Semiblind Hyperspectral Unmixing in the Presence of Spectral Library Mismatches

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

Abstract—The dictionary-aided sparse regression (SR) approach has recently emerged as a promising alternative to hyperspectral unmixing (HU) 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 tailored-designed for practical implementation 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—Semiblind hyperspectral unmixing, compressive sensing, dictionary mismatch, \( \ell_0 \) quasi-norm sparsity promoting, robust dictionary pruning

I. INTRODUCTION

Hyperspectral unmixing (HU) aims at decomposing pixels of an 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], [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 (U.S.G.S.) 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 the pixels. Such a dictionary-aided semiblind formulation is fundamentally identical to the well-known basis selection or sparse regression 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 sparse regression were proposed in [5]–[8]. The algorithms in [5] and [6], [7] treat the HU problem as a single pixel-based sparse regression (SR) problem and a multiple pixel-based collaborative sparse regression (CSR) problem, respectively. Classic \( \ell_1 \) norm and \( \ell_2/\ell_1 \) 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 \( \ell_1 \) norm and \( \ell_2/\ell_1 \) 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 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 above 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

Part of this work was published in WHISPERS 2014 [1]. This work was supported by the Hong Kong Research Grants Council (RGC) General Research Fund (GRF), under Project CUHK 14205414, and also in part by ARO STIR W911NF-15-1-0384.

X. Fu is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, 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 wkmn@ee.cuhk.edu.hk. J.M. Bioucas-Dias is with the Instituto de Telecomunicações and Instituto Superior Técnico, 1049-1, Lisbon, Portugal, e-mail bioucas@let.ist.ut.pt. T.-H. Chan is with the MediaTek Inc., Hsinchu, Taiwan, e-mail thchan@ieee.org.
finds its application in CS [21]. In dictionary-aided semiblind HU, MUSIC proves to be useful in pre-selecting 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; particularly, MUSIC-based dictionary pruning is sensitive to spectral signature mismatches, as will be seen in the simulations.

**Contributions:** In this work, 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 \( \ell_2/\ell_p \) (\( 0 < p < 1 \)) quasi-norm as the sparsity-promoting function, instead of the convex \( \ell_2/\ell_1 \) 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 sparse regression has received enormous attention in various fields, there are comparatively fewer works that study sparse regression 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 \( \ell_1 \)-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], [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 \( \ell_2/\ell_1 \) norm and its smoothed counterparts. We also note that \( \ell_p \) quasi-norm based sparse regression was applied to HU for single pixel-based unmixing without considering dictionary mismatches [31]. Here, our focus is collaborative sparse regression using multiple pixels, which is known to have both theoretical and practical advantages over the single pixel-based algorithms; we adopt the nonconvex \( \ell_2/\ell_p \) 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 \( x \in \mathbb{R}^n \) and \( X \in \mathbb{R}^{m \times n} \) mean that \( x \) and \( X \) are a real-valued \( n \)-dimensional vector and a real-valued \( m \times n \) matrix, respectively (resp.). The notation \( x \geq 0 \) (resp. \( X \geq 0 \)) means that \( x \) (resp. \( X \)) is element-wise non-negative. The \( i \)th column of a matrix \( X \in \mathbb{R}^{m \times n} \) is denoted by \( x_i \in \mathbb{R}^n \), and the \( j \)th row of \( X \) by \( x^j \). The superscripts “\(^T\)”,” “\(^-1\)” and “\(^\dagger\)” stand for the transpose, matrix inverse and matrix pseudo-inverse, resp. The orthogonal projectors onto the range space of \( X \) and its orthogonal complement are denoted by \( P_X = XX^T X \) and \( P_X = I - P_X \), resp. The \( \ell_p \) norm of a vector \( x \in \mathbb{R}^p \), \( p \geq 1 \), is denoted by \( \|x\|_p = (\sum_{i=1}^m |x_i|^p)^{1/p} \). The \( \ell_p \) quasi-norm, \( 0 < p < 1 \), is denoted by the same notation as above. The mixed \( \ell_p/\ell_q \)-norm or \( \ell_p/\ell_q \)-quasi-norm is denoted by \( \|X\|_{p,q} = (\sum_{i=1}^m \|x_i\|_q^p)^{1/p} \). The Frobenius norm is denoted by \( \|X\|_F = \|X\|_{2,2} \).

