0.01, 1] with the uniform log spacing. We measure the probabilities of detection and false alarm as the average with respect to the whole present and absent agents instead of a single one because the identification performance may vary among agents. Note that the definitions of $\mathrm{P_D}$ and $\mathrm{P_F}$ are different from those found in standard detection literatures (e.g., [19]). The choice of γ is closely related with the accuracy of the abundance estimation. To avoid confusion and observe the connection between the detection performance and the estimation accuracy, instead of plotting the receiver operating curve (ROC), the curves of $\mathrm{P_D}$ and $\mathrm{P_F}$ versus γ are shown in Figs. 3, 4. It is observed that both the NCLS and the OMP methods cannot beat SBM in terms of the identification performance.

Fig. 5 shows the histogram plotted by counting the abundance estimates of all the three present agents obtained in 10 000 trials. The y-axis denotes the frequency of the estimates occurring in a narrow interval so the profile of the histogram is an estimate of the posterior distribution of the abundance. It can be seen that the mode of the posterior distribution estimated by the NCLS algorithm has a significant bias from the true abundance. Fig. 6 is plotted in the same way for the 27 absent agents. It is observed that SBM achieves a higher probability of producing estimate $\hat{w}_i = 0$ for those zero-valued true abundances.

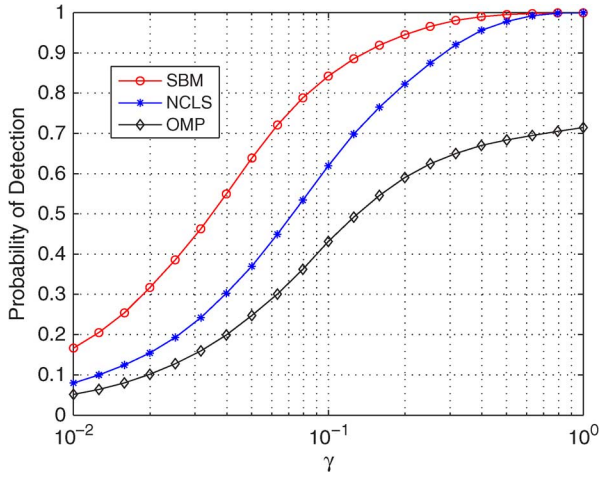


Fig. 3. Probability of detection obtained by the SBM, NCLS, and OMP methods.

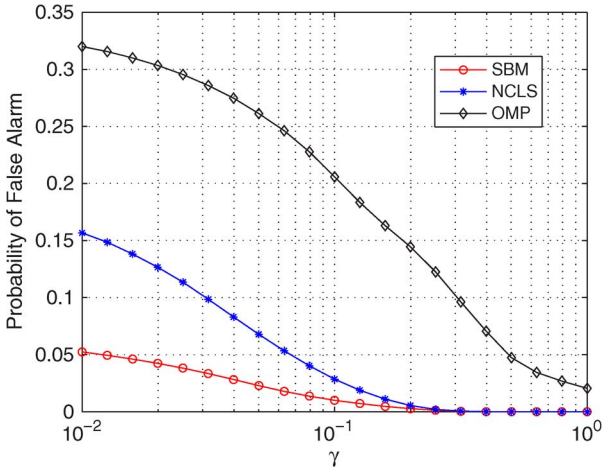


Fig. 4. Probability of false alarm obtained by the SBM, NCLS, and OMP methods.

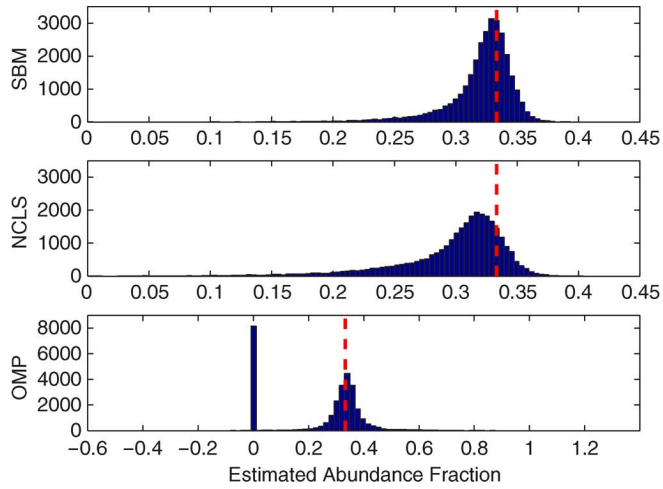


Fig. 5. Histogram of the abundance estimates for the agents present in the mixture, with the dashed line denoting the true abundance fraction.

