Semiblind Hyperspectral Unmixing in the Presence of Spectral Library Mismatches

Xiao Fu, *Member, IEEE*, Wing-Kin Ma, *Senior Member, IEEE*,
José M. Bioucas-Dias, *Senior Member, IEEE*, and Tsung-Han Chan

***Abstract*—The dictionary-aided sparse regression (SR) approach has recently emerged as a promising alternative to hyperspectral unmixing in remote sensing. By using an available spectral library as a dictionary, the SR approach identifies the underlying materials in a given hyperspectral image by selecting a small subset of spectral samples in the dictionary to represent the whole image. A drawback with the current SR developments is that an actual spectral signature in the scene is often assumed to have zero mismatch with its corresponding dictionary sample, and such an assumption is considered too ideal in practice. In this paper, we tackle the spectral signature mismatch problem by proposing a dictionary-adjusted nonconvex sparsity-encouraging regression (DANSER) framework. The main idea is to incorporate dictionary-correcting variables in an SR formulation. A simple and low per-iteration complexity algorithm is tailor-designed for practical realization of DANSER. Using the same dictionary-correcting idea, we also propose a robust subspace solution for dictionary pruning. Extensive simulations and real-data experiments show that the proposed method is effective in mitigating the undesirable spectral signature mismatch effects.**

***Index Terms*—Compressive sensing (CS), dictionary mismatch, robust dictionary pruning, semiblind hyperspectral unmixing (HU), ℓ_p quasi-norm sparsity promoting.**

I. Introduction

HYPERSPECTRAL unmixing (HU) aims at decomposing pixels of a hyperspectral image (HSI) into constituent spectra that represent some pure materials. HU is useful in a number of applications, such as environment surveillance, agriculture, mine detection, and food and medicine analytics. As one of the core developments in signal and image processing for HSIs, various HU algorithms have been developed in the past two decades from different perspectives, such as Bayesian inference, nonnegative matrix factorization, convex analysis, pure-pixel pursuit, and many more (see, e.g., [2] and [3] for recent overviews).

Recently, a class of HU algorithms based on spectral libraries has attracted much attention. A spectral library is a collection of spectral signatures of materials acquired in controlled or ideal environments, e.g., in laboratories. There are several publicly available libraries, provided by government agencies and research institutes. For example, the U.S. Geological Survey (USGS) library [4] contains remotely sensed and extracted spectral signatures of over 1300 materials. Such rich knowledge of materials' spectra in the existing libraries provides new opportunities for HU. By using an existing library as a dictionary and by assuming the linear mixture model, we can treat HU as a problem of selecting a small number of spectra from the dictionary to represent all of the pixels. Such a dictionary-aided semiblind formulation is fundamentally identical to the well-known basis selection or sparse regression (SR) problem in compressive sensing (CS), and thus, many well-developed tools from CS can be applied. Fundamentally, there are several advantages with dictionary-aided semiblind HU. First, unlike many blind HU approaches (which do not use dictionaries), dictionary-aided methods do not require assumptions such as the pure-pixel assumption and the sum-to-one abundance conditions. Second, dictionary-aided methods may not require knowledge of the number of materials contained in the HSIs.

Several dictionary-aided HU algorithms based on SR were proposed in [5]–[8]. The algorithms in [5] and [6], [7] treat the HU problem as a single-pixel-based SR problem and a multiple-pixel-based collaborative SR (CSR) problem, respectively. Classic ℓ_1 norm and ℓ_2/ℓ_1 mixed-norm minimization-based sparse optimization methods are employed to tackle the formulated problems there. The corresponding optimization problems are convex and thus can be solved efficiently, e.g., by some specialized alternating direction method of multipliers (ADMM) based algorithms [9]. Readers are also referred to the literature such as [10]–[15] for concurrent developments of SR in dictionary-aided semiblind HU. Three main difficulties have been observed when applying the SR and CSR algorithms, however, [5]–[7]: first, the spectral library members (i.e., the recorded material spectra) exhibit very high mutual coherence. As is known in CS [16]–[19], high mutual coherence may lead to poor performance when applying ℓ_1 norm and ℓ_2/ℓ_1 mixed-norm minimization-based sparse optimization. Second, the size of a spectral library is often very large. Consequently, we are faced with a large-scale problem, for which computational

Manuscript received June 24, 2015; revised November 24, 2015 and March 6, 2016; accepted April 11, 2016. Date of publication May 10, 2016; date of current version August 2, 2016. This work was supported in part by the Hong Kong Research Grants Council General Research Fund under Project CUHK 14205414, in part by the Army Research Office Short-Term Innovative Research Program under Project ARO STIR W911NF-15-1-0384, and in part by the Portuguese Science and Technology Foundation under Project UID/EEA/50008/2013. Part of this work was published at the 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2014) in [1].

X. Fu is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: xfu@umn.edu).

W.-K. Ma is with the Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong (e-mail: wkma@ee.cuhk.edu.hk).

J. M. Bioucas-Dias is with the Instituto de Telecomunicações and Instituto Superior Técnico, Universidade de Lisboa, 1049-1 Lisbon, Portugal (e-mail: bioucas@lx.it.pt).

T.-H. Chan is with MediaTek Inc., Hsinchu 30078, Taiwan (e-mail: thchan@ieee.org).

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/TGRS.2016.2557340

efficiency becomes an issue. Third, there may be mismatches between the actual spectral signatures in the scene and the dictionary samples due to various reasons. Such dictionary mismatches can adversely affect the performance of a dictionary-aided semiblind HU algorithm.

The first two difficulties mentioned previously have been tackled by employing a dictionary-pruning method based on multiple signal classification (MUSIC) [8]. MUSIC is a classical subspace method in sensor array processing [20] and recently finds its application in CS [21]. In dictionary-aided semiblind HU, MUSIC proves to be useful in preselecting some relevant spectra from a large spectral library. As a result, a size-reduced dictionary can be constructed for the SR and CSR algorithms to perform semiblind HU. After dictionary pruning, both the mutual coherence of the dictionary and the complexity of the subsequent semiblind HU algorithm can be reduced.

However, the third difficulty, spectral signature mismatches, is still not addressed. In practice, the mismatch problem arises for several reasons. First, the materials' spectra may vary from time to time and from site to site, subject to diverse physical conditions, e.g., strength of sunlight and temperature [22], [23]. Second, the calibration procedure for spectral signatures may introduce errors. Third, the spatial resolutions of spectra in the dictionary can be different from those of the image, and that can also result in modeling errors. Spectral mismatches can result in rather damaging effects on the existing semiblind HU algorithms; in particular, MUSIC-based dictionary pruning is sensitive to spectral signature mismatches, as will be seen in the simulations.

Contributions: In this paper, we propose a dictionary-aided HU framework that takes spectral signature mismatches into consideration. Our first contribution lies in developing a new dictionary-aided HU algorithm. The formulation leading to the new algorithm uses insights of CSR but has two key differences: 1) we model spectral signature mismatches as bounded error vectors and attempt to compensate those errors in the formulation. 2) We employ the nonconvex ℓ_2/ℓ_p $(0 < p < 1)$ quasi-norm as the sparsity-promoting function, instead of the convex ℓ_2/ℓ_1 mixed-norm as in CSR [7]. The second endeavor is motivated by the fact that quasi-norm-based sparse optimization has been demonstrated to exhibit better sparsity-promoting performance in certain difficult situations, e.g., the high-coherence dictionary case [24]–[26]. Since our formulation considers dictionary adjustment, it is more complicated to handle than the previous CSR work. We derive the new algorithm by a careful design of alternating optimization, and its upshot is that the solution update at each iteration involves simple matrix operations.

The second contribution is a spectral mismatch-robust solution to dictionary pruning. We give a robust MUSIC formulation, wherein the goal is to identify spectral signature samples that are close to the true materials' signatures, rather than being exactly equal. At first look, the robust MUSIC method seems to be computationally expensive; specifically, for every dictionary sample, we need to solve an optimization problem. We show that, however, the optimization problem in robust MUSIC can be converted to a single-variable optimization problem and solved with a very low computational cost. Simulations and real-data experiment are used to show the effectiveness of the proposed algorithm.

The robust MUSIC approach was presented at IEEE WHISPERS 2014 [1]. In this journal version, we additionally include the proof of the optimality of robust MUSIC. Also, the dictionary-adjusted sparse HU algorithm is completely new.

Related Works: While the topic of CS and SR has received enormous attention in various fields, there are comparatively fewer works that study SR in the presence of dictionary mismatches. Those works usually appear in signal processing, and the application is not HU. In [27], perturbations of dictionaries were modeled as Gaussian noise, and an ℓ_1-norm regularized total least squares criterion was proposed; there, the focus was the single-measurement vector case (or the single-pixel case in our problem), and constraints on the unknowns were not considered. In [28] and [29], dictionary perturbations were modeled as scaling factors on each dictionary atom, and the formulated problem is convex. The algorithm in [30] attacked the dictionary mismatch problem in CSR-based direction-of-arrival finding. There, the mismatch was characterized by a subspace of a structured matrix, and the optimization surrogates there are also convex ℓ_2/ℓ_1 norm and its smoothed counterparts. We also note that ℓ_p quasi-norm-based SR was applied to HU for single-pixel-based unmixing without considering dictionary mismatches [31]. Here, our focus is CSR using multiple pixels, which is known to have both theoretical and practical advantages over the single-pixel-based algorithms; we adopt the nonconvex ℓ_2/ℓ_p quasi-norm, where $0 < p < 1$, as our sparsity-promoting function, since it has proven to show better performance in various applications; and we model spectral mismatches as deterministic bounded errors, which does not require statistical assumptions and may be more flexible.

Notation: The notations $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ mean that $\boldsymbol{x}$ and $\boldsymbol{X}$ are a real-valued n-dimensional vector and a real-valued $m \times n$ matrix, respectively. The notation $\boldsymbol{x} \geq \boldsymbol{0}$ (respectively, $\boldsymbol{X} \geq \boldsymbol{0}$) means that $\boldsymbol{x}$ (respectively, $\boldsymbol{X}$) is elementwise nonnegative. The ith column of a matrix $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ is denoted by $\boldsymbol{x}_i \in \mathbb{R}^m$, and the jth row of $\boldsymbol{X}$ is denoted by $\boldsymbol{x}^j$. The superscripts "T," "-1," and "$\dagger$" stand for the transpose, matrix inverse, and matrix pseudoinverse, respectively. The orthogonal projectors onto the range space of $\boldsymbol{X}$ and its orthogonal complement are denoted by $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^\dagger\boldsymbol{X}^T$ and $\boldsymbol{P}_{\boldsymbol{X}}^\perp = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}}$, respectively. The ℓ_p norm of a vector $\boldsymbol{x} \in \mathbb{R}^n$ $(p \geq 1)$ is denoted by $\|\boldsymbol{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. The ℓ_p quasi-norm $(0 < p < 1)$ is denoted by the same aforementioned notation. The mixed ℓ_p/ℓ_q norm or ℓ_p/ℓ_q quasi-norm is denoted by $\|\boldsymbol{X}\|_{q,p} =) \sum_{i=1}^m \|\boldsymbol{x}^i\|_q^p)^{1/p}$. The Frobenious norm is denoted by $\|\boldsymbol{X}\|_F = \|\boldsymbol{X}\|_{2,2}$.

II. Background

A. Signal Model and Dictionary-Aided Semiblind HU

Consider a remotely sensed scene that is composed of mixtures of N different materials. Assuming linear mixtures, the measured HSI can be modeled as

$$\boldsymbol{y}[\ell] = \sum_{n=1}^{N} \boldsymbol{a}_n s_n[\ell] + \boldsymbol{v}[\ell], \quad \ell = 1, \ldots, L \tag{1}$$

where $\boldsymbol{y}[\ell] \in \mathbb{R}^M$ denotes the hyperspectral measurement at the ℓth pixel of the image, with M being the number of spectral bands; each $\boldsymbol{a}_n \in \mathbb{R}^M$, $n = 1, \ldots, N$, represents the spectral signature of a particular material, indexed by n here; $s_n[\ell] \geq 0$ is the abundance of material n at pixel ℓ; $\boldsymbol{v}[\ell] \in \mathbb{R}^M$ is a noise vector; and L is the number of pixels. For convenience, we will write (1) in a matrix form

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{V} \tag{2}$$

where $\boldsymbol{Y} = [\boldsymbol{y}[1], \ldots, \boldsymbol{y}[L]]$, $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$, $\boldsymbol{S} = [\boldsymbol{s}[1], \ldots, \boldsymbol{s}[L]]$, $\boldsymbol{s}[\ell] = [s_1[\ell], \ldots, s_N[\ell]]^T$, and $\boldsymbol{V} = [\boldsymbol{v}[1], \ldots, \boldsymbol{v}[L]]$.