**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 hyperspectral image can be modeled as

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

where \( y[\ell] \in \mathbb{R}^M \) denotes the hyperspectral measurement at the \( \ell \)th pixel of the image, with \( M \) being the number of spectral bands; each \( a_n \in \mathbb{R}^M \), \( n = 1, \ldots, N \), represents the spectral signature of a particular material, indexed by \( n \) here;
for some prespecified constant \( \lambda > 0 \) is the abundance of material \( n \) at pixel \( \ell \); \( v[\ell] \in \mathbb{R}^M \) is a noise vector; \( L \) is the number of pixels. For convenience, we will write (1) in a matrix form

\[
Y = AS + V,
\]

where \( Y = [y[1], \ldots, y[L]], A = [a_1, \ldots, a_N], S = [s[1], \ldots, s[L]], s[\ell] = [s_1[\ell], \ldots, s_N[\ell]]^T, \) and \( V = [v[1], \ldots, v[L]] \).

In HU, we aim by identifying \( A \) and \( S \) from \( Y \). 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 is that the dictionary covers the spectral space of a 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,

\[
a_n \in \{d_1, \ldots, d_K\}, \text{ for every } n = 1, \ldots, N.
\]

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

\[
a_n = d_{k_n}, \text{ for some } k_n \in \{1, \ldots, K\}.
\]

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

\[
Y = DC + V,
\]

where \( C \in \mathbb{R}^{K \times L} \) is a row-sparse matrix; to be specific, the \( k_n \)-th row of \( C \), \( n = 1, \ldots, N \), is the \( k \)-th row of \( S \), and the other rows of \( C \) are all zeros.

Let us consider the sparse regression approach—currently the main approach for dictionary-aided semiblind HU. The idea is to exploit the sparsity of \( C \), thereby attempting to recover the indices \( k_1, \ldots, k_N \) correctly and the abundance matrix \( S \) 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:

\[
\min_{C \in \mathbb{R}^{K \times L}} \|Y - DC\|_F^2 + \lambda \|C\|_{2,1}
\]

s.t. \( C \geq 0 \),

for some prespecified constant \( \lambda > 0 \). Here, notice that \( \|C\|_{2,1} = \sum_{i=1}^K \|c_i\|_2 \), which aims at promoting row sparsity of \( C \). As can be seen in Problem (5), CSR seeks to find a nonnegative row-sparse \( C \) that provides a good approximation to \( Y = DC \). 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 sparse regression 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 \( Y = AS \). Let \( U_S \in \mathbb{R}^{M \times N} \) denote a matrix that contains the first \( N \) left singular vectors of \( Y \). It can be shown that in the noiseless case and under some mild assumptions\(^1\), we have

\[
P_{U_S}^\perp d_k = 0 \iff d_k = a_{k_n} \text{ for some } n \in \{1, \ldots, N\}.
\]

The physical meaning of (6) is that if a spectral sample \( d_k \) 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 above observation suggests that we can correctly identify the indices \( k_1, \ldots, k_N \) by the simple closed-form equations on 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

\[
\gamma_{\text{MUSIC}}(k) = \frac{d_k^T P_{U_S} d_k}{\|d_k\|^2_F}.
\]

2. Determine \( \hat{\Lambda} = \{k_1, \ldots, k_N\} \) such that for \( n = 1, \ldots, N \), we have \( \gamma_{\text{MUSIC}}(k_n) < \gamma_{\text{MUSIC}}(j) \) for all \( j \notin \hat{\Lambda} \).

The above 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 \( U_S \) from the noisy \( Y \). 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 sparse regression 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.

III. PROPOSED APPROACH

The crucial assumption with dictionary-aided semiblind HU is that there is no spectrum mismatches; that is, we can

\(^1\)Specifically, we require that \( S \) has full row rank, and that \( \text{spark}(D) > N + 1 \), where \( \text{spark}(X) = r \) means that \( r \) is the smallest number such that any \( r \) columns of \( X \) 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.
always find a dictionary sample that exactly matches an actual spectral signature in the scene; cf. Eq. (3). As discussed in the Introduction, this may not be 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):

\[ \mathbf{d}_k = \mathbf{a}_n + \varepsilon_n, \quad n = 1, \ldots, N, \quad (8) \]

for some \( \varepsilon_n \in \mathbb{R}^M \) that characterizes the mismatch between the presumed and actual spectra of each material. Particularly, every spectral error \( \varepsilon_n \) is assumed to be bounded:

\[ \|\varepsilon_n\|_2 \leq \delta, \quad n = 1, \ldots, N, \]

for some \( \delta > 0 \). Physically, our model assumes that the dictionary still covers all 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 \( \delta^2 \) in terms of magnitude.