The above results are obtained by using $m = 216$ observations (spectral bands). In some scenarios, the spectral resolution could be lower. By changing the interval of the wavelength, the identification performance under different numbers of observations is measured. The mean square error (MSE) of the estimated abundance fractions for agents that are present is calculated by running 10 000 trials:

$$\mathrm{MSE} = \left\{ \sum_{n=1}^{10\,000} \sum_{i=1, w_i \neq 0}^{30} [\hat{w}_i(n) - w_i(n)]^2 \right\} / 10\,000. \quad (24)$$

The curves in Fig. 7 show that SBM has the smallest MSE even under very low spectral resolution.

All the above results show that the performance of the OMP method is undesirable. We notice that the maximum coherence of the spectral signatures between different agents in the spectral library is up to 0.998. The condition number of the matrix -$\mathbf{\Phi}$ is over 2820. This highly ill-conditioned situation can not satisfy the sparse recovery condition studied in [13] for the OMP algorithm even with high SNR. This is the main reason why OMP fails to find a good approximation for the sparse solution in our application.

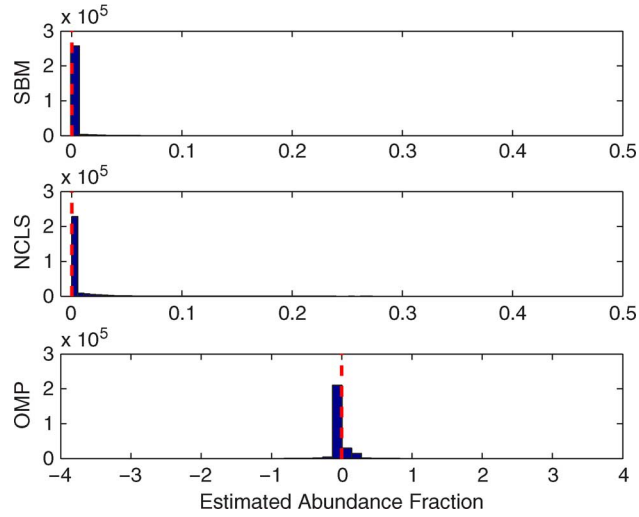


Fig. 6. Histogram of the abundance estimates for the absent agents, with the dashed line denoting the true abundance fraction.

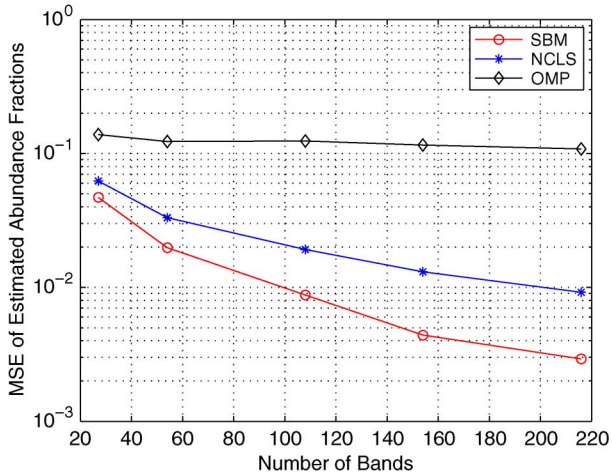


Fig. 7. Mean square error (MSE) of the estimated abundance fractions for the present agents.

B. *Experiments With Mismatched Spectral Signatures*

Numerical studies are carried out to demonstrate the agent identification performance when the spectral signatures are not

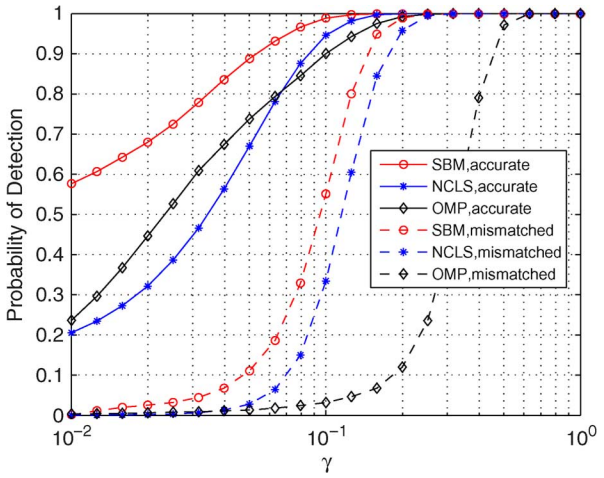


Fig. 8. Probability of detection obtained by the SBM, NCLS, and OMP methods with accurate and mismatched signatures.