In HU, we aim by identifying $\boldsymbol{A}$ and $\boldsymbol{S}$ from $\boldsymbol{Y}$. This amounts to a blind separation problem where hyperspectral signal-specific properties—such as pure pixel and sum-to-one abundance conditions—are often utilized to attack the problem in many existing and concurrent HU studies. Dictionary-aided semiblind HU takes a different strategy. Motivated by the fact that many spectral libraries (e.g., the USGS library [4]) have been built in the past decades, its principle is to use one such spectral library as a dictionary to infer what are the underlying spectral signatures, and hence materials, in the scene. To put this into context, define

$$\boldsymbol{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_K] \in \mathbb{R}^{M \times K}$$

as a spectral dictionary, where each $\boldsymbol{d}_k \in \mathbb{R}^M$ is a previously recorded spectral sample for a specific material and K denotes the dictionary size or the number of spectral samples. A dictionary often contains a wide variety of samples of materials, and as such, K is large. The key assumption with dictionary-aided semiblind HU is that the dictionary covers the spectral signatures of all materials in the scene; that is to say

$$\boldsymbol{a}_n \in \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_K\}, \text{ for every } n = 1, \ldots, N.$$

Alternatively, we can write, for each $n = 1, \ldots, N$

$$\boldsymbol{a}_n = \boldsymbol{d}_{k_n}, \text{ for some } k_n \in \{1, \ldots, K\}. \tag{3}$$

Consequently, the signal model in (2) can be written as

$$\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{C} + \boldsymbol{V} \tag{4}$$

where $\boldsymbol{C} \in \mathbb{R}^{K,L}$ is a row-sparse matrix; to be specific, the k_nth row of $\boldsymbol{C}$, $n = 1, \ldots, N$, is the kth row of $\boldsymbol{S}$, and the other rows of $\boldsymbol{C}$ are all zeros.

Let us consider the SR approach—currently the main approach for dictionary-aided semiblind HU. The idea is to exploit the sparsity of $\boldsymbol{C}$, thereby attempting to recover the indices $k_1, \ldots, k_n$ correctly and the abundance matrix $\boldsymbol{S}$ accurately. There is more than one way to formulate such a sparse promoting problem (see, e.g., [2], [3], and the references therein), and here, we are interested in the CSR formulation [7], [8]. The CSR formulation is given as follows:

$$\begin{aligned} \min_{\boldsymbol{C} \in \mathbb{R}^{K \times L}} \quad & \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{C}\|_F^2 + \lambda \|\boldsymbol{C}\|_{2,1} \\ \text{s.t.} \quad & \boldsymbol{C} \geq \boldsymbol{0} \end{aligned} \tag{5}$$

for some prespecified constant $\lambda > 0$. Here, notice that $\|\boldsymbol{C}\|_{2,1} = \sum_{i=1}^K \|\boldsymbol{c}^i\|_2$, which aims at promoting row-sparsity of $\boldsymbol{C}$. As can be seen in Problem (5), CSR seeks to find a nonnegative row-sparse $\boldsymbol{C}$ that provides a good approximation to $\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{C}$. Problem (5) is convex, and a fast algorithm based on ADMM has been derived for Problem (5) [7].

B. Dictionary Pruning Using the Subspace Approach

As discussed in the introduction, large dictionary size and high mutual coherence with the dictionary are two main difficulties encountered in CSR and other SR methods, and these two difficulties may be circumvented by applying dictionary pruning. Here, we are interested in a subspace-based dictionary-pruning method called MUSIC [8]. This subspace method may be best described by studying the noiseless case $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{S}$. Let $\boldsymbol{U}_S \in \mathbb{R}^{M \times N}$ denote a matrix that contains the first N left singular vectors of $\boldsymbol{Y}$. It can be shown that, in the noiseless case and under some mild assumptions,[1] we have

$$\boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{d}_k = \boldsymbol{0} \Longleftrightarrow \boldsymbol{d}_k = \boldsymbol{a}_{k_n} \text{ for some } n \in \{1, \ldots, N\}. \tag{6}$$

The physical meaning of (6) is that, if a spectral sample $\boldsymbol{d}_k$ in the dictionary is also one of the spectral signatures in the scene, then it must be perpendicular to the orthogonal complement signal subspace. Also, the converse is true. From an algorithm viewpoint, the aforementioned observation suggests that we can correctly identify the indices $k_1, \ldots, k_N$ by the simple closed-form equations at the left-hand side (LHS) of (6)—at least in the noiseless case.

In practice, where noise is present, the LHS of (6) may not be exactly all zero. Under such circumstances, the following procedure can be used to estimate $k_1, \ldots, k_N$.

1) For $k = 1, \ldots, K$, calculate

$$\gamma_{\text{MUSIC}}(k) = \frac{\boldsymbol{d}_k^T \boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|_2^2}. \tag{7}$$

2) Determine $\hat{\Lambda} = \{\hat{k}_1, \ldots, \hat{k}_N\}$ such that, for $n = 1, \ldots, N$, we have $\gamma_{\text{MUSIC}}(\hat{k}_n) < \gamma_{\text{MUSIC}}(j)$ for all $j \notin \hat{\Lambda}$.

The aforementioned procedure is known as MUSIC [8], [21]. Also, note that we may use some other hyperspectral subspace identification algorithms, e.g., HySiMe [32], to estimate the signal subspace matrix $\boldsymbol{U}_S$ from the noisy $\boldsymbol{Y}$. MUSIC can, in principle, be used to perform dictionary-based semiblind HU. However, because of its sensitivity to colored noise and modeling error that are usually present in real data, it is used as a preprocessing algorithm for CSR (or other SR methods) in practice. Specifically, MUSIC is used to discard a large number of spectral samples that yield large residuals $\gamma_{\text{MUSIC}}(k)$. The remaining spectral samples then form a (much) smaller dictionary for CSR to operate. Such a dictionary-pruning procedure has been found to be able to improve the HU performance and speed up the process quite significantly (see [8] for details).

[1] Specifically, we require that $\boldsymbol{S}$ has full row rank and that spark$(\boldsymbol{D}) > N+1$, where spark$(\boldsymbol{X}) = r + 1$ means that r is the smallest number such that any r columns of $\boldsymbol{X}$ are linearly independent. Intuitively, these requirements mean that the abundance maps of the different materials are sufficiently different and that any N spectral samples in the dictionary are sufficiently different.

III. Proposed Approach

The crucial assumption with dictionary-aided semiblind HU is that there are no spectrum mismatches, i.e., we can always find a dictionary sample that *exactly* matches an actual spectral signature in the scene [cf., (3)]. As discussed in the introduction, this may be not the case in reality. In this section, we will propose a dictionary-aided semiblind HU that takes into account the presence of spectrum mismatches.

A. Dictionary-Adjusted Nonconvex Sparsity-Encouraging Regression (DANSER)

We assume the following spectrum mismatch model in place of (3):

$$\boldsymbol{d}_{k_n} = \boldsymbol{a}_n + \boldsymbol{\varepsilon}_n, \quad n = 1, \ldots, N \tag{8}$$

for some $\boldsymbol{\varepsilon}_n \in \mathbb{R}^M$ that characterizes the mismatch between the presumed and actual spectra of each material. In particular, every spectral error $\boldsymbol{\varepsilon}_n$ is assumed to be bounded, i.e.,

$$\|\boldsymbol{\varepsilon}_n\|_2 \leq \delta, \quad n = 1, \ldots, N$$

for some $\delta > 0$. Physically, our model assumes that the dictionary still covers all of the actual spectral signatures in the scene, but their "best matched" spectral samples in the dictionary are subject to certain perturbations. Also, such perturbations do not go worse than δ^2 in terms of magnitude.

Our rationale is to adjust the dictionary in the CSR formulation. Specifically, we write

$$\boldsymbol{d}'_k = \boldsymbol{d}_k + \boldsymbol{e}_k, \quad k = 1, \ldots, K$$

where each $\boldsymbol{e}_k \in \mathbb{R}^M$ is a dictionary correction variable, and we assume $\|\boldsymbol{e}_k\|_2 \leq \delta$. Following the CSR formulation in (5), we propose a new formulation as follows:

$$\begin{aligned} \min_{\boldsymbol{D}' \in \mathbb{R}^{M \times K}, \boldsymbol{C} \in \mathbb{R}^{K \times L}} \quad & \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{D}'\boldsymbol{C}\|_F^2 + \lambda \|\boldsymbol{C}\|_{2,p}^p \\ \text{s.t.} \quad & \|\boldsymbol{d}'_k - \boldsymbol{d}_k\|_2 \leq \epsilon, \ k = 1, \ldots, K, \\ & \boldsymbol{C} \geq \boldsymbol{0} \end{aligned} \tag{9}$$

where $0 < p < 1$, $\lambda > 0$, and $\epsilon > 0$ are prespecified, and note that $\|\boldsymbol{C}\|_{2,p}^p = \sum_{i=1}^K \|\boldsymbol{c}^i\|_2^p$. Comparing the original CSR formulation in (5) and the aforementioned formulation, we see two differences. First, Problem (9) adjusts the dictionary to attempt to neutralize the spectrum mismatches. Second, Problem (9) employs a nonconvex row-sparsity-promoting function $\|\boldsymbol{C}\|_{2,p}^p$. The reason is that nonconvex ℓ_p quasi-norms may exhibit better sparsity-promoting performance than the ℓ_1-norm, as reported in the sparse optimization context [24], [26], [33], and we endeavor to explore such an opportunity to improve SR performance in the HU application. The formulation in (9) or its variants will be called DANSER in the sequel.

B. Efficient Algorithm for DANSER

Having expressed the DANSER formulation in the last section, we turn our attention to algorithm design for DANSER. A simple approach to handle DANSER is to apply alternating optimization: fix $\boldsymbol{D}'$ and optimize Problem (9) with respect to (w.r.t.) $\boldsymbol{C}$ at one time, fix $\boldsymbol{C}$ and optimize Problem (9) w.r.t. $\boldsymbol{D}'$ at another time, and repeat the aforementioned cycle until some stopping criterion holds. While this approach is doable, our algorithm design experience is that it can lead to a computationally expensive algorithm. For instance, the optimization of Problem (9) w.r.t. $\boldsymbol{D}'$ involves joint adjustment of all of the dictionary samples in an inseparable manner, which is computationally involved for large dictionary sizes. Also, the nonconvex row-sparsity-promoting function $\|\boldsymbol{C}\|_{2,p}^p$ used in Problem (9) introduces difficulties in the optimization of Problem (9) w.r.t. $\boldsymbol{C}$.

In view of the aforementioned issues, we formulate a modified version of Problem (9)

$$\begin{aligned} \min_{\boldsymbol{D}', \boldsymbol{H}, \boldsymbol{C}} \quad & \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{C}\|_F^2 + \frac{\mu}{2}\|\boldsymbol{H} - \boldsymbol{D}'\|_F^2 + \lambda \sum_{k=1}^K \left(\|\boldsymbol{c}^k\|_2^2 + \tau\right)^{\frac{p}{2}} \\ \text{s.t.} \quad & \|\boldsymbol{d}'_k - \boldsymbol{d}_k\|_2 \leq \epsilon, \quad k = 1, \ldots, K, \\ & \boldsymbol{C} \geq \boldsymbol{0} \end{aligned} \tag{10}$$

where $\mu, \tau > 0$ and $\boldsymbol{H}$ is a slack variable. In particular, it can be verified that, if $\mu = +\infty$ and $\tau = 0$, then Problem (10) and Problem (9) are essentially the same. It should be noted that we have applied the variable splitting technique in Problem (10) (specifically, to the variable $\boldsymbol{C}$), which is a commonly used trick in contexts such as image reconstruction [34]–[36].

The modified DANSER formulation in (10) can be handled in a low per-iteration complexity fashion. To describe it, let us first consider the following lemma [37]–[39].