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

\[ \mathbf{d}'_k = \mathbf{d}_k + \mathbf{e}_k, \quad k = 1, \ldots, K, \]

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

\[
\begin{align*}
\min_{\mathbf{D}' \in \mathbb{R}^{M \times K}, \mathbf{C} \in \mathbb{R}^{K \times L}} & \quad \frac{1}{2} \| \mathbf{Y} - \mathbf{H} \mathbf{C} \|_F^2 + \mu \| \mathbf{H} - \mathbf{D}' \|_F^2 + \lambda \sum_{k=1}^K (\|\mathbf{e}_k\|_2^2 + \tau)^{p/2} \\
\text{s.t.} & \quad \|\mathbf{d}'_k - \mathbf{d}_k\|_2 \leq \epsilon, \quad k = 1, \ldots, K, \\
& \quad \mathbf{C} \succeq 0,
\end{align*}
\]

where \( \mu, \tau > 0 \), and \( \mathbf{H} \) 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 \( \mathbf{C} \)), 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 \( \phi_p(w) = \frac{2-p}{2} \left( \frac{p}{2} w \right)^{\frac{p}{2-2}} + \tau w \), where \( 0 < p < 2 \), \( \tau > 0 \). The function \( \phi_p(w) \) is strictly convex on \( w \geq 0 \). Also, \( \phi_p(w) \) satisfies the following identity

\[ (x^2 + \tau)^{p/2} = \min_{w \geq 0} w \cdot x^2 + \phi_p(w) \]

and the solution to the problem above is uniquely given by

\[ w_{\text{opt}} = \frac{p}{2} \left( x^2 + \tau \right)^{\frac{2-p}{2}}. \]

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

\[
\begin{align*}
\min_{\mathbf{H}, \mathbf{C}, \mathbf{D}', \{w_k\}} & \quad \frac{1}{2} \| \mathbf{Y} - \mathbf{H} \mathbf{C} \|_F^2 + \mu \| \mathbf{H} - \mathbf{D}' \|_F^2 + \lambda \sum_{k=1}^K \left( w_k \|\mathbf{e}_k\|_2^2 + \phi_p(w_k) \right) \\
\text{s.t.} & \quad \|\mathbf{d}'_k - \mathbf{d}_k\|_2 \leq \epsilon, \quad k = 1, \ldots, K, \\
& \quad \mathbf{C} \succeq 0, \\
& \quad w_k \geq 0, \quad k = 1, \ldots, K,
\end{align*}
\]

Now, our strategy is to perform alternating optimization w.r.t. \( \mathbf{H}, \mathbf{D}', \{w_k\}, \mathbf{e}', \ldots, \mathbf{e}^K \). 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. \( \mathbf{H} \). One can easily see that the solution is

\[ \mathbf{H} := (\mu \mathbf{D}' + \mathbf{Y} \mathbf{C}^T) (\mathbf{C} \mathbf{C}^T + \mu \mathbf{I})^{-1}. \]

Second, the optimization w.r.t. \( \mathbf{D}' \) is separable w.r.t. \( d_1', \ldots, d_K' \), i.e., for \( k = 1, \ldots, K \), we have
\[
\min_{d_k'} \| d_k' - h_k \|_2^2
\]
\[
\text{s.t. } \| d_k' - d_k \|_2 \leq \epsilon.
\]
Problem (14) is a projection problem and the solution is
\[
d_{k}' := \begin{cases} h_k, & \| h_k - d_k \|_2 \leq \epsilon \\ d_k + \epsilon \frac{h_k - d_k}{\| h_k - d_k \|_2}, & \text{otherwise} \end{cases},
\]
(15)