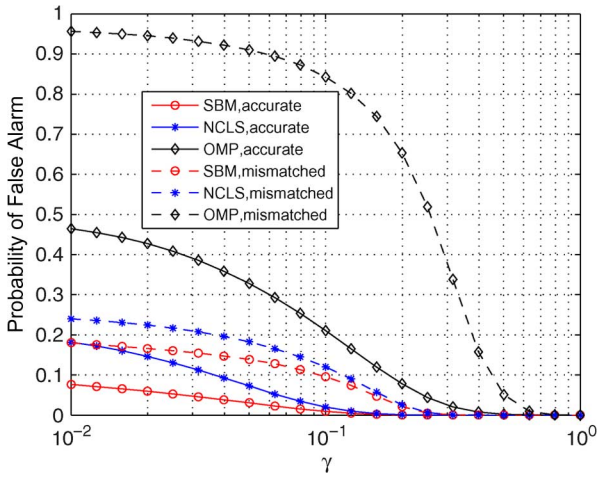


Fig. 9. Probability of false alarm obtained by the SBM, NCLS, and OMP methods with accurate and mismatched signatures.

exactly consistent with the spectra generating the mixture. The THz absorption spectra (within [0.18, 0.22] THz) of nine bacteria are used to form the spectral library in this experiment. As shown in Fig. 1, the spectra of one bacterium collected in two measurements are not exactly the same. In simulations, one of them is used to form the mixture while another one is saved in the signature library to simulate the mismatches occurring in real applications. The probabilities of detection and false alarm with accurate and mismatched signatures are compared in Figs. 8, 9. It is observed that all the methods exhibit some performance degradation with a decrease of the probability of detection and an increase of the probability of false alarm when mismatched signatures are used. We also notice that SBM outperforms the other two approaches even in the case of signature mismatch.

IV. Conclusion

A new agent identification method using a sparse Bayesian model has been introduced and investigated in this paper. By making use of the sparsity constraint of the abundance fraction vector, the proposed method yields significantly improved identification performance. The nonnegativity of the abundance fractions is taken into account and the convergence of the iterative algorithm is proved. The influence of the spectral signature mismatch has also been studied. The sparse agent identification scheme is applied to two sets of real data to demonstrate the performance in comparison with a few benchmark methods.

Appendix

For self completeness, the derivations for the weight posterior, the marginal likelihood and the hyperparameters are provided in detail in Appendices A and B although similar mathematics can be found in literatures (e.g., [9], [10], [20]). Moreover, the convergence of the iterative algorithm introduced in Section II-D is proved in Appendix C.

A. Derivation of the Weight Posterior (10) and the Marginal Likelihood (15)

By rearranging the exponential terms in (8) and (9), the product of the likelihood and the weight prior can be written as

$$\begin{aligned} &p(\mathbf{t}|\mathbf{w},\sigma^2)\cdot p(\mathbf{w}|\boldsymbol{\alpha}) \\ &\quad = |2\pi\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\right] \\ &\qquad \times |2\pi\sigma^2\mathbf{I}_m|^{-\frac{1}{2}}|2\pi\mathbf{A}^{-1}|^{-\frac{1}{2}}|2\pi\boldsymbol{\Sigma}|^{\frac{1}{2}} \\ &\qquad \times \exp\left[-\frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}_{\mathbf{t}}^{-1}\mathbf{t}\right] \end{aligned} \tag{25}$$

where

$$\begin{aligned} \boldsymbol{\mu} &= \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}^T\mathbf{t} \\ \boldsymbol{\Sigma} &= \left(\sigma^{-2}\boldsymbol{\Phi}^T\boldsymbol{\Phi}+\mathbf{A}\right)^{-1} \\ \boldsymbol{\Sigma}_{\mathbf{t}} &= \left[\sigma^{-2}\mathbf{I}_m-\sigma^{-4}\boldsymbol{\Phi}\left(\mathbf{A}+\sigma^{-2}\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^T\right]^{-1}. \end{aligned} \tag{26}$$

Using the Woodbury matrix identity [22], we get

$$\boldsymbol{\Sigma}_{\mathbf{t}} = \sigma^2\mathbf{I}_m+\boldsymbol{\Phi}\mathbf{A}^{-1}\boldsymbol{\Phi}^T. \tag{27}$$