Lemma 1: Let $\phi_p(w) = ((2-p)/2)((2/p)w)^{p/(p-2)} + \tau w$, where $0 < p < 2$ and $\tau > 0$. The function $\phi_p(w)$ is strictly convex on $w \geq 0$. Also, $\phi_p(w)$ satisfies the following identity:

$$(x^2 + \tau)^{\frac{p}{2}} = \min_{w \geq 0} w \cdot x^2 + \phi_p(w)$$

and the solution to the aforementioned problem is uniquely given by

$$w_{\text{opt}} = \frac{p}{2}(x^2 + \tau)^{\frac{p-2}{2}}. \tag{11}$$

By Lemma 1, Problem (10) can be equivalently expressed as

$$\begin{aligned} \min_{\boldsymbol{H}, \boldsymbol{C}, \boldsymbol{D}', \{w_k\}} \quad & \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{C}\|_F^2 + \frac{\mu}{2}\|\boldsymbol{H} - \boldsymbol{D}'\|_F^2 \\ & + \lambda \sum_{k=1}^K \left(w_k \left\|\boldsymbol{c}^k\right\|_2^2 + \phi_p(w_k)\right) \\ \text{s.t.} \quad & \|\boldsymbol{d}'_k - \boldsymbol{d}_k\|_2 \leq \epsilon, k = 1, \ldots, K, \\ & \boldsymbol{C} \geq \boldsymbol{0}, \quad w_k \geq 0, k = 1, \ldots, K. \end{aligned} \tag{12}$$

Now, our strategy is to perform alternating optimization w.r.t. $\boldsymbol{H}$, $\boldsymbol{D}'$, $\{w_k\}$, $\boldsymbol{c}^1, \ldots, \boldsymbol{c}^K$. As we will see soon, the merit of doing so is that every update admits a computationally light solution.

First, we examine the optimization w.r.t. $\boldsymbol{H}$. One can easily see that the solution is

$$\boldsymbol{H} := \left(\mu \boldsymbol{D}' + \boldsymbol{Y}\boldsymbol{C}^T\right)\left(\boldsymbol{C}\boldsymbol{C}^T + \mu \boldsymbol{I}\right)^{-1}. \tag{13}$$

Second, the optimization w.r.t. $\boldsymbol{D}'$ is separable w.r.t. $\boldsymbol{d}'_1, \ldots, \boldsymbol{d}'_K$; i.e., for $k = 1, \ldots, K$, we have

$$\begin{aligned} \min_{\boldsymbol{d}'_k} \quad & \|\boldsymbol{d}'_k - \boldsymbol{h}_k\|_2^2 \\ \text{s.t.} \quad & \|\boldsymbol{d}'_k - \boldsymbol{d}_k\|_2 \leq \epsilon. \end{aligned} \tag{14}$$

Problem (14) is a projection problem, and the solution is

$$\boldsymbol{d}'_k := \begin{cases} \boldsymbol{h}_k, & \|\boldsymbol{h}_k - \boldsymbol{d}_k\|_2 \leq \epsilon \\ \boldsymbol{d}_k + \epsilon \frac{\boldsymbol{h}_k - \boldsymbol{d}_k}{\|\boldsymbol{h}_k - \boldsymbol{d}_k\|_2}, & \text{otherwise.} \end{cases} \tag{15}$$

Third, to obtain the solution w.r.t. $\boldsymbol{c}^k$, let us first rewrite the optimization w.r.t. $\boldsymbol{C}$ as

$$\begin{aligned} \min_{\boldsymbol{C}} \quad & \left\|\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{H}}\boldsymbol{C}\right\|_F^2 \\ \text{s.t.} \quad & \boldsymbol{C} \geq \boldsymbol{0} \end{aligned}$$

where

$$\tilde{\boldsymbol{Y}} = \begin{bmatrix} \sqrt{\frac{1}{2}}\boldsymbol{Y} \\ \boldsymbol{0} \end{bmatrix} \quad \tilde{\boldsymbol{H}} = \begin{bmatrix} \sqrt{\frac{1}{2}}\boldsymbol{H} \\ \text{Diag}(\boldsymbol{\theta}) \end{bmatrix}$$

and $\boldsymbol{\theta} := [\sqrt{w_1\lambda}, \ldots, \sqrt{w_K\lambda}]^T$. Then, the subproblem w.r.t. $\boldsymbol{c}^k$ can be expressed as

$$\begin{aligned} \min_{\boldsymbol{c}^k} \quad & \left\|\tilde{\boldsymbol{Y}}_k - \tilde{\boldsymbol{h}}_k\boldsymbol{c}^k\right\|_F^2 \\ \text{s.t.} \quad & \boldsymbol{c}^k \geq \boldsymbol{0} \end{aligned} \tag{16}$$

where

$$\tilde{\boldsymbol{Y}}_k = \begin{bmatrix} \sqrt{\frac{1}{2}}\boldsymbol{Y} - \sum_{j\neq k}\sqrt{\frac{1}{2}}\boldsymbol{h}_j\boldsymbol{c}^j \\ \boldsymbol{0} \end{bmatrix} \quad \tilde{\boldsymbol{h}}_k = \begin{bmatrix} \sqrt{\frac{1}{2}}\boldsymbol{h}_k \\ \sqrt{w_k\lambda}\boldsymbol{f}_k \end{bmatrix}$$

in which $\boldsymbol{f}_k$ is the kth column of the $K \times K$ identity matrix. Problem (16) is known to have a simple solution [40], [41], given by

$$(\boldsymbol{c}^k)^T := \left(\frac{\left[\tilde{\boldsymbol{Y}}^T\tilde{\boldsymbol{H}}\right]_{:,k} - \boldsymbol{C}^T\left[\tilde{\boldsymbol{H}}^T\tilde{\boldsymbol{H}}\right]_{:,k} + (\boldsymbol{c}^k)^T\left[\tilde{\boldsymbol{H}}^T\tilde{\boldsymbol{H}}\right]_{k,k}}{\left[\tilde{\boldsymbol{H}}^T\tilde{\boldsymbol{H}}\right]_{k,k}} \right)_+ \tag{17}$$

where $(x)_+ = \max\{0, x\}$. Notice that using the update (17) is desirable: the large matrix products $\tilde{\boldsymbol{Y}}^T\tilde{\boldsymbol{H}}$ and $\tilde{\boldsymbol{H}}^T\tilde{\boldsymbol{H}}$ both only need to be calculated once before updating $\boldsymbol{c}^1, \ldots, \boldsymbol{c}^K$. Finally, by Lemma 1, the solution w.r.t. $\{w_k\}$ is

$$w_k = \left(\frac{p}{2}\right)\left(\|\boldsymbol{c}^k\|_2^2 + \tau\right)^{\frac{(p-2)}{2}}, \quad k = 1, \ldots, K. \tag{18}$$

The alternating optimization process described previously is summarized in Algorithm 1, and we simply call it DANSER. The DANSER algorithm has the following solution convergence guarantee.

Proposition 1: Every limit point of the solution sequence produced by DANSER (Algorithm 1) is a stationary point of Problem (10).

The proof of the aforementioned proposition is relegated to Appendix A. Proposition 1 indicates that, although we have been dealing with Problem (10) indirectly, a stationary point of Problem (10) may be expected. Following Proposition 1, we can stop DANSER by checking the relative or absolute change of the solution $\boldsymbol{C}$. Notice that, since Problem (10) is nonconvex, a good initialization would help DANSER converge to a better solution. In practice, one can use the CSR solution mentioned in Section II-A to initialize DANSER.

Algorithm 1: DANSER

input : $(\lambda, \tau, p, \mu, \epsilon)$; $\boldsymbol{D}$; $\boldsymbol{C}$ (initialization); $\boldsymbol{Y}$.

$w_k = (p/2)(\|\boldsymbol{c}^k\|_2^2 + \tau)^{(p-2)/2}$ for $k = 1, \ldots, K$.

repeat

  Unmixing: construct $\boldsymbol{\theta} := [\sqrt{w_1\lambda}, \ldots, \sqrt{w_k\lambda}]^T$;

$$\tilde{\boldsymbol{Y}} = \begin{bmatrix} \sqrt{\frac{1}{2}}\boldsymbol{Y} \\ \boldsymbol{0} \end{bmatrix}, \quad \tilde{\boldsymbol{H}} = \begin{bmatrix} \sqrt{\frac{1}{2}}\boldsymbol{H} \\ \text{Diag}(\boldsymbol{\theta}) \end{bmatrix};$$

  let $\boldsymbol{F} := \tilde{\boldsymbol{Y}}^T\tilde{\boldsymbol{H}}$ and $\boldsymbol{G} := \tilde{\boldsymbol{H}}^T\tilde{\boldsymbol{H}}$;

  for $k = 1 : K$ **do**

    update $\boldsymbol{c}^k$ by

$$\boldsymbol{c}^k := \left(\frac{\boldsymbol{F}_{:,k} - \boldsymbol{C}^T\boldsymbol{G}_{:,k} + (\boldsymbol{c}^k)^T\boldsymbol{G}_{k,k}}{\boldsymbol{G}_{k,k}}\right)_+;$$

  end

  Dictionary Adjusting: update $\boldsymbol{H}$ by

$$\boldsymbol{H} := (\mu\boldsymbol{D}' + \boldsymbol{Y}\boldsymbol{C}^T)\left(\boldsymbol{C}\boldsymbol{C}^T + \mu\boldsymbol{I}\right)^{-1};$$

  Update Slack Variable:

  for $k = 1 : K$ **do**

$$\boldsymbol{d}'_k := \begin{cases} \boldsymbol{h}_k, & \|\boldsymbol{h}_k - \boldsymbol{d}_k\|_2 \leq \epsilon \\ \boldsymbol{d}_k + \epsilon\frac{\boldsymbol{h}_k - \boldsymbol{d}_k}{\|\boldsymbol{h}_k - \boldsymbol{d}_k\|_2}, & \text{o.w.} \end{cases};$$

  end

  Reweighting: update $\{w_k\}$ by

$$w_k = (p/2)(\|\boldsymbol{c}^k\|_2^2 + \tau)^{(p-2)/2}, \quad k = 1, \ldots, K;$$

until *some stopping criterion is satisfied*;

output: $\boldsymbol{C}$.

Remark 1: By analyzing the per-iteration complexity of DANSER, one can verify that the complexities of many operations scale with K (i.e., the size of the dictionary) or higher. For example, to solve (13), the operations $\boldsymbol{C}\boldsymbol{C}^T$ and $\boldsymbol{Y}\boldsymbol{C}^T$ cost $\mathcal{O}(K^2L)$ and $\mathcal{O}(MKL)$ flops, respectively, and the matrix inversion requires $\mathcal{O}(K^3)$ flops. Plus, although solving the problems w.r.t. $\boldsymbol{c}^k$ is easy, these procedures have to be repeated K times at each iteration. Practically, it is therefore motivated to use a dictionary with a smaller size, or, to prune the dictionary in advance. However, due to the existence of spectral signature mismatches, directly applying MUSIC as in [8] for this purpose is not appropriate any more. To address this problem, a robust dictionary-pruning method will be proposed in the next section.

C. Robust MUSIC for Dictionary Pruning

Consider the MUSIC procedure back in Section II-B. In particular, recall that the metric

$$\gamma_{\text{MUSIC}}(k) = \frac{\boldsymbol{d}_k^T\boldsymbol{P}_{\hat{\boldsymbol{U}}_S}^{\perp}\boldsymbol{d}_k}{\|\boldsymbol{d}_k\|_2^2} \tag{19}$$

should yield a small value when $\boldsymbol{d}_k$ exactly matches an actual spectral signature in the scene, and this property has been used as the way to prune the dictionary in the MUSIC procedure. Now, in the presence of dictionary mismatches, we propose to replace (19) by the following robust MUSIC (RMUSIC) metric

$$\gamma_{\text{RMUSIC}}(k) = \min_{\boldsymbol{\xi}\in\mathbb{R}^M} \quad \frac{(\boldsymbol{d}_k-\boldsymbol{\xi})^T \boldsymbol{P}^{\perp}_{\boldsymbol{U}_S}(\boldsymbol{d}_k-\boldsymbol{\xi})}{\|\boldsymbol{d}_k-\boldsymbol{\xi}\|_2^2} \tag{20a}$$

$$\text{s.t.} \quad \|\boldsymbol{\xi}\|_2 \le \epsilon \tag{20b}$$

where $\epsilon > 0$ is prespecified. The idea is the same as the DANSER development in the aforementioned sections—adjust the dictionary to find a better match, this time in a subspace sense.