Third, to obtain the solution w.r.t. \( c_k \), let us first re-write the optimization w.r.t. \( C \) as
\[
\min_C \| \hat{\mathbf{Y}} - \hat{\mathbf{H}} C \|_F^2
\]
\[
\text{s.t. } C \geq 0,
\]
where
\[
\hat{\mathbf{Y}} = \begin{bmatrix} \sqrt{2} \mathbf{Y} \\ 0 \end{bmatrix}, \quad \hat{\mathbf{H}} = \begin{bmatrix} \sqrt{2} \mathbf{H} \\ \text{Diag}(\theta) \end{bmatrix},
\]
and \( \theta := [\sqrt{w_1 \lambda}, \ldots, \sqrt{w_K \lambda}]^T \). Then, the subproblem w.r.t. \( c_k \) can be expressed as
\[
\min_{c_k} \| \hat{\mathbf{Y}}_k - \hat{h}_k c_k \|_F^2
\]
\[
\text{s.t. } c_k \geq 0,
\]
where
\[
\hat{\mathbf{Y}}_k = \begin{bmatrix} \sqrt{2} \mathbf{Y} - \sum_{j \neq k} \sqrt{2} h_j c_j \\ 0 \end{bmatrix}, \quad \hat{h}_k = \begin{bmatrix} \sqrt{2} h_k \\ \sqrt{w_k \lambda} f_k \end{bmatrix},
\]
where \( f_k \) is the \( k \)th column of the \( K \times K \) identity matrix. Problem (16) is known to have a simple solution [40], [41], given by
\[
(c_k)_{T} := \left( \begin{bmatrix} \hat{\mathbf{Y}}^T \hat{\mathbf{H}} \end{bmatrix}_{:,k} - C^T [\hat{\mathbf{H}}^T \hat{\mathbf{H}}]_{:,k} + (c_{k})^T [\hat{\mathbf{H}}^T \hat{\mathbf{H}}]_{:,k,k} \right) / [\hat{\mathbf{H}}^T \hat{\mathbf{H}}]_{:,k,k}
\]
(17)
where \( (x)_{+} = \max\{0, x\} \). Notice that using the update (17) is desirable: the large matrix products \( \mathbf{Y}^T \hat{\mathbf{H}} \) and \( \hat{\mathbf{H}}^T \hat{\mathbf{H}} \) both only need to be calculated once before updating \( c_{1}, \ldots, c_{K} \).

Finally, by Lemma 1, the solution w.r.t. \( \{w_k\} \) is
\[
w_k = (p/2)(\| c_k \|_2^2 + \tau)^{(p-2)/2}, \quad k = 1, \ldots, K.
\]
(18)

The alternating optimization process described above 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 above 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 \( C \). 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.

**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 \( CC^T \) and \( \mathbf{Y}C^T \) cost \( O(K^2 L) \) and \( O(MKL) \) flops, respectively; and the matrix inversion requires \( O(K^3) \) flops. Thus, although solving the problems w.r.t. \( c_k \) 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 prun 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 subsection.

**Algorithm 1: DANSER**

Input: \( (\lambda, \tau, p, \mu, \epsilon); \mathbf{D}; C \) (initialization); \( \mathbf{Y} \).

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

2. repeat

3. Unmixing: construct \( \theta := [\sqrt{w_1 \lambda}, \ldots, \sqrt{w_K \lambda}]^T \);

4. \( \hat{\mathbf{Y}} = \begin{bmatrix} \sqrt{2} \mathbf{Y} \\ 0 \end{bmatrix}, \quad \hat{\mathbf{H}} = \begin{bmatrix} \sqrt{2} \mathbf{H} \\ \text{Diag}(\theta) \end{bmatrix} \);

5. let \( F := \hat{\mathbf{Y}}^T \hat{\mathbf{H}} \) and \( G := \hat{\mathbf{H}}^T \hat{\mathbf{H}} \);

6. for \( k = 1 : K \) do

7. update \( c_k \) by

8. \[
(c_k)_{T} := \left( \begin{bmatrix} \hat{\mathbf{Y}}^T \hat{\mathbf{H}} \end{bmatrix}_{:,k} - C^T [\hat{\mathbf{H}}^T \hat{\mathbf{H}}]_{:,k} + (c_{k})^T [\hat{\mathbf{H}}^T \hat{\mathbf{H}}]_{:,k,k} \right) / [\hat{\mathbf{H}}^T \hat{\mathbf{H}}]_{:,k,k}
\]

9. end

Dictionary Adjusting: update \( H \) by

10. \( H := (\mu \mathbf{D}' + \mathbf{Y} C^T) \left( CC^T + \mu \mathbf{I} \right)^{-1} \).

Update Slack Variable:

11. for \( k = 1 : K \) do

12. \[
d_{k}' := \begin{cases} h_k, & \| h_k - d_k \|_2 \leq \epsilon \\ d_k + \epsilon \frac{h_k - d_k}{\| h_k - d_k \|_2}, & \text{otherwise}. \end{cases}
\]

13. end

Reweighting: update \( \{w_k\} \) by

14. \[
w_k = (p/2)(\| c_k \|_2^2 + \tau)^{(p-2)/2}, \quad k = 1, \ldots, K.
\]

15. until some stopping criterion is satisfied;