It is easy to show that

$$\begin{aligned} &|2\pi\sigma^2\mathbf{I}_m|^{-\frac{1}{2}}|2\pi\mathbf{A}^{-1}|^{-\frac{1}{2}}|2\pi\boldsymbol{\Sigma}|^{\frac{1}{2}} \\ &\qquad = |2\pi\sigma^2\mathbf{I}_m|^{-\frac{1}{2}}|\mathbf{I}_n+\sigma^{-2}\mathbf{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Phi}|^{-\frac{1}{2}}. \end{aligned} \tag{28}$$

According to Sylvester's determinant theorem [22], we have

$$|\mathbf{I}_n+\sigma^{-2}\mathbf{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Phi}| = |\mathbf{I}_m+\sigma^{-2}\boldsymbol{\Phi}\mathbf{A}^{-1}\boldsymbol{\Phi}^T|. \tag{29}$$

Therefore, (28) is equal to $|2\pi\boldsymbol{\Sigma}_{\mathbf{t}}|^{-(1/2)}$. According to Bayes' rule, we know that

$$p(\mathbf{t}|\mathbf{w},\sigma^2)\cdot p(\mathbf{w}|\boldsymbol{\alpha}) = p(\mathbf{w}|\mathbf{t},\boldsymbol{\alpha},\sigma^2)\cdot p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2). \tag{30}$$

Comparing (25) and (30), it is clear that the expressions for the weight posterior distribution and the marginal likelihood are given by (10) and (15), respectively.

B. Derivation of the Hyperparameters (16)

In order to calculate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the hyperparameters $\boldsymbol{\alpha}$ and σ^2 need to be estimated first. They can be obtained by maximizing $p(\boldsymbol{\alpha}, \sigma^2|\mathbf{t})$, which is proportional to $p(\mathbf{t}|\boldsymbol{\alpha}, \sigma^2)p(\boldsymbol{\alpha})p(\sigma^2)$. When log-uniform hyperpriors are assumed, the optimization reduces to maximizing the marginal likelihood $p(\mathbf{t}|\boldsymbol{\alpha}, \sigma^2)$. Ignoring the terms irrelevant with the hyperparameters, the logarithm of the marginal likelihood (15) is

$$f = -\frac{1}{2}\left(\ln|\boldsymbol{\Sigma}_\mathbf{t}| + \mathbf{t}^T\boldsymbol{\Sigma}_\mathbf{t}^{-1}\mathbf{t}\right). \tag{31}$$

Let $\beta = \sigma^{-2}$. Based on (29), we can derive

$$\begin{aligned}|\boldsymbol{\Sigma}_\mathbf{t}| &= \beta^{-m}|\mathbf{I}_n + \beta\mathbf{A}^{-1}\boldsymbol{\Phi}^T\boldsymbol{\Phi}| \\ &= \beta^{-m}|\mathbf{A}|^{-1}|\boldsymbol{\Sigma}|^{-1}.\end{aligned} \tag{32}$$

Therefore, the first term in the right side of (31) is rewritten as

$$\ln|\boldsymbol{\Sigma}_\mathbf{t}| = -m\ln\beta - \ln|\boldsymbol{\Sigma}| - \ln|\mathbf{A}|. \tag{33}$$

By replacing $\boldsymbol{\Sigma}_\mathbf{t}$ by the form expressed in (26), the second term of (31) becomes

$$\mathbf{t}^T\boldsymbol{\Sigma}_\mathbf{t}^{-1}\mathbf{t} = \beta\|\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{\mu}\|^2 + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}. \tag{34}$$

Substituting (33) and (34) into (31), we get the log marginal likelihood as follows:

$$f = \frac{m}{2}\ln\beta + \frac{1}{2}\ln|\boldsymbol{\Sigma}| + \frac{1}{2}\sum_{i=1}^{n}\ln\alpha_i - \frac{\beta}{2}\|\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{\mu}\|^2 - \frac{1}{2}\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}. \tag{35}$$

The derivative of f with respect to α_i is

$$\begin{aligned}\frac{\partial f}{\partial\alpha_i} &= \frac{1}{2}\frac{\partial\ln|\boldsymbol{\Sigma}|}{\partial\alpha_i} + \frac{1}{2}\frac{\partial\sum_{j=1}^{n}\ln\alpha_j}{\partial\alpha_i} - \frac{1}{2}\frac{\partial\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}}{\partial\alpha_i} \\ &= -\frac{1}{2}\mathrm{Tr}(\mathbf{I}_i\boldsymbol{\Sigma}) + \frac{1}{2\alpha_i} - \frac{1}{2}\boldsymbol{\mu}^T\mathbf{I}_i\boldsymbol{\mu}\end{aligned} \tag{36}$$

where Tr denotes the trace of the matrix and $\mathbf{I}_i$ represents a diagonal matrix with all entries equal to 0 except the ith diagonal element. Setting (36) to zero yields

$$\alpha_i = \frac{1}{\mu_i^2 + \Sigma_{ii}} \Rightarrow \alpha_i = \frac{\gamma_i}{\mu_i^2} \tag{37}$$

where $\gamma_i = 1 - \alpha_i\Sigma_{ii}$. In practice, the latter form of the above update equations shows a higher convergence rate.