The key issue with realizing RMUSIC lies in solving Problem (20). Problem (20) is a single-ratio fractional quadratic program, which is quasi-convex and can be solved, e.g., by the Dinkelbach algorithm or its variants [42], [43]. While this means that we can implement RMUSIC by applying some existing optimization algorithms, we have to solve K such quasi-convex problems—which is still inefficient for large K. However, by carefully examining the problem structure, we find that this particular problem can be solved quite easily. To see this, let us re-express $\gamma_{\text{RMUSIC}}(k)$ as

$$\begin{aligned}\gamma_{\text{RMUSIC}}(k) &= \min_{\|\boldsymbol{\xi}\|_2\le\epsilon} \frac{\left\|\boldsymbol{P}^{\perp}_{\boldsymbol{U}_S}(\boldsymbol{d}_k-\boldsymbol{\xi})\right\|_2^2}{\|\boldsymbol{P}^{\perp}_{\boldsymbol{U}_S}(\boldsymbol{d}_k-\boldsymbol{\xi})\|_2^2 + \|\boldsymbol{P}_{\boldsymbol{U}_S}(\boldsymbol{d}_k-\boldsymbol{\xi})\|_2^2} \\ &= \min_{\|\boldsymbol{\xi}\|_2\le\epsilon} \frac{\eta_k^2(\boldsymbol{\xi})}{\eta_k^2(\boldsymbol{\xi})+1} \end{aligned} \tag{21}$$

where we recall that $\boldsymbol{P}_{\boldsymbol{U}_S} = \boldsymbol{U}_S\boldsymbol{U}_S^T$ is the orthogonal projector onto the range space of $\boldsymbol{U}_S$ and

$$\eta_k(\boldsymbol{\xi}) = \frac{\left\|\boldsymbol{P}^{\perp}_{\boldsymbol{U}_S}(\boldsymbol{d}_k-\boldsymbol{\xi})\right\|_2}{\|\boldsymbol{P}_{\boldsymbol{U}_S}(\boldsymbol{d}_k-\boldsymbol{\xi})\|_2}. \tag{22}$$

Since the objective function of (21) is a monotonically increasing function of $\eta^2(\boldsymbol{\xi}) \in [0,\infty)$, computing $\gamma_{\text{RMUSIC}}(k)$ is the same as finding the minimal value of $\eta_k(\boldsymbol{\xi})$ subject to $\|\boldsymbol{\xi}\|_2 \le \epsilon$. Let us denote

$$\eta_k^{\star} = \min_{\|\boldsymbol{\xi}\|_2\le\epsilon} \frac{\left\|\boldsymbol{P}^{\perp}_{\boldsymbol{U}_S}(\boldsymbol{d}_k-\boldsymbol{\xi})\right\|_2}{\|\boldsymbol{P}_{\boldsymbol{U}_S}(\boldsymbol{d}_k-\boldsymbol{\xi})\|_2}. \tag{23}$$

We show the following.

Proposition 2: The optimal value of Problem (23) can be found by solving a single-variable problem

$$\eta_k^{\star} = \min_{0\le\theta\le\epsilon} \frac{\left|\|\boldsymbol{P}^{\perp}_{\boldsymbol{U}_S}\boldsymbol{d}_k\|_2-\theta\right|}{\|\boldsymbol{P}_{\boldsymbol{U}_S}\boldsymbol{d}_k\|_2+\sqrt{\epsilon^2-\theta^2}}. \tag{24}$$

The proof of Proposition 2 is relegated to Appendix B. The message revealed here is quite intriguing—the originally quasi-convex problem can be recast as a simple single-variable problem that can be easily solved, e.g., by grid search or bisection. Practically, this means that the RMUSIC strategy can be implemented quite efficiently.

As in the previous MUSIC work [8], we use RMUSIC to perform dictionary pruning for DANSER. Specifically, we use RMUSIC to select a number of $\tilde{K}$ ($\tilde{K} < K$) spectral samples from $\boldsymbol{D}$, form a size-$\tilde{K}$ dictionary, denoted by $\tilde{\boldsymbol{D}}$ here, and then use $\tilde{\boldsymbol{D}}$ as a pruned dictionary to run DANSER. We summarize this procedure in Algorithm 2, and we call the procedure RMUSIC-DANSER.

Algorithm 2: RMUSIC-DANSER

input : $\boldsymbol{Y}$; $\boldsymbol{D}$; ϵ; $\tilde{K}$.
apply HySiMe [32] on $\boldsymbol{Y}$ to obtain $\boldsymbol{U}_S$;
for $k = 1 : K$ **do**
  $\eta_k^{\star} = \min_{0\le\theta\le\epsilon} \frac{\left|\|\boldsymbol{P}^{\perp}_{\boldsymbol{U}_S}\boldsymbol{d}_k\|_2-\theta\right|}{\|\boldsymbol{P}_{\boldsymbol{U}_S}\boldsymbol{d}_k\|_2+\sqrt{\epsilon^2-\theta^2}}$;
  $\gamma_{\text{RMUSIC}}(k) = \frac{(\eta_k^{\star})^2}{(\eta_k^{\star})^2+1}$;
end
determine $\hat{\Lambda} = \{\hat{k}_1,\ldots,\hat{k}_{\tilde{K}}\}$ such that $\gamma_{\text{RMUSIC}}(i) < \gamma_{\text{RMUSIC}}(j)$ for any $i \in \hat{\Lambda}$ and $j \notin \hat{\Lambda}$;
determine $\tilde{\boldsymbol{D}} = \left[\boldsymbol{d}_{\hat{k}_1}, \boldsymbol{d}_{\hat{k}_2}, \ldots, \boldsymbol{d}_{\hat{k}_{\tilde{K}}}\right]$;
feed $\tilde{\boldsymbol{D}}$ and $\boldsymbol{Y}$ to DANSER (Algorithm 1);
output: $\boldsymbol{C}$.

IV. Computer Simulations

In this section, we use synthetic HSIs to show the effectiveness of the proposed approach. In our simulations, the "ground-truth" spectra are randomly selected from a subset of the USGS library that has 332 spectral signatures; in this subset, the angles between any two endmembers are larger than 3°, and the 2-norms of the endmembers are larger than one. The "available dictionary" $\boldsymbol{D}$ is formed by the same subset of spectra, but a perturbation (i.e., $\boldsymbol{e}_k$ for $k = 1,\ldots,K$) is intentionally added to each spectrum. To quantify the "mismatch level" of the available dictionary, we define the *dictionary to modeling error ratio* (DMER) as follows:

$$\text{DMER(dB)} = 10\log_{10}\left(\|\boldsymbol{d}_{k^\star}\|_2^2/\delta^2\right)$$

where $k^\star = \arg\min_{k=1,\ldots,K}\|\boldsymbol{d}_k\|_2$ and $\delta = \max_{k=1,\ldots,K}\|\boldsymbol{e}_k\|_2$. The perturbation terms are first generated following a zero-mean i.i.d. Gaussian distribution and then scaled such that the DMER is satisfied. The abundances are generated following the uniform Dirichlet distribution. Throughout this section, we set the number of pixels to be $L = 5\,000$. We also define the signal-to-noise ratio (SNR) as $\text{SNR} = \sum_{\ell=1}^{L}\|\boldsymbol{A}\boldsymbol{s}[\ell]\|_2^2/ML\sigma^2$ to quantify the noise level, where σ^2 denotes the variance of the additive noise, which is also assumed to be zero-mean i.i.d. Gaussian. The choice of the parameter ϵ is as follows:

$$\epsilon = \frac{1-\alpha}{1+\alpha}\|\boldsymbol{d}_{k^\star}\|_2 \tag{25}$$

where $\alpha \in [0,1]$ is given. The parameter α controls the normalized correlation between the RMUSIC/DANSER-resulted dictionary member $\boldsymbol{d}_k - \boldsymbol{\xi}$ and the original dictionary member $\boldsymbol{d}_k$. Specifically, under $\|\boldsymbol{\xi}\|_2 \le \epsilon$, it can be shown that the choice of ϵ in (25) leads to $(\boldsymbol{d}_{k^\star}-\boldsymbol{\xi})^T\boldsymbol{d}_{k^\star}/\|\boldsymbol{d}_{k^\star}-\boldsymbol{\xi}\|_2\|\boldsymbol{d}_{k^\star}\|_2 \ge \alpha$; the derivation is shown in Appendix C. Note that ϵ need not be identical across different dictionary members under the

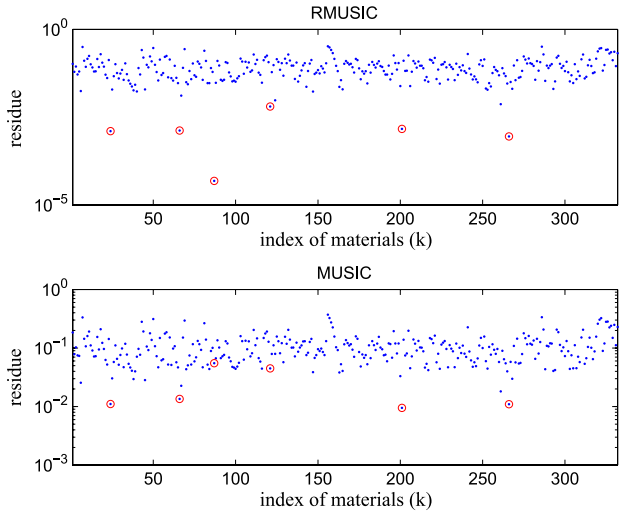


Fig. 1. Projection residues of MUSIC and RMUSIC.

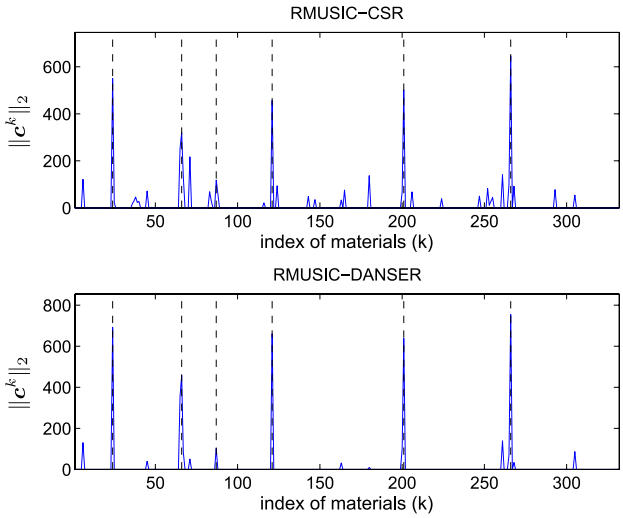


Fig. 2. 2-norms of $\boldsymbol{c}^k$'s of CSR and DANSER. The black dash lines correspond to the indices of the ground-truth materials' spectra in the dictionary; for the RMUSIC-pruned spectra, we set $\|\boldsymbol{c}^k\|_2 = 0$.

RMUSIC/DANSER framework: if some dictionary members have larger spectral signature mismatches, larger ϵ's may be given to them.

Figs. 1 and 2 show an illustrative example. Fig. 1 shows the residues of applying MUSIC and RMUSIC to prune the dictionary $\boldsymbol{D}$. Here, we randomly pick $N = 6$ spectra as the ground-truth materials and then use the described dictionary $\boldsymbol{D}$ to observe the performance of MUSIC and RMUSIC. The parameter of RMUSIC is set to be $\alpha = 0.85$, and we set DMER $=$ 20 dB and SNR $=$ 35 dB in this case. We see that MUSIC has difficulty in distinguishing several ground-truth spectra from the other dictionary members (to be precise, the third and fourth materials' spectra), but RMUSIC can clearly differentiate the same spectra from the irrelevant spectra. Fig. 2 compares the unmixing performance of DANSER and CSR using the same case, where the pruned dictionary with 40 spectra is obtained by RMUSIC. Here, the CSR part is performed by the CLSUnSAL algorithm [7], which is considered as a state of the art. For

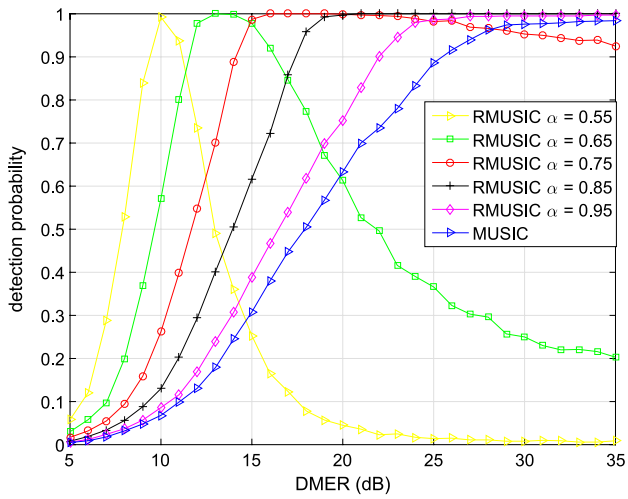


Fig. 3. Detection probabilities of RMUSIC/MUSIC under various DMERs and different α's. SNR $=$ 35 dB, $N = 8$, the pruned dictionary size is $\tilde{K} = 40$, and the original dictionary size $K = 332$.