Output: \( C \).
Consider the MUSIC procedure back in Section II.B. In particular, recall that the metric
\[ \gamma_{\text{MUSIC}}(k) = \frac{d_k^T P_{U_S^T} d_k}{||d_k||_2^2} \] (19)
should yield a small value when \( d_k \) 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
\[ \gamma_{\text{RMUSIC}}(k) = \min_{\xi \in \mathbb{R}^n} \left( \frac{(d_k - \xi)^T P_{U_S^T} (d_k - \xi)}{||d_k - \xi||_2^2} \right) \]
\[ \text{s.t. } ||\xi||_2 \leq \epsilon, \] (20a)
where \( \epsilon > 0 \) is prespecified. The idea is the same as the DANSER development in the above subsections—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
\[ \gamma_{\text{RMUSIC}}(k) = \min_{||\xi||_2 \leq \epsilon} \left( \frac{||P_{U_S^T} (d_k - \xi)||_2^2}{||d_k - \xi||_2^2} \right) \]
\[ = \min_{||\xi||_2 \leq \epsilon} \left( \frac{\eta^2(\xi)}{\eta^2(\xi) + 1} \right), \] (21)
where we recall that \( P_{U_S} = U_S U_S^T \) is the orthogonal projector onto the range space of \( U_S \), and
\[ \eta_k(\xi) = \frac{||P_{U_S^T} (d_k - \xi)||_2}{||d_k - \xi||_2}. \] (22)
Since the objective function of (21) is a monotone increasing function of \( \eta^2(\xi) \in [0, \infty) \), computing \( \gamma_{\text{RMUSIC}}(k) \) is the same as finding the minimal value of \( \eta_k(\xi) \) subject to \( ||\xi||_2 \leq \epsilon \). Let us denote
\[ \eta^*_k = \min_{||\xi||_2 \leq \epsilon} \frac{||P_{U_S^T} (d_k - \xi)||_2}{||d_k - \xi||_2}. \] (23)
We show that

**Proposition 2** The optimal value of Problem (23) can be found by solving a single-variable problem
\[ \eta^*_k = \min_{0 \leq \theta \leq \epsilon} \frac{||P_{U_S^T} d_k||_2 - \theta}{||P_{U_S^T} d_k||_2 + \sqrt{\epsilon^2 - \theta^2}}. \] (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 \( K (K < K) \) spectral samples from \( D \), form a size-\( K \) dictionary, denoted by \( \tilde{D} \) here, and then use \( \tilde{D} \) 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

1. apply HySiMe \[32\] on \( Y \) to obtain \( U_S \);
2. for \( k = 1 : K \) do
3. \[ \eta^*_k = \min_{0 \leq \theta \leq \epsilon} \frac{||P_{U_S^T} d_k||_2 - \theta}{||P_{U_S^T} d_k||_2 + \sqrt{\epsilon^2 - \theta^2}}; \]
4. \[ \gamma_{\text{RMUSIC}}(k) = \frac{||P_{U_S^T} d_k||_2}{||d_k||_2}; \]
5. end
6. determine \( \tilde{\Lambda} = \{k_1, \ldots, k_{\tilde{K}}\} \) such that
7. \[ \gamma_{\text{RMUSIC}}(i) < \gamma_{\text{RMUSIC}}(j) \text{ for any } i \in \tilde{\Lambda} \text{ and } j \notin \tilde{\Lambda}; \]
8. feed \( \tilde{D} \) and \( Y \) to DANSER (Algorithm 1);
9. output: \( C \).

**IV. COMPUTER SIMULATIONS**

In this section, we use synthetic hyperspectral images to show the effectiveness of the proposed approach. In our simulations, the ‘ground-truth’ spectra are randomly selected from a subset of the U.S.G.S. library that has 332 spectral signatures; in this subset, the angles between any two endmembers are larger than \( 3^\circ \), and the 2-norms of the endmembers are larger than one. The ‘available dictionary’, \( D \), is formed by the same subset of spectra, but a perturbation (i.e., \( e_k \) 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:

\[ \text{DMER(dB)} = 10 \log_{10} \left( \frac{||d_k||_2^2}{\delta^2} \right), \]