The derivative of f with respect to β is

$$\frac{\partial f}{\partial\beta} = \frac{1}{2}\frac{\partial\ln|\boldsymbol{\Sigma}|}{\partial\beta} + \frac{m}{2\beta} - \frac{1}{2}\|\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{\mu}\|^2. \tag{38}$$

By

$$(\mathbf{A} + \beta\boldsymbol{\Phi}^T\boldsymbol{\Phi})\boldsymbol{\Sigma} = \mathbf{I}_n \Rightarrow \boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\Sigma} = \beta^{-1}(\mathbf{I}_n - \mathbf{A}\boldsymbol{\Sigma}) \tag{39}$$

we know that

$$\frac{\partial\ln|\boldsymbol{\Sigma}|}{\partial\beta} = -\mathrm{Tr}(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\Sigma}) = -\beta^{-1}\sum_{i=1}^{n}(1 - \alpha_i\Sigma_{ii}). \tag{40}$$

Substituting (40) into (38) and setting (38) to zero, we obtain

$$\sigma^2 = \beta^{-1} = \frac{\|\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{\mu}\|^2}{m - \sum_{i=1}^{n}\gamma_i}. \tag{41}$$

C. Proof of the Convergence of the Iterative Algorithm in Section II-D

Treating the abundance fraction vector $\mathbf{w}$ as a hidden variable, $\boldsymbol{\alpha}$ and σ^2 are chosen to maximize the log marginal likelihood $L(\boldsymbol{\alpha}, \sigma^2) = \ln p(\mathbf{t}|\boldsymbol{\alpha}, \sigma^2) = \ln[\int p(\mathbf{t}|\mathbf{w}, \sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})\,\mathrm{d}\mathbf{w}]$ by the EM algorithm [9], [21].

First, we consider the update of $\boldsymbol{\alpha}$ assuming a fixed σ^2. Define

$$\Delta(\boldsymbol{\alpha}|\boldsymbol{\alpha}_l) = \int p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}_l, \sigma^2)\ln\frac{p(\mathbf{t}|\mathbf{w}, \sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})}{p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}_l, \sigma^2)p(\mathbf{t}|\boldsymbol{\alpha}_l, \sigma^2)}\,\mathrm{d}\mathbf{w}. \tag{42}$$

According to Jensen's inequality [21], we have

$$L(\boldsymbol{\alpha}_l, \sigma^2) + \Delta(\boldsymbol{\alpha}|\boldsymbol{\alpha}_l) \le L(\boldsymbol{\alpha}, \sigma^2). \tag{43}$$

This inequality implies that any $\boldsymbol{\alpha}$ which can increase the left-hand side will definitely increase the right-hand side. By the EM algorithm, $\boldsymbol{\alpha}$ is chosen to maximize $L(\boldsymbol{\alpha}_l, \sigma^2) + \Delta(\boldsymbol{\alpha}|\boldsymbol{\alpha}_l)$ instead of $L(\boldsymbol{\alpha}, \sigma^2)$. Therefore, ignoring independent terms in $\Delta(\boldsymbol{\alpha}|\boldsymbol{\alpha}_l)$, $\boldsymbol{\alpha}$ is updated by

$$\boldsymbol{\alpha}_{l+1} = \arg\max_{\boldsymbol{\alpha}} Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}_l) \tag{44}$$

where

$$\begin{aligned}Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}_l) &= \int p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}_l, \sigma^2)\ln[p(\mathbf{w}|\boldsymbol{\alpha})]\,\mathrm{d}\mathbf{w} \\ &= \mathrm{E}_{\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}_l, \sigma^2}\ln[p(\mathbf{w}|\boldsymbol{\alpha})].\end{aligned} \tag{45}$$