DANSER, we set $p = 0.5$ and $\lambda = 0.04$ for this case. For CSR, the regularization parameter is $\lambda = 0.005$. In this example and the forthcoming simulations and real-data experiment, we feed the solution of CSR to DANSER as initialization. We see that RMUSIC-DANSER yields much row-sparser $\boldsymbol{C}$ than that of RMUSIC-CSR, and all of the desired spectra have been successively identified by DANSER.

In the following, we use Monte Carlo simulations to evaluate the performance of the proposed algorithms. Two performance discriminators will be used throughout this section. First, to measure the dictionary-pruning performance, we define the following detection probability:

$$\Pr\left\{\Lambda \subset \hat{\Lambda}\right\}$$

where $\Lambda = \{k_1, \ldots, k_N\}$ denotes the index set that indicates the ground-truth spectra and $\hat{\Lambda} \subseteq \{1, \ldots, K\}$ denotes an index selection subset outputted by a dictionary-pruning algorithm. Also, we will use $\tilde{K}$ to denote the size of the pruned dictionary. Second, to measure the unmixing performance, we calculate in the following the signal to reconstruction error (SRE) [5]–[7]:

$$\text{SRE(dB)} = 10\log_{10}\left(\frac{\|\boldsymbol{S}\|_F^2}{\left\|\boldsymbol{C} - \hat{\boldsymbol{C}}\right\|_F^2}\right)$$

where $\boldsymbol{C}$ is the true row-sparse abundance matrix (see (2) in Section II-A), and $\hat{\boldsymbol{C}}$ is the output of an unmixing algorithm.

In Fig. 3, we show the index set detection probabilities of MUSIC and RMUSIC under various DMERs. In each trial, $N = 8$ materials are randomly picked. The SNR in this simulation is set to be 35 dB, and $\tilde{K} = 40$ is employed. The results are averaged from 1000 trials. One can see that MUSIC is sensitive to dictionary mismatches even under high DMERs, and MUSIC is not able to identify all of the true materials from the dictionary. Generally, using RMUSIC with $\alpha = 0.85$ and 0.95 both yield much better detection probabilities than MUSIC under all DMERs. Interestingly, one can see that RMUSIC

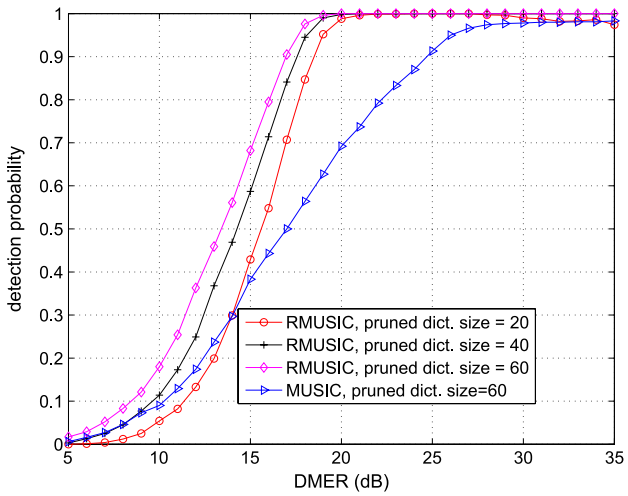


Fig. 4. Detection probabilities of RMUSIC/MUSIC under various DMERs and different $\tilde{K}$'s (size of the pruned dictionary). $\alpha = 0.85$, $N = 8$, the original dictionary size $K = 332$, and SNR $= 35$ dB.

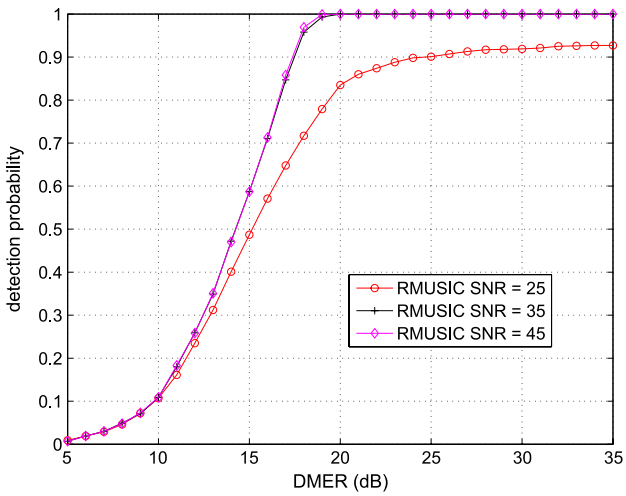


Fig. 5. Detection probabilities of RMUSIC under various DMERs and different SNRs. $\alpha = 0.85$, $N = 8$, the pruned dictionary size is $\tilde{K} = 40$, and the original dictionary size $K = 332$.

with $\alpha = 0.75$ admits very good detection probabilities when DMER ≤ 20 dB; however, when the DMER is higher, using a small α leads to a slight performance degradation. The reason is that a smaller α implies that one is allowed to adjust $\boldsymbol{d}_k$'s more significantly in RMUSIC. Hence, several similar $\boldsymbol{d}_k$'s may be confused with each other. Such a trend is more obvious with smaller α—we see that RMUSIC with $\alpha = 0.55$ and $\alpha = 0.65$ cannot yield reasonable results when DMER gets higher. This observation suggests that a more conservative choice of α should be safer for implementing RMUSIC in practice.

Fig. 4 shows the detection probabilities of RMUSIC and MUSIC under different $\tilde{K}$'s (the size of the pruned dictionary). Setting $\tilde{K}$ to be small may be easier for the SR stage but is considered more aggressive—some spectra corresponding to the ground-truth materials may also be discarded. We see that, when DMER ≥ 15 dB, RMUSIC with $\tilde{K} = 20$ yields higher detection probabilities than that of MUSIC with $\tilde{K} = 60$, and that RMUSIC with a larger $\tilde{K}$ has a better detection performance.

Figs. 5 and 6 show the performance of RMUSIC under various SNRs and different number of underlying ground-truth materials, respectively. From these figures, one can see how this algorithm is scaled by different parameters.

Beginning from Fig. 7, we show the SRE performance of the CSR-based HU algorithms. Specifically, we compare the SREs yielded by the proposed RMUSIC-DANSER and by MUSIC-CSR [8]. We also benchmark our algorithm using RMUSIC-CSR for fairness, since we now have seen that RMUSIC yields much better dictionary-pruning performance. In all of the following simulations, we fix $p = 0.5$, $\mu = 10^5$, and $\tau = 10^{-5}$ for DANSER, no matter how the simulation settings change; the sparsity-controlling parameter λ's for DANSER and CSR are also fixed to be 0.5 and 0.1 unless specified. We stop DANSER if $\|\boldsymbol{C}^{(i)} - \boldsymbol{C}^{(i-1)}\|_F \leq 10^{-5}$, where $\boldsymbol{C}^{(i)}$ denotes the solution at iteration i, or if the number of iterations reaches 5000. The results in all of the following figures of this section are averaged from 50 independent trials.

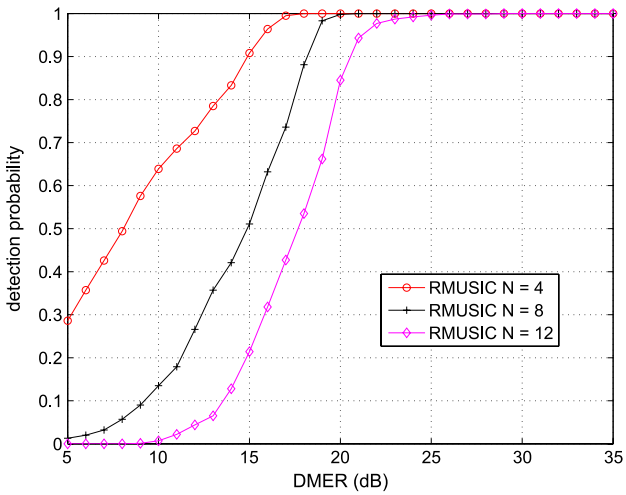


Fig. 6. Detection probabilities of RMUSIC under various DMERs and different N's. $\alpha = 0.85$, $N = 8$, the pruned dictionary size is $\tilde{K} = 60$, the original dictionary size $K = 332$, and SNR $= 35$ dB.

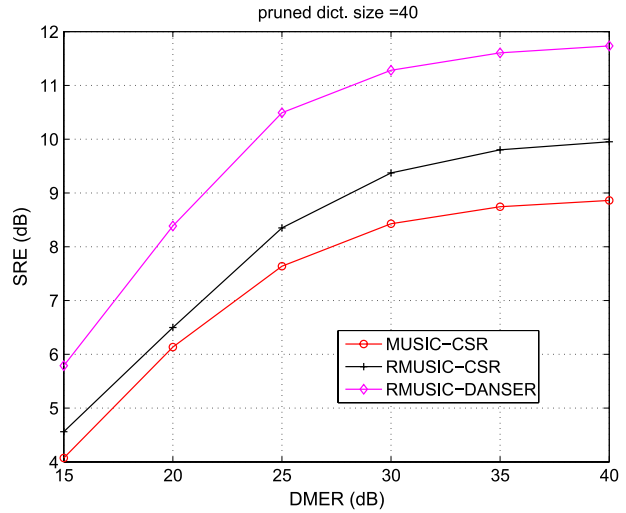


Fig. 7. SREs of the algorithms under different DMERs. $\alpha = 0.85$, $N = 8$, the pruned dictionary size is $\tilde{K} = 40$, the original dictionary size $K = 332$, and SNR $= 35$ dB.

TABLE I
RUNTIMES (IN SECONDS) OF DANSER AND CSR UNDER VARIOUS DMERS. $\alpha = 0.85$, $N = 8$, THE PRUNED DICTIONARY SIZE IS $\tilde{K} = 40$, THE ORIGINAL DICTIONARY SIZE $K = 332$, AND SNR $= 35$ dB

Algorithm	DMER (dB)					
	15	20	25	30	35	40
DANSER	15.9205	16.2639	13.0023	11.0784	10.7352	9.9090
CSR	0.8687	0.9123	1.3611	1.3472	1.6298	1.4699

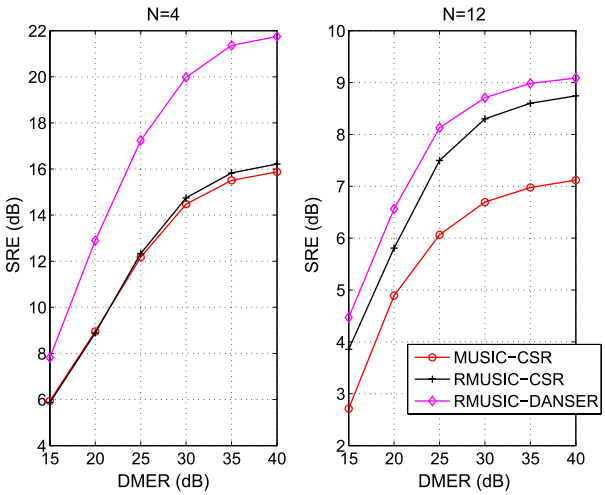


Fig. 8. SREs of the algorithms under different N's. $\alpha = 0.85$, $N = 8$, the pruned dictionary size is $\tilde{K} = 40$, the original dictionary size $K = 332$, and SNR $= 35$ dB.

Fig. 7 shows the SREs of the algorithms under different DMERs. We see that, under all DMERs, RMUSIC-DANSER yields the highest SREs. We see that RMUSIC-CSR also consistently yields better SRE performance than that of MUSIC-CSR—this suggests that RMUSIC itself can help improve the sparse unmixing performance. The runtime performance of DANSER and CSR (i.e., CLSUnSAL) is shown in Table I as a reference. We see that DANSER requires more time to converge compared to CSR, since it also adjusts the dictionary during its updates. Also, when the DMER gets higher, the convergence speed of DANSER improves by 1/3. This intuitively suggests that DANSER does put much effort on adjusting the dictionary (i.e., updating $\boldsymbol{H}$) when the DMER is low.

Figs. 8 and 9 show the performance of the algorithms under different numbers of materials and SNRs, respectively. We see that the results are similar to that in Fig. 7—the SRE performance of RMUSIC-DANSER is consistently higher than the other two under comparison. Note that, for the SNR $= 25$ dB case, we change λ of DANSER and CSR to be 1 and 0.5, respectively, to accommodate the situation where the data are more severely corrupted.