where \( k^* = \arg \min_{k=1, \ldots, K} ||d_k||_2 \) and \( \delta = \max_{k=1, \ldots, K} ||e_k||_2 \). 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 \( L = 5000 \). We also define the signal-to-noise ratio (SNR) as \( \text{SNR} = \frac{\sum_{k=1}^K ||A^k(\ell)||_2^2}{ML\sigma^2} \) to quantify the noise level, where \( \sigma^2 \) denotes the variance of the additive noise, which is also assumed to be zero-mean i.i.d. Gaussian. The choice of the parameter \( \epsilon \) is as follows
\[ \epsilon = \frac{1 - \alpha}{1 + \alpha} ||d_k||_2, \] (25)
where \( \alpha \in [0, 1] \) is given. The parameter \( \alpha \) controls the normalized correlation between the RMUSIC/DANSER-resulted
dictionary member \(d_k - \xi\) and the original dictionary member \(d_k\). Specifically, under \(|\xi|_2 \leq \epsilon\), it can be shown that the choice of \(\epsilon\) in (25) leads to \(\frac{(d_k - \xi)^\top d_k}{|d_k|_2^2} \geq \alpha\); the derivation is shown in Appendix C. Note that \(\epsilon\) need not be identical across different dictionary members under the RMUSIC/DANSER framework: If some dictionary members have larger spectral signature mismatches, larger \(\epsilon\)'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 \(D\). Here, we randomly pick \(N = 6\) spectra as the ground-truth materials, and then use the described dictionary \(D\) to observe the performance of MUSIC and RMUSIC. The parameter of RMUSIC is set to be \(\alpha = 0.85\), and we set DMER = 20dB and SNR = 35dB in this case. We see that MUSIC has difficulty in distinguishing several ground-truth spectra from the dictionary members (to be precise, the third and the 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 DANSER, we set \(p = 0.5\), \(\lambda = 0.04\) 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 \(C\) than that of RMUSIC-CSR, and all of the desired spectra have been successfully 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-true spectra, and \(\Lambda \subseteq \{1, \ldots, K\}\) denotes an index selection subset outputted by a dictionary pruning algorithm. Also, we will use \(K\) to denote the size of the pruned dictionary. Second, to measure the unmixing performance, we calculate the following the signal to reconstruction error (SRE) [5]–[7]:

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

where \(C\) is the true row-sparse abundance matrix (see (2) in Section II.A), and \(\hat{C}\) 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 35dB, and \(K = 40\) 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 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 with \(\alpha = 0.75\) admits very good detection probabilities when DMER \(\leq 20\)dB; however, when the DMER is higher, using a small \(\alpha\) leads to a slight performance degradation. The reason is that a smaller \(\alpha\) implies that one is allowed to adjust \(d_k\)'s more significantly in RMUSIC. Hence, several similar \(d_k\)'s may be confused with each other. Such a trend is more obvious with smaller \(\alpha\) — 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 \(\alpha\) should be safer for implementing RMUSIC in practice.

Fig. 4 shows the detection probabilities of RMUSIC and MUSIC under different \(K\)'s (the size of the pruned dictionary). Setting \(K\) to be small may be easier for the sparse regression...
stage, but is considered more aggressive—some spectra corresponding to the ground-truth materials may also be discarded. We see that when DMER $\geq 15$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 DMERs and different $\tilde{K}$’s (size of the pruned dictionary). $\alpha = 0.85$; $N = 8$; the original dictionary size $K = 332$.

Fig. 5 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 $H$) when the DMER is low.

TABLE I

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>DMER (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DANSER</td>
<td>15.9205</td>
</tr>
<tr>
<td>CSR</td>
<td>0.8687</td>
</tr>
</tbody>
</table>

Figs. 8 and 9 show the performance of the algorithms under different number of materials and SNRs, resp. 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 = 25dB case, we change $\lambda$ of DANSER and CSR to be 1 and 0.5, respectively, to accommodate the situation where the data is 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 $\tilde{K} = 60$. This results may shed some light on choosing $\tilde{K}$ in practice - using a large $\tilde{K}$ may safely capture all the true materials in the pruned dictionary, but it may also degrade the unmixing performance since the sparse regression-type algorithms are in general in favor of smaller $\tilde{K}$.

V. REAL DATA EXPERIMENT

In this section, we test the algorithms on the AVIRIS Cuprite data set which was captured in Nevada, 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 $\mu$m, with 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 × 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 U.S.G.S. library. It has been noticed that there are calibration mismatches between the real image spectra of this scene and the spectra available in the U.S.G.S. library [5]–[7]. Hence, we directly treat the 332 spectra from U.S.G.S. library as D that contains model mismatches — which differs from the simulation in the last section.

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 α = 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 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 include 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 into 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 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, λ, μ, τ) = (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 U.S.G.S. Tetra-corder software [44] and the estimated abundance maps of Alunite, Buddingtonite, Chalcedony, and Montmorillonite by RMUSIC-CSR and RMUSIC-DANSER in Fig. 13 - Fig. 16. As mentioned in [5]–[7], the classification maps are based on the older version Cuprite data captured in 1995, while the hyperspectral image 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. Particularly, the abundance map of Chalcedony yielded by RMUSIC-DANSER is visually observed to be closer to the corresponding Tetra-corder abundance map. Also, if we consider a library member associated with εₖ such that ∥εₖ∥₂ ≥ maxₑ₌₁,...,L ∥εₑ∥₂/100 as an ‘active endmember’, DANSER and CSR were found to yield 15 and 19 active endmembers, resp., in this experiment. Following [7], we also observe the average per-pixel number of endmembers yielded by the algorithms, which is defined as \( N_{PP} = (1/L) \sum_{ℓ=1}^{L} |N_k| \), in which \( N_k = \{ k \mid c_k[ℓ] > 0.05 \} \). In this experiment, \( N_{PP} \) of RMUSIC-CSR is 6.47, and that of RMUSIC-DANSER is 5.84. These observations support our claim that using \( ε_p \) quasi-norm yields sparser solution.