With (8) and (10), $Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}_l)$ is derived to be

$$\begin{aligned}Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}_l) &= \mathrm{E}_{\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}_l, \sigma^2}\left[-\frac{1}{2}\mathbf{w}^T\mathbf{A}\mathbf{w}\right] - \frac{1}{2}\ln|2\pi\mathbf{A}^{-1}| \\ &= \frac{1}{2}\left[\ln|\mathbf{A}| - \mathrm{Tr}(\mathbf{A}\boldsymbol{\Sigma}) - \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}\right] - \ln(2\pi)^{\frac{n}{2}}.\end{aligned} \tag{46}$$

By $\partial Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}_l)/\partial\alpha_i = 0$, we can derive

$$\alpha_i = \frac{1}{\mu_i^2 + \Sigma_{ii}}. \tag{47}$$

In the current estimate $\boldsymbol{\mu}_l$, we assume that the negative entries are denoted by $\mu_{k,l} < 0$, $k = 1, 2, \ldots, K$. Forcing negative abundance estimations to zero, all the affected variables are

expressed by new notations with the superscript $\prime$. The change of Q is

$$Q'(\boldsymbol{\alpha}_{l+1}|\boldsymbol{\alpha}_l) - Q(\boldsymbol{\alpha}_{l+1}|\boldsymbol{\alpha}_l) = \frac{1}{2}\sum_{k=1}^{K} \ln \frac{\alpha'_{k,l+1}}{\alpha_{k,l+1}} \tag{48}$$

where

$$\alpha_{k,l+1} = \frac{1}{\mu_{k,l}^2 + \Sigma_{kk,l}}, \quad k = 1, 2, \ldots, K \tag{49}$$

and

$$\alpha'_{k,l+1} = \frac{1}{\mu_{k,l}'^2 + \Sigma'_{kk,l}}, \quad k = 1, 2, \ldots, K. \tag{50}$$

Removing signatures corresponding to $\mu_{k,l} < 0$ is equivalent to $\mu_{k,l} < \mu'_{k,l} = 0$ and $\alpha_{k,l+1} < \alpha'_{k,l+1} = \infty$, so we can infer that

$$Q'(\boldsymbol{\alpha}_{l+1}|\boldsymbol{\alpha}_l) > Q(\boldsymbol{\alpha}_{l+1}|\boldsymbol{\alpha}_l). \tag{51}$$

Since $\boldsymbol{\alpha}_{(l+1)}$ is chosen to maximize $Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}_l)$, we have

$$Q(\boldsymbol{\alpha}_{l+1}|\boldsymbol{\alpha}_l) \geq Q(\boldsymbol{\alpha}_l|\boldsymbol{\alpha}_l). \tag{52}$$

The above inequalities lead to

$$\Delta(\boldsymbol{\alpha}_l|\boldsymbol{\alpha}_l) \leq \Delta(\boldsymbol{\alpha}_{l+1}|\boldsymbol{\alpha}_l) < \Delta'(\boldsymbol{\alpha}_{l+1}|\boldsymbol{\alpha}_l). \tag{53}$$

This inequality implies that $L(\boldsymbol{\alpha}, \sigma^2)$ is nondecreasing as $\boldsymbol{\alpha}$ is updated for each iteration even if we force the negative abundance estimations to zero.

Second, we consider the update of σ^2 for given $\boldsymbol{\alpha}$. With the same idea, σ^2 is chosen by

$$\sigma_{l+1}^2 = \arg\max_{\sigma^2} \quad Q\left(\sigma^2|\sigma_l^2\right) \tag{54}$$

where

$$\begin{aligned} Q\left(\sigma^2|\sigma_l^2\right) &= \int p\left(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}, \sigma_l^2\right) \ln\left[p(\mathbf{t}|\mathbf{w}, \sigma^2)\right] \mathrm{d}\mathbf{w} \\ &= \mathrm{E}_{\mathbf{w}|\mathbf{t},\boldsymbol{\alpha},\sigma_l^2} \ln\left[p(\mathbf{t}|\mathbf{w}, \sigma^2)\right] \\ &= \frac{\|\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{\mu}\|^2 + \mathrm{Tr}(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\Sigma})}{-2\sigma^2} - \ln\left(2\pi\sigma^2\right)^{\frac{m}{2}}. \end{aligned} \tag{55}$$

By differentiating $Q(\sigma^2|\sigma_l^2)$ with respect to σ^2, the update equation for σ^2 can be derived as

$$\sigma^2 = \frac{\|\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{\mu}\|^2}{m - \sum_{i=1}^{n} \gamma_i}. \tag{56}$$