Fig. 10 shows the SREs of the algorithms under different values of $\tilde{K}$. An interesting observation is that using $\tilde{K} = 20$ yields much better unmixing performance than using $\tilde{K} = 60$. These results may shed some light on choosing $\tilde{K}$ in practice—using a large $\tilde{K}$ may safely capture all of the true materials in the pruned dictionary, but it may also degrade the unmixing performance since the SR-type algorithms are, in general, in favor of smaller $\tilde{K}$.

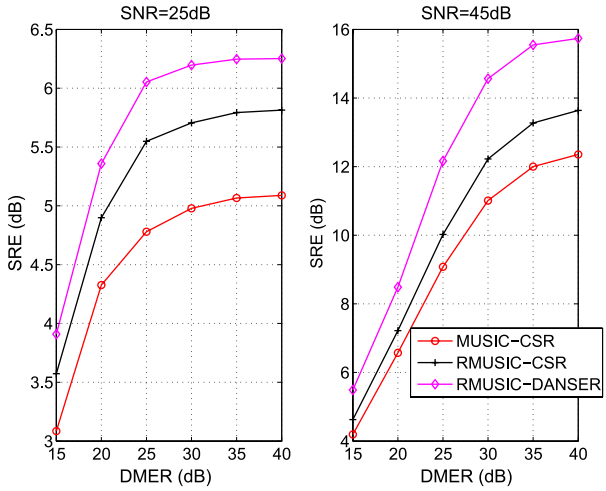


Fig. 9. SREs of the algorithms under different SNRs. $\alpha = 0.85$, $N = 8$, the pruned dictionary size is $\tilde{K} = 40$, and the original dictionary size $K = 332$.

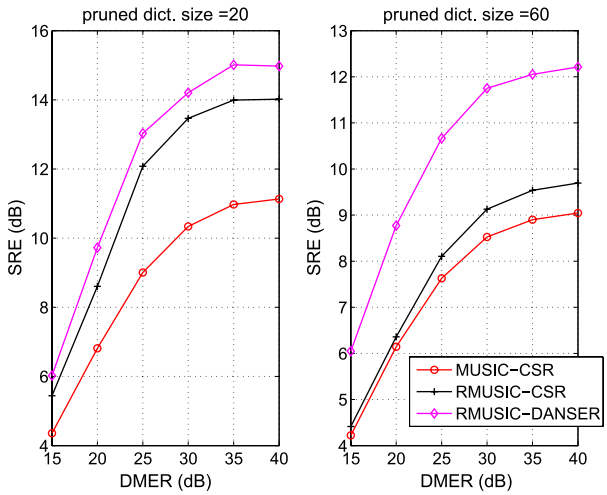


Fig. 10. SREs of the algorithms under different $\tilde{K}$s. $\alpha = 0.85$, $N = 8$, the original dictionary size $K = 332$, and SNR $= 35$ dB.

V. REAL-DATA EXPERIMENT

In this section, we test the algorithms on the AVIRIS Cuprite data set which was captured in Nevada in 1997 (see http://aviris.jpl.nasa.gov/html/aviris.freedata.html). This data set has been studied for years, and the abundance maps of several prominent materials are well recognized. The scene originally has 224 spectral bands between 0.4 and 2.5 μm, with a nominal spectral resolution of 10 nm. Low SNR bands, i.e., bands 1–2, 105–115, 150–170, and 223–224, have been removed, resulting in a total of 188 spectral bands. We take a subimage of the whole data set, which consists of 250 $\times$ 191 pixels (see Fig. 11 for this subimage at spectral band 30). Like the previous computer simulations, the dictionary chosen is a subset of 332 spectral signatures from the USGS library. It has been noticed that there are calibration mismatches between the real image spectra of this scene and the spectra available in the USGS library [5]–[7]. Hence, we directly treat the 332 spectra from USGS library as $\boldsymbol{D}$ that contains model mismatches—which differs from the simulation in the last section.

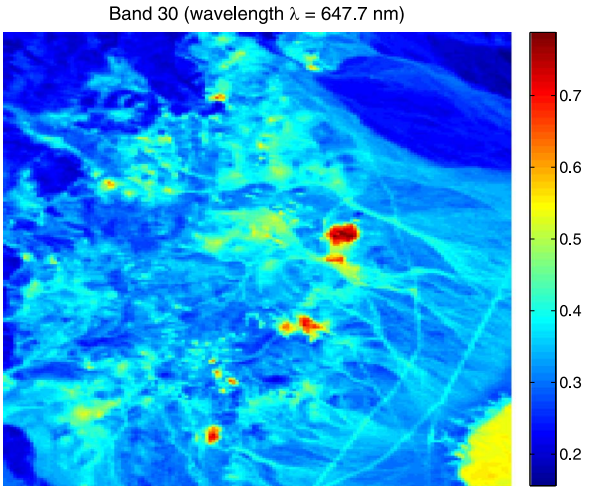


Fig. 11. Band 30 (wavelength $\lambda = 647.7$ nm) of the subimage of AVIRIS Cuprite Nevada data set that is used in the experiment of this section.

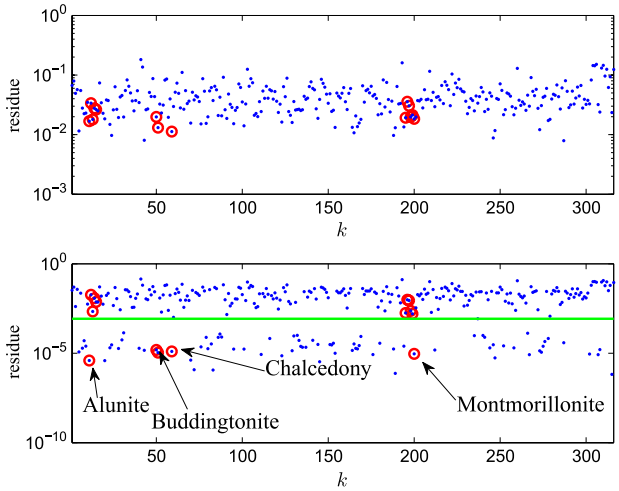


Fig. 12. (Top) Music and (bottom) RMUSIC residues of the real data.

We first apply RMUSIC and MUSIC on this data set. We adopt the following way to evaluate the performance: from the previous studies in [5]–[7], we know that *alunite*, *buddingtonite*, *chalcedony*, and *montmorillonite* are prominent materials in this data set. Fig. 12 shows the residues obtained by applying MUSIC (top) and RMUSIC (bottom). For RMUSIC, we set $\alpha = 0.85$. The red circles correspond to the library members associated with alunite, buddingtonite, chalcedony, and montmorillonite. For the real-data experiment, we see that RMUSIC and MUSIC perform worse than in the simulated cases due to various types of modeling error. We also see that many other materials admit residue levels comparable to those by the materials of interest. Nevertheless, we still find the result by RMUSIC interesting: first, we see that the four endmembers corresponding to the materials of interest indeed have small RMUSIC residues. Second, the residues by RMUSIC can be clearly separated into two groups, and the group with smaller residues includes the spectra of the materials that we wish to identify. Such a result is useful in practice: since the residues associated with the spectra are clearly divided to two clusters, it is easy to decide which spectra should be kept in the pruned dictionary. In this experiment, we simply keep the spectra below the green line, which is drawn by visual inspection, and this results in a pruned dictionary with $\tilde{K} = 79$ spectra. Compared to the original size $K = 332$, RMUSIC successfully reduces the dictionary size by 75% while preserving the spectra associated with the prominent materials.

We follow the method in [5]–[7] to compare the abundance map estimation results of RMUSIC-CSR and RMUSIC-DANSER. We set the parameters of DANSER to be $(p, \lambda, \mu, \tau) = (0.5, 3, 10^3, 10^{-5})$ and the regularization parameter of CSR to be 0.01, respectively. Specifically, we plot the classification maps yielded by the USGS Tetracorder software [44] and the estimated abundance maps of alunite, buddingtonite, chalcedony, and montmorillonite by RMUSIC-CSR and RMUSIC-DANSER in Figs. 13–16. As mentioned in [5]–[7], the classification maps are based on the older version Cuprite data captured in 1995, while the HSI was captured in 1997, which means that the details of the new data may not be fully revealed by the classification maps—but it still makes a good reference for visual evaluation. We see that, for alunite and buddingtonite, RMUSIC-CSR and RMUSIC-DANSER yield similar abundance maps. However, for chalcedony and montmorillonite, the abundances given by RMUSIC-DANSER generally have stronger intensities all over the area of interest. In particular, the abundance map of chalcedony yielded by RMUSIC-DANSER is visually observed to be closer to the corresponding Tetracorder abundance map. Also, if we consider a library member associated with $\boldsymbol{c}^\ell$ such that $\|\boldsymbol{c}^\ell\|_2 \geq \max_{\ell=1,\ldots,L} \|\boldsymbol{c}^\ell\|_2/100$ as an "active endmember," DANSER and CSR were found to yield 15 and 19 active endmembers, respectively, in this experiment. Following [7], we also observe the average per-pixel number of endmembers yielded by the algorithms, which is defined as $\hat{N}_{\mathrm{PP}} = (1/L)\sum_{\ell=1}^{L} |\mathcal{N}_\ell|$, in which $\mathcal{N}_\ell = \{k \mid c_k[\ell] > 0.05\}$. In this experiment, $\hat{N}_{\mathrm{PP}}$ of RMUSIC-CSR is 6.47, and that of RMUSIC-DANSER is 5.84. These observations support our claim that using ℓ_p quasi-norm yields sparser solution.

VI. Conclusion

In this paper, we have developed a dictionary-aided semiblind HU method that takes into account the spectral signature mismatch problem. We have proposed a dictionary-adjusted CSR formulation with a nonconvex collaborative sparsity-promoting regularizer. By a careful reformulation, an alternating optimization algorithm with simple per-iteration updates was proposed. A new dictionary-pruning algorithm based on a spectral mismatch-robust MUSIC criterion was also proposed. Simulations and real-data experiments showed that the proposed algorithms are promising in improving the HU performance compared to the prior works.

Appendix

A. Proof of Proposition 1

First, we claim that any limit point of the solution sequence generated by the DANSER algorithm in Algorithm 1 is a

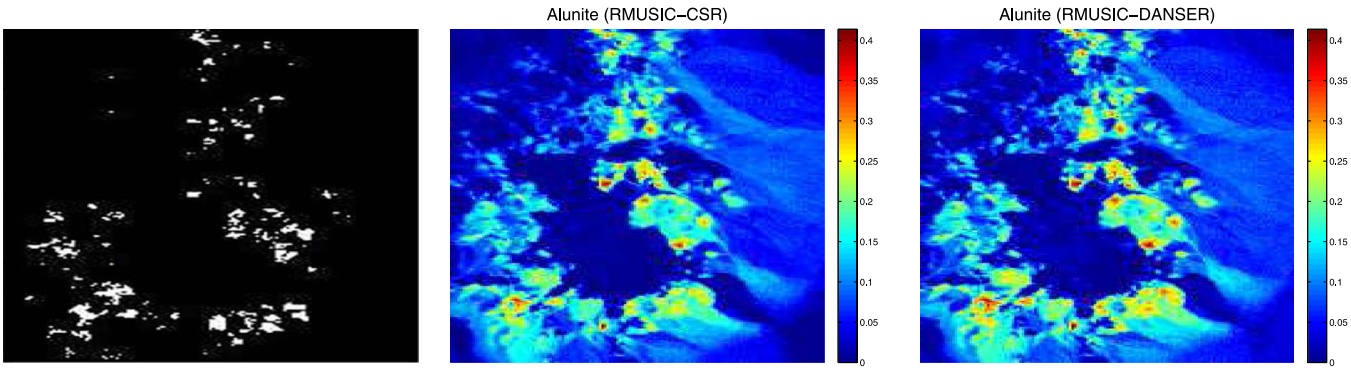


Fig. 13. USGS Tetracorder abundance map (left) and estimated abundance maps of the alunite by RMUSIC-CSR and RMUSIC-DANSER, respectively.