VI. CONCLUSION

In this work, we have developed a dictionary-aided semilind HU method that takes into account the spectral signature mismatch problem. We proposed a dictionary-adjusted semilind 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
Fig. 13. The U.S.G.S. Tetracorder abundance map (left) and the estimated abundance map of the Alunite by RMUSIC-CSR and RMUSIC-DANSER, respectively.

Fig. 14. The U.S.G.S. Tetracorder abundance map (left) and the estimated abundance map of the Buddingtonite by RMUSIC-CSR and RMUSIC-DANSER, respectively.

Fig. 15. The U.S.G.S. Tetracorder abundance map (left) and estimated abundance map of the Chalcedony by RMUSIC-CSR and RMUSIC-DANSER, respectively.

Fig. 16. The U.S.G.S. Tetracorder abundance map (left) and estimated abundance map of the Montmorillonite by RMUSIC-CSR and RMUSIC-DANSER, respectively.
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 stationary point of Problem (12) or Problem (10). The proof is as follows. For notational convenience, let \( X = [C, H^T, (D')^T] \), \( w = [w_1, \ldots, w_K]^T \), and denote

\[
g(X, w) = h(X) + \lambda \sum_{k=1}^{K} \left( w_k \parallel e^k \parallel^2_2 + \phi_p(w_k) \right),
\]

\[
f(X) = h(X) + \lambda \sum_{k=1}^{K} \left( \parallel e^k \parallel^2_2 + \tau \right)^{p/2}
\]

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

\[
h(X) = \frac{1}{2} \parallel Y - HC \parallel^2_F + \frac{\mu}{2} \parallel H - D' \parallel^2_F,
\]

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

\[
f(X) = \min_{w \geq 0} g(X, w).
\]

Now, let \( (X^*, w^*) \) be a stationary point of Problem (12), which, by definition, satisfies

\[
(\nabla_w g(X^*, w^*))^T (w - w^*) \geq 0, \forall w \geq 0,
\]

\[
\text{Tr} \left( (\nabla_X g(X^*, w^*))^T (X - X^*) \right) \geq 0, \forall X \in \mathcal{X},
\]

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

\[
g(X^*, w^*) = \min_{w \geq 0} g(X^*, w),
\]

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

\[
\text{Tr} \left( (\nabla_X f(X^*))^T (X - X^*) \right) \geq 0, \forall X \in \mathcal{X}.
\]

The above equation is identical to the definition for \( X^* \) to be a stationary point of Problem (10). Hence, we have proven that for any stationary point \( (X^*, w^*) \) of Problem (12), the part \( X^* \) is a stationary point of Problem (10).

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

B. Proof of Proposition 2

Recall that we aim at solving

\[
\min_{\parallel \xi \parallel_2 \leq \epsilon} \eta_k(\xi),
\]

where

\[
\eta_k(\xi) = \frac{\parallel P_{U^k} d_k - \xi \parallel_2}{\parallel P_{U^k} d_k - \xi \parallel_2},
\]

By the triangle inequality, we have

\[
\eta_k(\xi) \geq \frac{\parallel P_{U^k} d_k \parallel - \parallel P_{U^k} \xi \parallel_2}{\parallel P_{U^k} d_k \parallel_2 + \parallel P_{U^k} \xi \parallel_2},
\]

where equality above holds if and only if i) \( P_{U^k} \xi = \beta P_{U^k} d_k \), \( \beta \geq 0 \), and ii) \( P_{U^k} \xi = -\alpha P_{U^k} d_k, \alpha \geq 0 \). The two conditions above can be satisfied simultaneously by setting

\[
\xi = -\frac{\alpha}{\parallel P_{U^k} d_k \parallel_2} P_{U^k} d_k + \frac{\beta}{\parallel P_{U^k} d_k \parallel_2} P_{U^k}^+ d_k,
\]

for some \( \alpha, \beta \geq 0 \). Also, note that \( \parallel \xi \parallel_2 \leq \epsilon \) is equivalent to

\[
\alpha^2 + \beta^2 \leq \epsilon^2.
\]