This is the same as that derived in Appendix B. Forcing the negative abundance estimations to zero, $Q(\sigma^2|\sigma_l^2)$ at the $(l+1)$th step is changed by

$$Q'\left(\sigma_{l+1}^2|\sigma_l^2\right) - Q\left(\sigma_{l+1}^2|\sigma_l^2\right) = -\frac{m}{2} \ln \frac{\sigma_{l+1}'^2}{\sigma_{l+1}^2} \tag{57}$$

where

$$\sigma_{l+1}^2 = \frac{\|\mathbf{t} - \boldsymbol{\Phi}_l\boldsymbol{\mu}_l\|^2}{m - \mathrm{Tr}(\mathbf{I} - \mathbf{A}_l\boldsymbol{\Sigma}_l)} \tag{58}$$

and

$$\sigma_{l+1}'^2 = \frac{\left\|\mathbf{t} - \boldsymbol{\Phi}_l'\boldsymbol{\mu}_l'\right\|^2}{m - \mathrm{Tr}\left(\mathbf{I} - \mathbf{A}_l'\boldsymbol{\Sigma}_l'\right)}. \tag{59}$$

By (47), it is easy to show that $1 - \alpha_i\Sigma_{ii} \geq 0$ and hence $\mathrm{Tr}(\mathbf{I} - \mathbf{A}_l'\boldsymbol{\Sigma}_l') \leq \mathrm{Tr}(\mathbf{I} - \mathbf{A}_l\boldsymbol{\Sigma}_l)$. Moreover, since the true abundances are nonnegative, the residual obtained by $\left\|\mathbf{t} - \boldsymbol{\Phi}_l'\boldsymbol{\mu}_l'\right\|^2$ must be smaller than $\|\mathbf{t} - \boldsymbol{\Phi}_l\boldsymbol{\mu}_l\|^2$. These inequalities lead to

$$Q'\left(\sigma_{l+1}^2|\sigma_l^2\right) > Q\left(\sigma_{l+1}^2|\sigma_l^2\right) \geq Q\left(\sigma_l^2|\sigma_l^2\right). \tag{60}$$

Based on (53) and (60), we know that $L(\boldsymbol{\alpha}, \sigma^2)$ is nondecreasing as $\boldsymbol{\alpha}$ and σ^2 are updated for each iteration considering the nonnegativity of the abundance fractions. For the Gaussian marginal likelihood, $L(\boldsymbol{\alpha}, \sigma^2)$ is bounded from above. Therefore, we can arrive at the conclusion that the iterative algorithm in Section II-D converges.

REFERENCES

[1] J. F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 9, no. 10, pp. 2009–2025, Oct. 1998.

[2] C. I. Chang and D. C. Heinz, "Constrained subpixel target detetion for remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1144–1159, May 2000.

[3] N. Dobigeon, J. Y. Tourneret, and C. I. Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, Jul. 2008.

[4] W. Wang, T. Adalý, and D. Emge, "Subspace partitioning for target detection and identification," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1250–1259, Apr. 2009.

[5] C. Kwan, B. Ayhan, G. Chen, J. Wang, B. Ji, and C. Chang, "A novel approach for spectral unmixing, classification, and concentration estimation of chemical and biological agents," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 409–419, Feb. 2006.

[6] C. Kwan, G. Schmera, J. M. Smulko, L. B. Kish, P. Heszler, and C. Granqvist, "Advanced agent identification with fluctuation-enhanced sensing," *IEEE Sens. J.*, vol. 8, no. 6, pp. 706–713, Jun. 2008.

[7] L. L. Scharf and B. Friedlander, "Matched subspace detector," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[8] W. Wang, T. Adalý, and D. Emge, "Unsupervised target detection using canonical correlation analysis and its application to Raman spectroscopy," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Thessaloniki, Greece, Aug. 2007, pp. 247–252.

[9] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res. 1*, pp. 211–244, Jun. 2001.

[10] M. E. Tipping, "Bayesian inference: An introduction to principles and practice in machine learning," *Adv. Lectures Mach. Learn.*, pp. 41–62, Mar. 2004.

[11] M. A. T. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Trans. Pattern Anal. Mach. Learn.*, vol. 25, no. 9, pp. 1150–1159, Sep. 2003.

[12] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

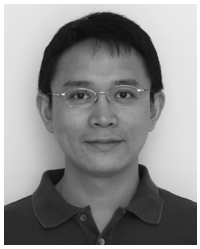
[13] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[14] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *J. Fourier Anal. Applicat.*, vol. 14, no. 5–6, pp. 877–905, Dec. 2008.