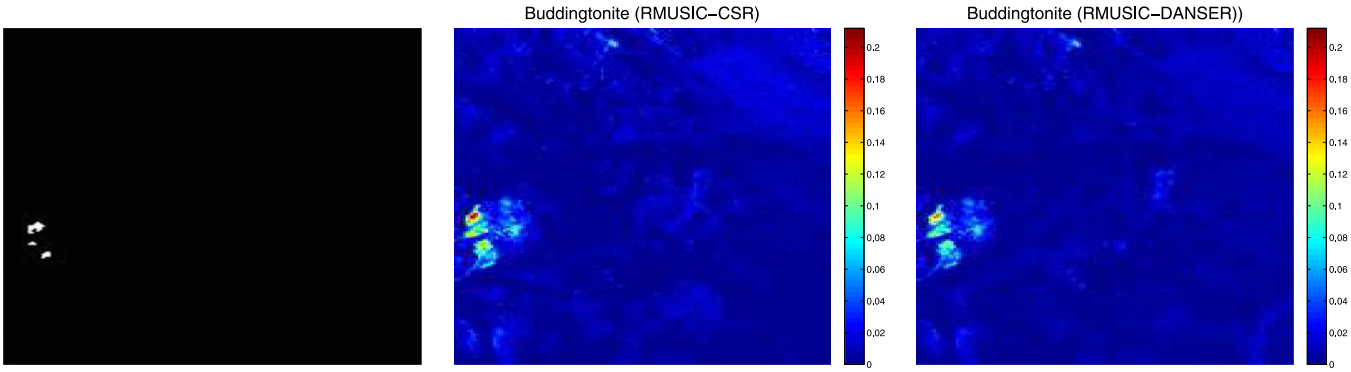


Fig. 14. USGS Tetracorder abundance map (left) and estimated abundance maps of the buddingtonite by RMUSIC-CSR and RMUSIC-DANSER, respectively.

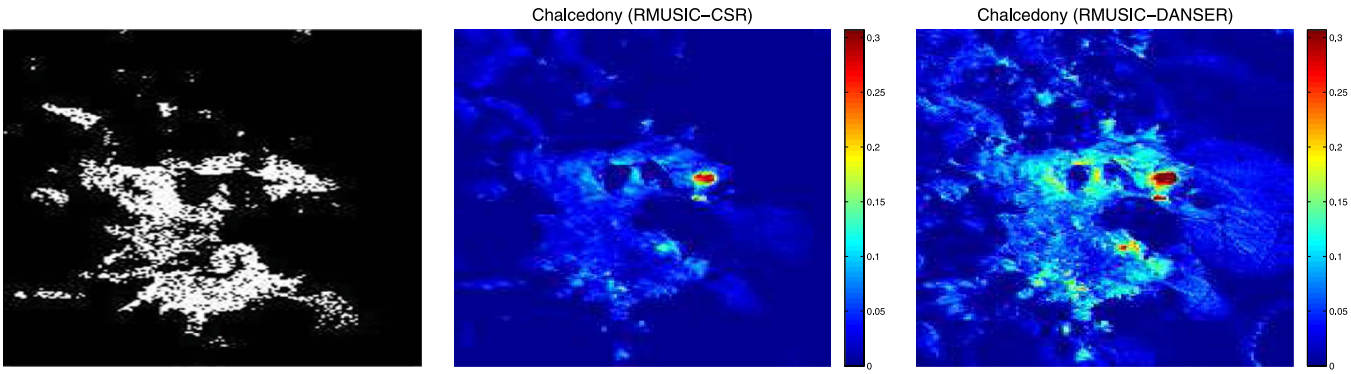


Fig. 15. USGS Tetracorder abundance map (left) and estimated abundance maps of the chalcedony by RMUSIC-CSR and RMUSIC-DANSER, respectively.

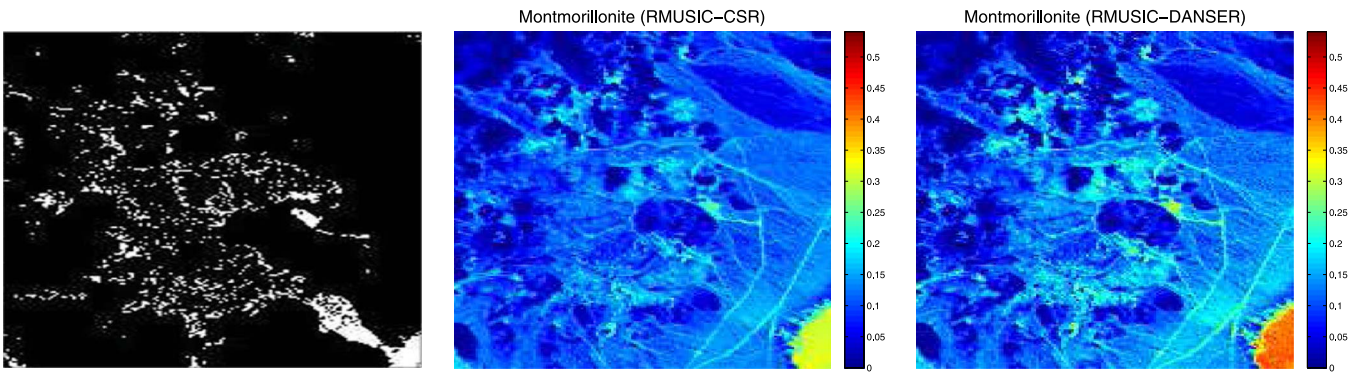


Fig. 16. USGS Tetracorder abundance map (left) and estimated abundance maps of the montmorillonite by RMUSIC-CSR and RMUSIC-DANSER, respectively.

stationary point of Problem (12) (but not Problem (10) at this moment). The claim is obtained by applying a general alternating optimization (AO) result in [45, Prop. 2.7.1], which says that every limit point of a solution sequence generated by an AO algorithm is a stationary point of its tackled problem if each partial optimization problem in AO is strictly convex and has a continuously differentiable objective function within the interior of its feasible set. In our case, one can see that the partial optimizations of Problem (12) w.r.t. $\boldsymbol{H}$, $\boldsymbol{D}'$, $\boldsymbol{c}^1, \ldots, \boldsymbol{c}^K$, and $\{w_k\}$ satisfy the aforementioned condition.

Second, we claim that a stationary point of Problem (12) is also a stationary point of Problem (10). The proof is as follows. For notational convenience, let $\boldsymbol{X} = [\boldsymbol{C}, \boldsymbol{H}^T, (\boldsymbol{D}')^T]$, $\boldsymbol{w} = [w_1, \ldots, w_K]^T$, and denote

$$g(\boldsymbol{X}, \boldsymbol{w}) = h(\boldsymbol{X}) + \lambda \sum_{k=1}^{K} \left(w_k \left\| \boldsymbol{c}^k \right\|_2^2 + \phi_p(w_k) \right)$$

$$f(\boldsymbol{X}) = h(\boldsymbol{X}) + \lambda \sum_{k=1}^{K} \left(\|\boldsymbol{c}^k\|_2^2 + \tau \right)^{\frac{p}{2}}$$

as the objective functions of Problem (12) and Problem (10), respectively, where

$$h(\boldsymbol{X}) = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{C}\|_F^2 + \frac{\mu}{2}\|\boldsymbol{H} - \boldsymbol{D}'\|_F^2.$$

Also, recall from the development in Section III-B that

$$f(\boldsymbol{X}) = \min_{\boldsymbol{w} \geq \boldsymbol{0}} g(\boldsymbol{X}, \boldsymbol{w}).$$

Now, let $(\boldsymbol{X}^\star, \boldsymbol{w}^\star)$ be a stationary point of Problem (12), which, by definition, satisfies

$$\left(\nabla_{\boldsymbol{w}} g(\boldsymbol{X}^\star, \boldsymbol{w}^\star)\right)^T (\boldsymbol{w} - \boldsymbol{w}^\star) \geq 0, \forall\, \boldsymbol{w} \geq \boldsymbol{0} \quad (26a)$$

$$\mathrm{Tr}\left(\left(\nabla_{\boldsymbol{X}} g(\boldsymbol{X}^\star, \boldsymbol{w}^\star)\right)^T (\boldsymbol{X} - \boldsymbol{X}^\star) \right) \geq 0, \forall\, \boldsymbol{X} \in \mathcal{X} \quad (26b)$$

where $\nabla_{\boldsymbol{X}} g(\boldsymbol{X}, \boldsymbol{w})$ and $\nabla_{\boldsymbol{w}} g(\boldsymbol{X}, \boldsymbol{w})$ denote the gradient of $g(\boldsymbol{X}, \boldsymbol{w})$ w.r.t. $\boldsymbol{X}$ and $\boldsymbol{w}$, respectively, and $\mathcal{X}$ denotes the feasible set of $\boldsymbol{X}$ in Problem (12) or Problem (10). From (26a), we observe that

$$g(\boldsymbol{X}^\star, \boldsymbol{w}^\star) = \min_{\boldsymbol{w} \geq \boldsymbol{0}} g(\boldsymbol{X}^\star, \boldsymbol{w}) \quad (27)$$

and the argument is as follows: g is strictly convex w.r.t. $\boldsymbol{w} \geq \boldsymbol{0}$ by Lemma 1, and as a result of the optimality conditions of convex optimization, (26a) holds if and only if $\boldsymbol{w}^\star$ is the optimal solution to $\min_{\boldsymbol{w} \geq \boldsymbol{0}} g(\boldsymbol{X}^\star, \boldsymbol{w})$. Equation (27) implies that $f(\boldsymbol{X}^\star) = g(\boldsymbol{X}^\star, \boldsymbol{w}^\star)$. Consequently, we can rewrite (26b) as

$$\mathrm{Tr}\left((\nabla_{\boldsymbol{X}} f(\boldsymbol{X}^\star))^T (\boldsymbol{X} - \boldsymbol{X}^\star) \right) \geq 0, \forall\, \boldsymbol{X} \in \mathcal{X}.$$

The aforementioned equation is identical to the definition for $\boldsymbol{X}^\star$ to be a stationary point of Problem (10). Hence, we have proven that, for any stationary point $(\boldsymbol{X}^\star, \boldsymbol{w}^\star)$ of Problem (12), the part $\boldsymbol{X}^\star$ is a stationary point of Problem (10).

Finally, combining the aforementioned two claims leads to the conclusion in Proposition 1.

B. Proof of Proposition 2

Recall that we aim at solving

$$\min_{\|\boldsymbol{\xi}\|_2 \leq \epsilon} \eta_k(\boldsymbol{\xi}) \quad (28)$$

where

$$\eta_k(\boldsymbol{\xi}) = \frac{\left\| \boldsymbol{P}_{\boldsymbol{U}_S}^{\perp}(\boldsymbol{d}_k - \boldsymbol{\xi}) \right\|_2}{\|\boldsymbol{P}_{\boldsymbol{U}_S}(\boldsymbol{d}_k - \boldsymbol{\xi})\|_2}. \quad (29)$$

By the triangle inequality, we have

$$\eta_k(\boldsymbol{\xi}) \geq \frac{\left| \left\| \boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{d}_k \right\|_2 - \left\| \boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{\xi} \right\|_2 \right|}{\|\boldsymbol{P}_{\boldsymbol{U}_S} \boldsymbol{d}_k\|_2 + \|\boldsymbol{P}_{\boldsymbol{U}_S} \boldsymbol{\xi}\|_2} \quad (30)$$

where the aforementioned equality holds if and only if 1) $\boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{\xi} = \beta \boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{d}_k$, $\beta \geq 0$ and 2) $\boldsymbol{P}_{\boldsymbol{U}_S} \boldsymbol{\xi} = -\alpha \boldsymbol{P}_{\boldsymbol{U}_S} \boldsymbol{d}_k$, $\alpha \geq 0$. The two aforementioned conditions can be satisfied simultaneously by setting

$$\boldsymbol{\xi} = -\frac{\alpha}{\|\boldsymbol{P}_{\boldsymbol{U}_S} \boldsymbol{d}_k\|_2} \boldsymbol{P}_{\boldsymbol{U}_S} \boldsymbol{d}_k + \frac{\beta}{\|\boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{d}_k\|_2} \boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{d}_k \quad (31)$$

for some $\alpha, \beta \geq 0$. Also, note that $\|\boldsymbol{\xi}\|_2 \leq \epsilon$ is equivalent to

$$\alpha^2 + \beta^2 \leq \epsilon^2. \quad (32)$$

By substituting (31) into $\eta_k(\boldsymbol{\xi})$ and by noting (32), we recast Problem (28) as

$$\min_{\substack{\alpha, \beta \geq 0 \\ \alpha^2 + \beta^2 \leq \epsilon^2}} \frac{\left| \left\| \boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{d}_k \right\|_2 - \beta \right|}{\|\boldsymbol{P}_{\boldsymbol{U}_S} \boldsymbol{d}_k\|_2 + \alpha}. \quad (33)$$

Consider two cases, namely, 1) $\|\boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{d}_k\|_2 \leq \epsilon^2$ and 2) $\|\boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{d}_k\|_2 > \epsilon^2$. For case 1), the optimal β is $\beta = \|\boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{d}_k\|_2$, and the optimal α may take any value in $[0, \sqrt{\epsilon^2 - \|\boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{d}_k\|_2^2}]$. For case 2), we observe the following: fixing β and α should be made as large as possible so as to reduce the objective value. Hence, we can substitute $\alpha = \sqrt{\epsilon^2 - \beta^2}$ (the largest possible α fixing β) into Problem (33) and simplify the problem to

$$\eta_k^\star = \min_{0 \leq \beta \leq \epsilon} \frac{\left| \left\| \boldsymbol{P}_{\boldsymbol{U}_S}^{\perp} \boldsymbol{d}_k \right\|_2 - \beta \right|}{\|\boldsymbol{P}_{\boldsymbol{U}_S} \boldsymbol{d}_k\|_2 + \sqrt{\epsilon^2 - \beta^2}} \quad (34)$$

which is exactly Problem (24).