Substituting (31) into \( \eta_k(\xi) \), and by noting (32), we recast Problem (28) as

\[
\min_{\alpha^2 + \beta^2 \leq \epsilon^2} \frac{\parallel P_{U^k}^+ d_k \parallel_2 - \beta}{\parallel P_{U^k} d_k \parallel_2 + \alpha}.
\]

Consider two cases, namely, i) \( \parallel P_{U^k}^+ d_k \parallel_2 \leq \epsilon^2 \), and ii) \( \parallel P_{U^k}^+ d_k \parallel_2 > \epsilon^2 \). For case i), the optimal \( \beta \) is \( \beta = \parallel P_{U^k}^+ d_k \parallel_2 \), and the optimal \( \alpha \) may take any value in \( 0, \sqrt{\epsilon^2 - \parallel P_{U^k}^+ d_k \parallel_2^2} \). For case ii), we observe the following: fixing \( \beta \), \( \alpha \) 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 \( \alpha \) fixing \( \beta \)) into Problem (33) and simplify the problem to

\[
\eta_k^* = \min_{\alpha^2 + \beta^2 \leq \epsilon^2} \frac{\parallel P_{U^k}^+ d_k \parallel_2 - \beta}{\parallel P_{U^k} d_k \parallel_2 + \sqrt{\epsilon^2 - \beta^2}},
\]

which is exactly Problem (24).

C. On the Choice of \( \epsilon \) in (25)

In this appendix we discuss the rationale of the choice of \( \epsilon \) in (25). It can be verified that for any \( \parallel \xi \parallel_2 \leq \epsilon \), (25) implies

\[
\frac{\parallel d_k \parallel_2^2 - \parallel \xi \parallel_2^2}{\parallel d_k \parallel_2^2 + \parallel \xi \parallel_2^2} \geq \alpha.
\]
By using the fact that
\[ \|d_k\|_2 + \|\xi\|_2 \geq \|d_k - \xi\|_2, \]
\[ \|d_k\|_2 - \|\xi\|_2 \leq \frac{\|d_k\|_2 - \|\xi\|_2}{\|d_k\|_2} = \frac{\|d_k - \xi\|_2}{\|d_k\|_2}, \]
which can be obtained by the triangle inequality and Cauchy-Schwartz inequality, resp., we see that (35) implies
\[ (d_k - \xi)^T d_k \geq \alpha. \]  
Eq. (36) suggests that the choice of \( \epsilon \) in (25) ensures that the dictionary endmember \( d_k \) and its corrected version \( d_k - \xi \) must have their normalized correlation no less than \( \alpha \).

REFERENCES

Xiao Fu (S'12-M'15) received his 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. In 2014, he received his Ph.D. degree in Electronic Engineering from the Chinese University of Hong Kong (CUHK), Shatin, N.T., Hong Kong. From 2005 to 2006, he was an Assistant Engineer at China Telecom Co. Ltd., Shenzhen, China. He is currently a Postdoctoral Associate at the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, United States. 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 co-authored 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, both in electronic engineering, from The Chinese University of Hong Kong (CUHK), 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 also an Assistant Professor with the Institute of Communications Engineering, National Tsing Hua University, Taiwan, in 2004 and his Ph.D. degree in electrical and computer engineering. Dr. Ma currently serves as Senior Area Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING and Associate Editor of Signal Processing. Also, he has previously served as Associate Editor and Guest Editor of several journals, which include 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 EUROSIPCO 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). Dr. Ma has received several awards, such as 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 Bioucas-Dias (SM’15) received the EE, MSc, PhD, and “Agregado” degrees from Instituto Superior Técnico (IST), Technical University of Lisbon (TULisbon, now University of Lisbon), Portugal, in 1985, 1991, 1995, and 2007, respectively, all in electrical and computer engineering. Since 1995, José Bioucas-Dias has been with the Department of Electrical and Computer Engineering, IST, where he was an Assistant Professor from 1995 to 2007 and an Associate Professor since 2007. Since 1993, he is also a Senior Researcher with the Pattern and Image Analysis group of the Instituto de Telecomunicações, which is a private non-profit 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.

Tsung-Han Chan received his B.S. degree from the Department of Electrical Engineering, Yuan Ze University, Taiwan, in 2004 and his Ph.D. degree from the Institute of Communications Engineering, National Tsing Hua University, Taiwan, in 2009. He is currently working as a Senior Engineer with MediaTek Inc., Hsinchu, Taiwan. He was a co-recipient 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. His research interests are in image processing and numerical optimization, with a recent emphasis on hyperspectral remote sensing, computer vision, and machine learning.