[15] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.
[16] E. J. Candès and Y. Plan, "Near-ideal model selection by l_1 minimization," *Ann. Statist.*, vol. 37, no. 5A, pp. 2145–2177, 2009.
[17] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.
[18] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling [a sensing/sampling paradigm that goes against the common knowledge in data acquisition]," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[19] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Englewood Cliffs, NJ: PTR Prentice-Hall, 1998.
[20] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Englewood Cliffs, NJ: PTR Prentice-Hall, 1998.
[21] S. Borman, "The expectation maximization algorithm, a short tutorial," [Online]. Available: http://www.seanborman.com/publications/EM-algorithm.pdf Jan. 2009
[22] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, MA: Cambridge Univ. Press, 1990.

Huiping Duan received the M.S. degree from Xidian University, Xi'an, China, in 2001 and the Ph.D. degree from Nanyang Technological University, Singapore, in 2008.

From June 2007 to January 2008, she was an Associate Scientist at Temasek Laboratories@NTU, Singapore. She has been a Research Scientist at L. C. Pegasus Corporation, Hillside, NJ, since 2009. Her research interests include array signal processing, statistical signal processing, pattern recognition, and spectral sensing.



Hongbin Li (M'99–SM'09) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, in 1999, all in electrical engineering.

From July 1996 to May 1999, he was a Research Assistant in the Department of Electrical and Computer Engineering, University of Florida. He was a Summer Visiting Faculty Member at the Air Force Research Laboratory in the summers of 2003, 2004 and 2009. Since July 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, where he is a Professor. His current research interests include statistical signal processing, wireless communications, and radars.

Dr. Li is a member of Tau Beta Pi and Phi Kappa Phi. He received the Harvey N. Davis Teaching Award in 2003 and the Jess H. Davis Memorial Award for excellence in research in 2001 from Stevens Institute of Technology, and the Sigma Xi Graduate Research Award from the University of Florida in 1999. He is currently a member of both the Sensor Array and Multichannel (SAM) Technical Committee and the Signal Processing Theory and Methods (SPTM) Technical Committee of the IEEE Signal Processing Society. He has been an editor or associate editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON SIGNAL PROCESSING, and served as a Guest Editor for *EURASIP Journal on Applied Signal Processing*, Special Issue on Distributed Signal Processing Techniques for Wireless Sensor Networks.

Jing Xie received the B.S. degree from Peking University, Beijing, China, in 2005 and the M.S. degree in physics from the Stevens Institute of Technology, Hoboken, NJ.

She is a research scientist at L.C. Pegasus Corp, Hillside, NJ. Her research interests focus on optical fiber sensors and THz detection of biological agents.

Nicolai S. Panikov is a Research Professor in the Biology Department, Northeastern University, Boston, MA, and a Principal Investigator with L. C. Pegasus Corporation, Hillside, NJ. He is former Head of the Department and vice-Director of the Winogradsky Institute of Microbiology, Russian Academy of Sciences, Moscow, Russia, and Faculty Member of the M. Lomonosov Moscow State University. In 1999, he moved to the U.S. and became a Professor in chemical biology at Stevens Institute of Technology, Hoboken, NJ, and a Senior Researcher at Dartmouth College, Hanover, NH. He is the author of three books, several patents, and more than 200 papers on microbial growth theory, environmental and polar microbiology, detection of hazardous microorganisms, development of instrumental techniques in biotechnology, subsurface and soil studies, and discovery and description of novel microorganisms in Arctic, wetlands, and soils as related to C and methane cycle, lignocellulose-degradation, and bioremediation.

Hong-Liang Cui (M'93) received the undergraduate education in applied physics with a concentration in laser optics from the Changchun Institute of Optics and Fine Mechanics, Changchun, China, receiving the B.Eng. degree, graduating at the top of his class.

He is a Professor of Applied Physics at the Polytechnic Institute of New York University. In 1981, he came to the United States for graduate study as one of the first group of Chinese physics students in the CUSPEA program, earning a Ph.D. degree in theoretical condensed matter physics in 1987, from Stevens Institute of Technology, Hoboken, NJ, where he had been on the physics faculty until 2009. His research efforts have been concentrated in the areas of solid-state electronics and nanoelectronics, optical communications and sensing, electromagnetic wave propagation, and interaction with matters such as chemical and bio-agents, and high-performance computing approach to modeling of physical devices and phenomena. For the last 20 years, his work has been funded by NSF, ARO, ONR, and DARPA, with a total funding of more than $15 M. He has published more than 200 research papers in peer-reviewed scientific journals, holds nine U.S. patents, and has guided more than 30 Ph.D. dissertations to completion.

Prof. Cui is a member of the American Physical Society, the Optical Society of America, and Sigma Xi.