C. On the Choice of ϵ in (25)

In this appendix, we discuss the rationale of the choice of ϵ in (25). It can be verified that, for any $\|\boldsymbol{\xi}\|_2 \leq \epsilon$, (25) implies

$$\frac{\|\boldsymbol{d}_k\|_2 - \|\boldsymbol{\xi}\|_2}{\|\boldsymbol{d}_k\|_2 + \|\boldsymbol{\xi}\|_2} \geq \alpha. \quad (35)$$

By using the fact that

$$\|\boldsymbol{d}_k\|_2 + \|\boldsymbol{\xi}\|_2 \geq \|\boldsymbol{d}_k - \boldsymbol{\xi}\|_2$$

$$\|\boldsymbol{d}_k\|_2 - \|\boldsymbol{\xi}\|_2 \leq \frac{\|\boldsymbol{d}_k\|_2^2 - \boldsymbol{d}_k^T \boldsymbol{\xi}}{\|\boldsymbol{d}_k\|_2} = \frac{\boldsymbol{d}_k^T) \boldsymbol{d}_k - \boldsymbol{\xi})}{\|\boldsymbol{d}_k\|_2}$$

which can be obtained by the triangle inequality and Cauchy–Schwartz inequality, respectively, we see that (35) implies

$$\frac{(\boldsymbol{d}_k - \boldsymbol{\xi})^T \boldsymbol{d}_k}{\|\boldsymbol{d}_k - \boldsymbol{\xi}\|_2 \|\boldsymbol{d}_k\|_2} \geq \alpha. \tag{36}$$

Equation (36) suggests that the choice of ϵ in (25) ensures that the dictionary endmember $\boldsymbol{d}_k$ and its corrected version $\boldsymbol{d}_k - \boldsymbol{\xi}$ must have their normalized correlation no less than α.

REFERENCES

[1] X. Fu, W.-K. Ma, J. Bioucas-Dias, and T.-H. Chan, "A robust subspace method for semiblind dictionary-aided hyperspectral unmixing," in *Proc. IEEE WHISPERS*, 2014, pp. 1–4.

[2] J. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[3] W.-K. Ma *et al.*, "A signal processing perspective on hyperspectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.

[4] R. Clark *et al.*, "USGS Digital Spectral Library splib06a: U.S. Geological Survey, Digital Data Series 231, 2007. [Online]. Available: http://speclab.cr.usgs.gov/spectral.lib06

[5] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[6] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.

[7] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.

[8] M.-D. Iordache, J. Bioucas-Dias, A. Plaza, and B. Somers, "MUSIC-CSR: Hyperspectral unmixing via multiple signal classification and collaborative sparse regression," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4364–4382, Jul. 2014.

[9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[10] K. E. Themelis *et al.*, "A novel hierarchical Bayesian approach for sparse semisupervised hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 585–599, Feb. 2012.

[11] W. Tang, Z. Shi, Y. Wu, and C. Zhang, "Sparse unmixing of hyperspectral data using spectral *a priori* information," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 770–783, Feb. 2015.

[12] W. Tang, Z. Shi, and Z. Duren, "Sparse hyperspectral unmixing using an approximate l0 norm," *Optik-Int. J. Light Electron Opt.*, vol. 125, no. 1, pp. 31–38, Jan. 2014.

[13] W. Tang, Z. Shi, and Y. Wu, "Regularized simultaneous forward-backward greedy algorithm for sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5271–5288, Sep. 2014.

[14] Z. Shi, W. Tang, Z. Duren, and Z. Jiang, "Subspace matching pursuit for sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3256–3274, Jun. 2014.

[15] R. Feng, Y. Zhong, and L. Zhang, "An improved nonlocal sparse unmixing algorithm for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 915–919, Apr. 2015.

[16] J. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, Mar. 2006.

[17] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505–519, Jan. 2010.

[18] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[19] 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.

[20] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[21] J. M. Kim, O. K. Lee, and J. C. Ye, "Compressive music: Revisiting the link between compressive sensing and array signal processing," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 278–301, Jan. 2012.

[22] B. Somers, G. P. Asner, L. Tits, and P. Coppin, "Endmember variability in spectral mixture analysis: A review," *Remote Sens. Environ.*, vol. 115, no. 7, pp. 1603–1616, Jul. 2011.

[23] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.

[24] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE ICASSP*, 2008, pp. 3869–3872.

[25] B. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, Jan. 1999.

[26] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inv. Prob.*, vol. 24, no. 3, May 2008, Art. no. 035020.

[27] H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.

[28] R. Gribonval, G. Chardon, and L. Daudet, "Blind calibration for compressed sensing by convex optimization," in *Proc. IEEE ICASSP*, 2012, pp. 2713–2716.

[29] C. Bilen, G. Puy, R. Gribonval, and L. Daudet, "Blind sensor calibration in sparse recovery using convex optimization," in *Proc. SAMPTA-10th Int. Conf. Sampling Theory Appl.*, 2013, pp. 1–5.

[30] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatches," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4997–5008, Oct. 2014.

[31] F. Chen and Y. Zhang, "Sparse hyperspectral unmixing based on constrained $\ell_p - \ell_2$ optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1142–1146, Sep. 2013.

[32] J. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[33] Y. Shen, J. Fang, and H. Li, "Exact reconstruction analysis of log-sum minimization for compressed sensing," *IEEE Signal Process. Lett.*, vol. 20, no. 12, pp. 1223–1226, Dec. 2013.

[34] R. Courant *et al.*, "Variational methods for the solution of problems of equilibrium and vibrations," *Bull. Amer. Math. Soc.*, vol. 49, no. 1, pp. 1–23, 1943.

[35] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imaging Sci.*, vol. 1, no. 3, pp. 248–272, 2008.

[36] Y. Xiao and J. Yang, "A Fast Algorithm for Total Variation Image Reconstruction From Random Projections, CERN, Geneva, Switzerland, Jan. 2010. [Online]. Available: http://cds.cern.ch/record/1232976

[37] X. Fu, K. Huang, W.-K. Ma, N. D. Sidiropoulos, and B. Rasmus, "Joint tensor factorization and outlying slab suppression with applications," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6315–6328, Dec. 2015.

[38] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.

[39] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1001–1009, Jul. 2001.

[40] A. Cichocki and A. H. Phan, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. 92, no. 3, pp. 708–721, 2009.

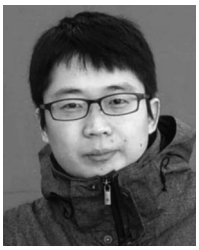
[41] R. Bro and N. D. Sidiropoulos, "Least squares algorithms under unimodality and non-negativity constraints," *J. Chemometrics*, vol. 12, no. 4, pp. 223–247, Jul./Aug. 1998.

[42] A.-L. Zhang, "Quadratic Fractional Programming Problems With Quadratic Constraints," Ph.D. dissertation, Dept. Appl. Math. Phys., Kyoto University, Kyoto, Japan,2008.

[43] W. Dinkelbach, "On nonlinear fractional programming," *Manag. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[44] R. N. Clark *et al.*, "Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems," *J. Geophys. Res., Planets*, vol. 108, no. E12, pp. 1–44, Dec. 2003.

[45] D. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA, Athena Scientific, 1999.



Xiao Fu (S'12–M'15) received the B.Eng. and M.Eng. degrees in communication and information engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2005 and 2010, respectively, and the Ph.D. degree in electronic engineering from The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, in 2014.

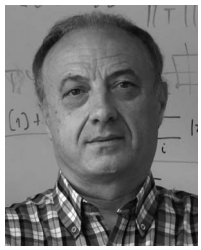
From 2005 to 2006, he was an Assistant Engineer with China Telecom Co. Ltd., Shenzhen, China. He is currently a Postdoctoral Associate with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA. His research interests include signal processing and machine learning, with a recent emphasis on factor analysis and its applications.

Dr. Fu received a Best Student Paper Award at IEEE ICASSP 2014 and was a finalist of the Best Student Paper Competition at IEEE SAM 2014. He also coauthored a paper that received a Best Student Paper Award at IEEE CAMSAP 2015.

Wing-Kin Ma (M'01–SM'11) received the B.Eng. degree in electrical and electronic engineering from the University of Portsmouth, Portsmouth, U.K., in 1995 and the M.Phil. and Ph.D. degrees in electronic engineering from The Chinese University of Hong Kong (CUHK), Shatin, N.T., Hong Kong, in 1997 and 2001, respectively.

He is currently an Associate Professor with the Department of Electronic Engineering, CUHK. From 2005 to 2007, he was an Assistant Professor with the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan. Prior to becoming a faculty member, he held various research positions with McMaster University, Hamilton, ON, Canada, CUHK, and the University of Melbourne, Parkville, Vic., Australia. His research interests are signal processing, communications, and optimization, with a recent emphasis on MIMO transceiver designs and interference management, blind separation and structured matrix factorization, and hyperspectral unmixing in remote sensing.

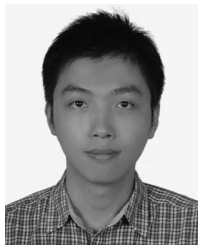
Dr. Ma currently serves as a Senior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and an Associate Editor of *Signal Processing*. Also, he has previously served as an Associate Editor and a Guest Editor of several journals, which include the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, and IEEE SIGNAL PROCESSING MAGAZINE. He was a tutorial speaker in EUSIPCO 2011 and ICASSP 2014. He is currently a member of the Signal Processing Theory and Methods Technical Committee (SPTM-TC) and the Signal Processing for Communications and Networking Technical Committee (SPCOM-TC). He has received several awards, such as the ICASSP Best Student Paper Awards in 2011 and 2014, respectively (by Dr. Ma's students), WHISPERS 2011 Best Paper Award, Research Excellence Award 2013–2014 by CUHK, and 2015 IEEE Signal Processing Magazine Best Paper Award.



José M. Bioucas-Dias (SM'15) received the E.E., M.Sc., Ph.D., and "Agregado" degrees in electrical and computer engineering from Instituto Superior Técnico (IST), Technical University of Lisbon (TULisbon, now University of Lisbon), Lisbon, Portugal, in 1985, 1991, 1995, and 2007, respectively.

Since 1995, he has been with the Department of Electrical and Computer Engineering, IST, where he was an Assistant Professor from 1995 to 2007 and where he has been an Associate Professor since 2007. Since 1993, he has also been a Senior Researcher with the Pattern and Image Analysis Group, Instituto de Telecomunicações, which is a private nonprofit research institution. His research interests include inverse problems, signal and image processing, pattern recognition, optimization, and remote sensing. His research work has been highly cited, and he is included in Thomson Reuters' Highly Cited Researchers 2015 list.

Dr. Bioucas-Dias was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (1997–2000) and IEEE TRANSACTIONS ON IMAGE PROCESSING (2010–2014), and he is a Senior Area Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING and an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was the General Cochair of the 3rd IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing (WHISPERS'2011), and he has been a member of program/technical committees of several international conferences.



Tsung-Han Chan received the B.S. degree in electrical engineering from Yuan Ze University, Taoyuan, Taiwan, in 2004 and the Ph.D. degree from the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan, in 2009.

He is currently a Senior Engineer with MediaTek Inc., Hsinchu. His research interests are image processing and numerical optimization, with a recent emphasis on hyperspectral remote sensing, computer vision, and machine learning.

Dr. Chan was a corecipient of a WHISPERS 2011 Best Paper Award. He was also recognized as an Outstanding Reviewer for Computer Vision and Pattern Recognition 2014 and a Best Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING 2014.